

Exact Thresholds for Noisy Non-Adaptive Group Testing

Junren Chen

Jonathan Scarlett

Abstract

In recent years, the mathematical limits and algorithmic bounds for probabilistic group testing having become increasingly well-understood, with exact asymptotic thresholds now being known in general scaling regimes for the noiseless setting. In the noisy setting where each test outcome is flipped with constant probability, there have been similar developments, but the overall understanding has lagged significantly behind the noiseless setting. In this paper, we substantially narrow this gap by deriving exact asymptotic thresholds for the noisy setting under two widely-studied random test designs: i.i.d. Bernoulli and near-constant tests-per-item. These thresholds are established by combining components of an existing information-theoretic threshold decoder with a novel analysis of maximum-likelihood decoding (upper bounds), and deriving a novel set of impossibility results by analyzing certain failure events for optimal maximum-likelihood decoding (lower bounds). Our results show that existing algorithmic upper bounds for the noisy setting are strictly suboptimal, and leave open the interesting question of whether our thresholds can be attained using computationally efficient algorithms.

Contents

1	Introduction	3
1.1	Problem Setup	3
1.2	Related Work	5
1.3	Overview of Our Results	7
2	Outline of Converse Proofs	11
3	Outline of Achievability Proofs	13
3.1	Information-Theoretic Tools	13
3.2	A Hybrid Decoding Rule	14
3.3	Analysis of the Low Overlap (High ℓ) Regime	15
3.4	Analysis of the High Overlap (Low ℓ) Regime	15
4	Conclusion	16

J. Chen is with the Department of Mathematics, The University of Hong Kong. The work was done when he was a visiting Ph.D. student in the Department of Computer Science, National University of Singapore (NUS). J. Scarlett is with the Department of Computer Science, Department of Mathematics, and Institute of Data Science, National University of Singapore (NUS). e-mail: chenjr58@connect.hku.hk; scarlett@comp.nus.edu.sg

A	Roadmap of the Proofs	17
A.1	Overview of the Appendices	18
A.2	Note on the Scaling of n	18
A.3	Useful Technical Lemmas	19
A.4	Table of Notation	20
B	Necessary and Sufficient Conditions for Maximum-Likelihood Decoding	20
C	Low-ℓ Converse Analysis for Bernoulli Designs	24
D	Low-ℓ Achievability Analysis for Bernoulli Designs	30
E	Low-ℓ Converse Analysis for Near-Constant Weight Designs	41
F	Low-ℓ Achievability Analysis for Near-Constant Weight Designs	50
G	Technical Lemmas for Near-Constant Weight Designs	60
H	High-ℓ Achievability Analysis for Both Designs	64
H.1	Initial Results	64
H.2	Bernoulli Design	66
H.3	Near-Constant Weight Design	66
I	High-ℓ Converse Analysis for Both Designs	67
I.1	Bernoulli Design	67
I.2	Near-Constant Weight Design	68
J	Proofs of Technical Results in the High-ℓ Analysis for the Near-Constant Weight Design	69
J.1	Proof of Lemma 11 (Concentration Bound for $\frac{\ell}{k} \rightarrow \alpha \in [0, 1]$)	70
J.2	Proof of Lemma 10 (Mutual Information for the Near-Constant Weight Design)	74
J.3	Proof of Lemma 13 (Characterization of $\mathbb{P}(V)$)	81

1 Introduction

Group testing is a fundamental combinatorial and statistical problem with roots in medical testing and a variety of more recent applications including DNA testing, data storage, communication protocols, and COVID-19 testing. Alongside its uses in applications, group testing has had a long history of mathematical developments characterizing the number of tests required for reliable recovery, with increased interest in recent years [6, 20].

In particular, following a number of recent advances (which we describe in more detail in Section 1.2), a complete characterization is now available for the asymptotic number of tests required to achieve an error probability approaching zero in the noiseless setting [3, 4, 11, 15, 17, 29, 39]. Moreover, efficient algorithms are available whose performance matches these fundamental limits, with the Definite Defective (DD) algorithm [5, 29] or individual testing [2] sufficing in “less sparse” settings, and a spatial coupling approach [17] closing the remaining gaps in “more sparse” settings.

While similar advances have also been made in *noisy* group testing, its understanding has lagged significantly behind the noiseless setting. In particular, under the widely-considered symmetric noise model, optimal thresholds on the number of tests are only known in a narrow range of (sufficiently sparse) scaling regimes [39], and the degree of (sub-)optimality of practical algorithms is unknown [24, 40, 41].

In this paper, we address this gap by obtaining *exact information-theoretic thresholds* for the symmetric noise setting under two of the most commonly-considered test designs: i.i.d. Bernoulli and near-constant tests-per-item. Our results provide a significant improvement over the existing state-of-the-art, and demonstrate that all of the existing bounds for practical algorithms (to our knowledge) are suboptimal. We believe that this is a significant step forward in understanding noisy group testing, though we highlight that there are still two open problems that are beyond the scope of this work: A tight converse for *general test designs*, and an *efficient algorithm* that can attain the information-theoretic thresholds that we derive.

1.1 Problem Setup

Let p denote the number of items, k the number of defective items, and n the number of tests. We consider the standard probabilistic group testing setup in which the defective set S is uniform over the $\binom{p}{k}$ subsets of $[p]$ of cardinality k , where $[p] := \{1, \dots, p\}$. Each test is represented by a length- p binary vector $X = (X_1, \dots, X_p)$, with $X_j = 1$ if item j is included in the test, and $X_j = 0$ otherwise. We adopt the standard symmetric noise model (e.g., see [13, 39]), in which the resulting test outcome is given by

$$Y = \bigvee_{j \in S} X_j \oplus Z, \tag{1}$$

where $Z \sim \text{Bernoulli}(\rho)$ for some $\rho \in (0, \frac{1}{2})$, and \oplus denotes modulo-2 addition. With n tests, there are n such test vectors $X^{(1)}, \dots, X^{(n)}$, and we represent these via an $n \times p$ matrix \mathbf{X} . We only consider non-adaptive testing, in which the entire matrix \mathbf{X} must be specified prior to observing

any test results; such designs are often preferred due to allowing the implementation of the tests in parallel. The test outcomes $Y^{(1)}, \dots, Y^{(n)}$ are represented via a length- n vector \mathbf{Y} , and we will sometimes use \mathbf{Z} to denote the corresponding vector of noise variables. We assume independent noise across tests, i.e., $Y^{(1)}, \dots, Y^{(n)}$ are conditionally independent given \mathbf{X} .

Given the test matrix \mathbf{X} and the corresponding outcomes \mathbf{Y} , a *decoder* forms an estimate \hat{S} , and the error probability is given by

$$P_e := \mathbb{P}[\hat{S} \neq S]. \quad (2)$$

Here the probability is taken with respect to the uniformly random defective set, the possibly-randomized tests, and the noise. As was done in prior works such as [17, 39], we consider the goal of attaining $P_e \rightarrow 0$ (with as few tests as possible), without seeking a precise convergence rate.

We will state our theorems in terms of bounds on n , but in our discussions and figures, we will instead use the notion of *rate* [6]:

$$\text{Rate} = \lim_{p \rightarrow \infty} \frac{\log_2 \binom{p}{k}}{n} \quad (\text{bits/test}), \quad (3)$$

where k and n implicitly depend on p . Thus, a higher rate amounts to fewer tests, and the rate is in $[0, 1]$ since the tests are binary-valued.

We focus on the sublinear sparsity regime, in which $k = \Theta(p^\theta)$ for some sparsity parameter $\theta \in (0, 1)$. While the linear regime $k = \Theta(p)$ is also of interest, there are strong hardness results in this regime showing that even in the noiseless setting, (almost) every item must be tested individually to attain $P_e \rightarrow 0$ [2], or similarly even to attain $P_e \not\rightarrow 1$ [11]. Thus, exact recovery is overly stringent for this regime, and one needs to move to other recovery criteria (e.g., approximate recovery [42]).

Finally, we describe the two randomized test designs that we will consider in this paper:

- Under the *Bernoulli design* (e.g., see [3, 13, 33]), each item is independently placed in each test with probability $\frac{\nu}{k}$ for some $\nu > 0$.
- Under the *near-constant weight design* (e.g., see [15, 29]), each item is independently placed in $\Delta = \frac{\nu n}{k}$ tests chosen uniformly at random with replacement, for some $\nu > 0$. (We leave integer rounding implicit, as it does not affect our results.) Since the same test may be chosen multiple times, the weight of each column in \mathbf{X} may be (slightly) below Δ .

Below we will outline the existing results regarding these designs, and then state and discuss our new results.

Note on notation. In most cases, capital letters (e.g., S and \mathbf{X}) are used for random variables (or random vectors/matrices), and lower-case letters (e.g., s and \mathbf{x}) are used for specific realizations. We use bold/non-bold symbols to distinguish multi-test/single-test quantities; for instance, even though $X^{(i)}$ is a vector, is non-bold because it corresponds to a single test. Asymptotic notation such as $O(\cdot)$ and $o(\cdot)$ is with respect to $p \rightarrow \infty$, with quantities such as k and n implicitly depending on p . We will often use the binary entropy function $H_2(\rho) = \rho \log \frac{1}{\rho} + (1 - \rho) \log \frac{1}{1 - \rho}$, the binary KL divergence $D(a \| b) = a \log \frac{a}{b} + (1 - a) \log \frac{1 - a}{1 - b}$, and the shorthand $a \star b = ab + (1 - a)(1 - b)$. The

function $\log(\cdot)$ has base e , and information quantities are measured in nats (except in our plots, where we convert to bits).

1.2 Related Work

Two important defining features of the group testing problem are probabilistic vs. combinatorial (i.e., whether the recovery guarantee is high-probability or definite) and non-adaptive vs. adaptive (i.e., whether or not the tests may be designed sequentially based on previous outcomes). We will mainly summarize existing work for the probabilistic non-adaptive setting with sublinear sparsity (i.e., $k = \Theta(n^\theta)$ for some $\theta \in (0, 1)$), since this is the focus of our work.

Noiseless group testing. Some of the main developments in the theory of noiseless probabilistic group testing in the sublinear regime are outlined as follows:

1. Information-theoretic characterization of the very sparse regime $k = O(1)$ [32];
2. Initial algorithmic bounds via the simple Combinatorial Orthogonal Matching Pursuit (COMP) algorithm with Bernoulli designs [13];
3. Improved algorithmic bounds via the more sophisticated (but still simple) Definite Defectives (DD) algorithm with Bernoulli designs [5];
4. Precise information-theoretic bounds for Bernoulli designs [39];
5. Ensemble tightness¹ of these information-theoretic bounds for Bernoulli designs [4];
6. Improved algorithmic bounds for COMP and DD via the near-constant weight design [29];
7. Precise information-theoretic bounds for the near-constant weight design, including a proof of their ensemble tightness [15];
8. A proof that the bounds in the preceding dot point are optimal among *all* designs, and a computationally efficient method for attaining this optimal threshold [17].

These results for the noiseless setting are summarized in Figure 1 (Left). We also briefly note that computational considerations for the standard (non spatially coupled) random designs were studied in [18, 28], with evidence both for and against the prospect of a statistical-computational gap.

Noisy group testing. In the noisy setting (with the symmetric noise model (1)), several counterparts to the above developments are available, but the list is significantly less complete:

- Regarding #1, the very sparse regime $k = O(1)$ has long been well-understood [32].
- Regarding #2, algorithmic bounds for COMP-like algorithms were established in [13, 40].
- As an attempt towards #4, information-theoretic upper bounds were established in [41], but with tightness only shown for very small values of θ .

¹By ensemble tightness, we mean establishing that the bounds derived are the best possible for the specific random design considered, implying an exact threshold. This leaves open the possibility that other designs may be better.

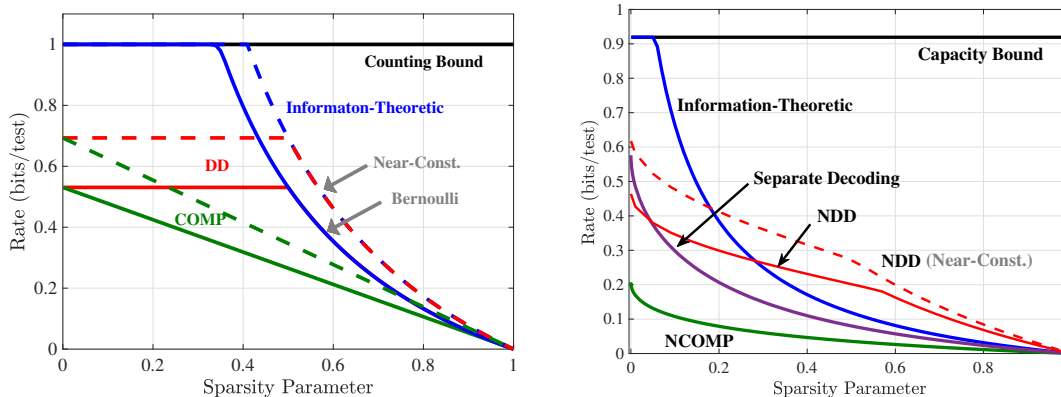


Figure 1: Existing bounds for the noiseless setting (Left) and the noisy setting with $\rho = 0.01$ (Right). The sparsity parameter $\theta \in (0, 1)$ is the value such that $k = \Theta(p^\theta)$, and the rate is the asymptotic limit of $\frac{1}{n} \log_2 \binom{p}{k}$. In the noiseless setting, the dashed curve labeled “Information-Theoretic” is not only optimal for the near-constant weight design, but also for arbitrary test designs. The curves “Counting Bound” and “Capacity Bound” represent simple algorithm-independent lower bounds on n for arbitrary test designs (e.g., see [10, 33]).

- Regarding #3, and #6, bounds were derived for a noisy DD-like algorithm in [41] for the Bernoulli design, and in [24] for the near-constant weight design. Exact thresholds were established in certain (dense) regimes for one-sided noise models, but not for the more commonly-considered symmetric noise model.

These existing results are summarized in Figure 1 (Right).

We will outline our contributions below, but they can be succinctly summarized as follows:

In this paper, we resolve the gaps in the entire list above for the noisy setting, except item #8.

In particular, we complete item #4 (precise-information-theoretic bounds; Bernoulli) which was only partially commenced in [39], and we complete items #5 (ensemble tightness; Bernoulli) and #7 (counterparts for the near-constant weight design), which have not been attempted before.

Other settings. We only provide a brief summary of other aspects of group testing that are less relevant to the present paper:

- For adaptive designs, near-optimal algorithms have long been known for the noiseless setting [27], and analogs were recently derived for the noisy setting [37, 43] and shown to significantly improve over the best non-adaptive results in denser (high θ) regimes. We note that even following the improved non-adaptive bounds we derive in the present paper, significant improvements via adaptivity still remain.
- Combinatorial non-adaptive group testing is a widely-studied problem [20–22, 30], and has also been studied under adversarial noise [14]. Under the condition of exact recovery, the stricter recovery criterion (namely, zero probability of error) increases the dependence on k

from linear to quadratic, meaning that it comes at a significant price in the number of tests. To our knowledge, precise constant factors have not been sought in this line of works.

- If the exact recovery criterion (2) is relaxed to approximate recovery, where αk false positives and false negatives are allowed in the reconstruction for some $\alpha > 0$, then the study of noisy (and noiseless) probabilistic group testing becomes much simpler, with [39] giving upper and lower bounds that match in the limit of small α . (With reference to the bound (4) below, the “difficult” second term disappears, and only the “simple” first term remains.) Analogous findings for the near-constant weight design can be inferred from [17, Sec. 5], and related studies of “all-or-nothing” phase transitions in group testing can be found in [35, 44].

1.3 Overview of Our Results

In this paper, we obtain exact information-theoretic thresholds for both the Bernoulli design and the near-constant weight design. To do so, we prove four results separately: Bernoulli achievability, Bernoulli converse, near-constant weight achievability, and near-constant-weight converse, where we use the term “converse” in a design-specific sense (i.e., ensemble tightness). We emphasize that all of our achievability results are based on a computationally intractable decoder that explicitly searches over all $\binom{p}{k}$ possible defective sets; thus, we only prove achievability in an information-theoretic sense, and we defer efficient algorithms to future work.

Our results are summarized in the four theorems stated below, whose proofs are outlined in Section 3 (achievability) and Section 2 (converse), with the formal details deferred to the appendices. We briefly recall the notation $a \star b = ab + (1 - a)(1 - b)$,² the binary KL divergence $D(a\|b) = a \log \frac{a}{b} + (1 - a) \log \frac{1-a}{1-b}$, and the binary entropy function $H_2(a) = a \log \frac{1}{a} + (1 - a) \log \frac{1}{1-a}$. The function $\log(\cdot)$ and the information measures are in units of nats, except where stated otherwise.

We note that both of the thresholds that we derive are somewhat complicated to state, but we will provide some intuition in Sections 2 and 3 regarding how the various terms arise. In all of our results, we make use of the design parameter $\nu > 0$, noise level $\rho \in (0, \frac{1}{2})$, and sparsity parameter $\theta \in (0, 1)$, which is the constant such that $k = \Theta(p^\theta)$. Since $\theta \in (0, 1)$, we have $k \rightarrow \infty$ and $k = o(p)$.

Threshold for Bernoulli designs: The threshold for Bernoulli design with i.i.d. Bernoulli($\frac{\nu}{k}$) entries is given by

$$n_{\text{Bern}}^* = \max \left\{ \frac{k \log \frac{p}{k}}{H_2(e^{-\nu} \star \rho) - H_2(\rho)}, \frac{k \log \frac{p}{k}}{(1 - \theta)\nu e^{-\nu} \min_{C > 0, \zeta \in (0, 1)} \max\{\frac{1}{\theta} f_1^{\text{Bern}}(C, \zeta, \rho), f_2^{\text{Bern}}(C, \zeta, \rho)\}} \right\}, \quad (4)$$

²This is interpreted as the probability of getting a “1” when a $\text{Bern}(1 - a)$ variable is generated and then flipped with probability b .

where we define

$$f_1^{\text{Bern}}(C, \zeta, \rho) = C \log C - C + C \cdot D(\zeta \| \rho) + 1, \quad (5)$$

$$f_2^{\text{Bern}}(C, \zeta, \rho) = g^{\text{Bern}}(C, \zeta, d_{\text{Bern}}^*, \rho), \quad (6)$$

and where $g^{\text{Bern}}(C, \zeta, \rho, d)$ and d_{Bern}^* are defined as

$$g^{\text{Bern}}(C, \zeta, d, \rho) = \rho d \log d + (\rho d - C(1 - 2\zeta)) \log \left(\frac{\rho d - C(1 - 2\zeta)}{1 - \rho} \right) + 1 - 2\rho d + C(1 - 2\zeta), \quad (7)$$

$$d_{\text{Bern}}^* = \frac{C(1 - 2\zeta) + \sqrt{C^2(1 - 2\zeta)^2 + 4\rho(1 - \rho)}}{2\rho}, \quad (8)$$

with the latter implicitly depending on (C, ζ, ρ) . As we will argue in the proof, d_{Bern}^* is the minimizer of g^{Bern} with respect to d .

The achievability and converse statements are formally given as follows.

Theorem 1. (Achievability via Bernoulli Designs) *Under the Bernoulli design with i.i.d. entries following $\text{Bernoulli}(\frac{\nu}{k})$ for fixed $\nu > 0$, there exists a decoding strategy such that $P_e \rightarrow 0$ as $p \rightarrow \infty$ provided that*

$$n \geq (1 + \eta)n_{\text{Bern}}^* \quad (9)$$

for an arbitrary constant $\eta > 0$.

Theorem 2. (Converse for Bernoulli Designs) *Under the Bernoulli design with i.i.d. entries following $\text{Bernoulli}(\frac{\nu}{k})$ for fixed $\nu > 0$, any decoding strategy must have $P_e \rightarrow 1$ as $p \rightarrow \infty$ whenever*

$$n \leq (1 - \eta)n_{\text{Bern}}^* \quad (10)$$

for an arbitrary constant $\eta > 0$.

Threshold for Near-Constant Weight Designs: The threshold for the near-constant weight design with $\Delta = \frac{\nu n}{k}$ placements per item is given by

$$n_{\text{NC}}^* = \max \left\{ \frac{k \log \frac{p}{k}}{H_2(e^{-\nu} \star \rho) - H_2(\rho)}, \frac{k \log \frac{p}{k}}{(1 - \theta)\nu e^{-\nu} \min_{C \in (0, e^\nu), \zeta \in (0, 1)} \max \left\{ \frac{1}{\theta} f_1^{\text{NC}}(C, \zeta, \rho, \nu), f_2^{\text{NC}}(C, \zeta, \rho, \nu) \right\}} \right\}, \quad (11)$$

where

$$f_1^{\text{NC}}(C, \zeta, \rho, \nu) = e^\nu D(Ce^{-\nu} \| e^{-\nu}) + C \cdot D(\zeta \| \rho) \quad (12)$$

$$f_2^{\text{NC}}(C, \zeta, \rho, \nu) = g^{\text{NC}}(C, \zeta, d_{\text{NC}}^*, \rho, \nu), \quad (13)$$

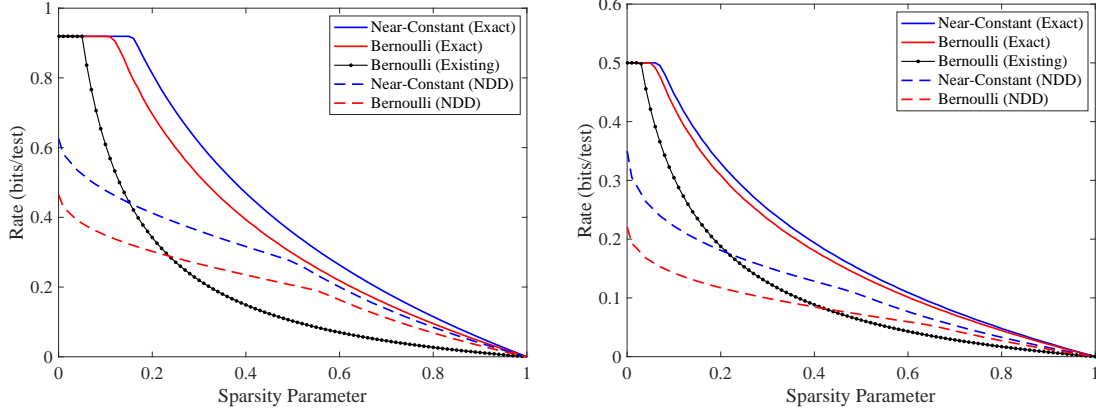


Figure 2: Our exact thresholds and the state-of-the-art existing results for noise levels $\rho = 0.01$ (Left) and $\rho = 0.11$ (Right). The existing Bernoulli achievability result is from [39], and the existing noisy definite defectives (NDD) curves are from [24, 41]. All curves are optimized over ν , except that of [39] which was only proved for $\nu = \ln 2$.

and where $g^{\text{NC}}(C, \zeta, d, \rho, \nu)$ and d_{NC}^* are given by

$$g^{\text{NC}}(C, \zeta, d, \rho, \nu) = e^\nu \cdot D(de^{-\nu} \| e^{-\nu}) + d \cdot D\left(\frac{1}{2} + \frac{C(1-2\zeta)}{2d} \parallel \rho\right) \quad (14)$$

$$d_{\text{NC}}^* = \arg \min_d g(C, \zeta, d, \rho, \nu) \text{ subject to } |C(1-2\zeta)| \leq d \leq e^\nu. \quad (15)$$

Here d_{NC}^* implicitly depends on (C, ζ, ν) and can in fact be expressed in closed form, but the expression is rather complicated so is omitted here (see (218) in Appendix E).

Theorem 3. (Achievability via Near-Constant Weight Designs) *Under the near-constant weight design where each item is placed $\Delta = \frac{\nu n}{k}$ tests uniformly at random with replacement with $\nu = \Theta(1)$, there exists a decoding strategy such that $P_e \rightarrow 0$ as $p \rightarrow \infty$ provided that*

$$n \geq (1 + \eta)n_{\text{NC}}^* \quad (16)$$

for an arbitrary constant $\eta > 0$.

Theorem 4. (Converse for Near-Constant Weight Designs) *Under the near-constant weight design where each item is placed $\Delta = \frac{\nu n}{k}$ tests uniformly at random with replacement with $\nu = \Theta(1)$, any decoding strategy must have $P_e \rightarrow 1$ as $p \rightarrow \infty$ whenever $p \rightarrow \infty$ whenever*

$$n \leq (1 - \eta)n_{\text{NC}}^* \quad (17)$$

for an arbitrary constant $\eta > 0$.

We note that the subscripts/superscripts “Bern” and “NC” (e.g., on f_1 and f_2) will be omitted throughout the proofs in the appendices, since the choice of design will be clear from the context.

Discussion. We illustrate our results numerically using the notion of rate (measured in bits/test), as defined in (3). The bounds for both designs are illustrated in Figure 2, where we observe a substantial improvement over the best known existing bounds. The horizontal part at small θ in Figure 2 corresponds to the capacity of the binary symmetric channel (namely, $\log 2 - H_2(\rho)$), and the corresponding impossibility result holds for arbitrary test designs and even adaptive algorithms (e.g., see [10, 33]). Thus, in this regime, we have particularly strong guarantees of asymptotic optimality. By comparison, for higher θ , we have not established optimality with respect to arbitrary designs, but we have at least established ensemble tightness (and hence exact thresholds) for the two random designs under consideration.

Among the practical algorithms, the most promising one for being asymptotically optimal in some regimes (namely, θ close to one) is noisy DD, since DD is known to exhibit such optimality in the noiseless setting [3] and under the one-sided “Z” and “reverse Z” noise models [24, 41]. However, our results indicate that under symmetric noise, the existing noisy DD bounds are strictly suboptimal for all $\theta \in (0, 1)$ (at least for the ρ values we considered). Hence, more sophisticated techniques appear to be needed to attain our thresholds in a computationally efficient manner, with the spatial coupling approach of [17] perhaps being a promising candidate.

Another interesting implication of our results concerns the optimal choice of ν for the two designs. In the noiseless setting, the following is well known [4, 29]:

- For the Bernoulli design, $\nu = \log 2$ is optimal for $\theta \leq \frac{1}{3}$, and $\nu = 1$ is optimal for almost all higher θ except for a narrow “transition region”.
- For the near-constant weight design, $\nu = \log 2$ is optimal for all θ .

In the noisy case, under both designs, the choice $\nu = \log 2$ is still optimal for sufficiently small θ such that the capacity bound $n = \frac{k \log \frac{2}{k}}{\log 2 - H_2(\rho)}(1 + o(1))$ is achieved. This has the natural interpretation that in this regime, ν should be chosen to make the tests maximally informative in the sense of attaining $H(Y) = \log 2$ (i.e., each test is equally likely to be positive or negative).

On the other hand, when θ is large enough, the division by $1 - \theta$ makes the second term in each bound become dominant, and optimizing ν eventually amounts to optimizing that term. (There is also a “transition region” where the optimal ν is chosen to equate the two terms.) In the Bernoulli design, the only ν dependence is via $\nu e^{-\nu}$, which is the same term that enters in the noiseless setting, so we again have $\nu = 1$ being optimal for large enough θ (see Figure 3 (Right) for an illustration). However, for the near-constant weight design, the choice of ν becomes more complicated even at high θ values. In particular, we found that in contrast to the noiseless setting, the choice $\nu = \log 2$ can be strictly suboptimal (again see Figure 3). Having said this, the degree of suboptimality is small, and we found that $\nu = \log 2$ is still a generally good choice.

While the proofs of our main results all have highly technical components and are quite lengthy, the general approaches taken are conceptually simple. We provide conceptual outlines in Sections 2 and 3, and provide the technical details in the appendices, starting with a roadmap in Appendix A. The outlines will mostly be the same for the two designs, despite the differing technical details.

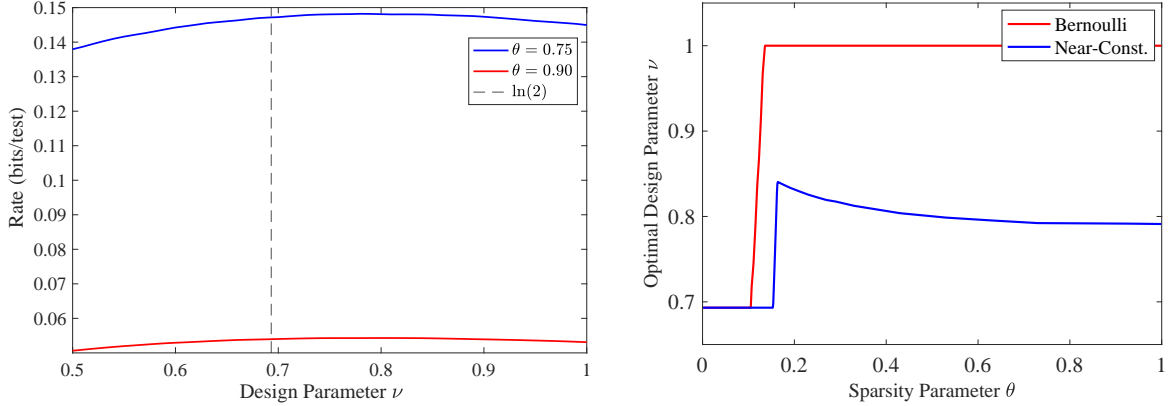


Figure 3: Illustrations regarding the optimal design parameter ν (varying with the sparsity parameter θ) with noise level $\rho = 0.01$. (Left) Example plots of the rate as a function of ν under the near-constant weight design. (Right) Optimal ν as a function of the sparsity parameter θ under both designs. We note that the sudden increases after the horizontal parts are not instantaneous; both curves are continuous with respect to θ .

2 Outline of Converse Proofs

We first outline the proofs of our converse results (i.e., lower bounds on the number of tests).

First term. The thresholds in (4) and (11) consist of two terms. The converse for the first term *maximized over ν* is known even for *arbitrary test designs* via a simple channel capacity argument [33, 38], and for arbitrary ν , the converse for Bernoulli designs is also known via [39]. We required non-minor additional technical effort to handle the near-constant weight design with arbitrary ν ,³ but the intuition behind the bound is simple: With parameter ν , each test has probability roughly $e^{-\nu}$ of containing no defectives, and thus a probability roughly $e^{-\nu} \star \rho$ of being positive. A standard information-theoretic argument shows that each test can only reveal $H_2(e^{-\nu} \star \rho) - H_2(\rho)$ bits of information. On the other hand, with $\binom{p}{k}$ possible defective sets, we require roughly $\log \binom{p}{k} = (k \log_2 \frac{p}{k})(1 + o(1))$ bits to identify the correct one. Comparing these quantities leads to the first term in (11). This part of the analysis will be detailed in Appendix I.

Discussion of second term. The second terms in (10) and (17) are dominant at high values of θ . Terms of this kind (albeit much simpler ones) are widely known in the noiseless setting [3, 4, 17, 29], and are based on the idea of that if a defective item is *masked* (i.e., every test it is in also contains at least one other defective), then even an optimal decoder will be unable to identify it. Intuitively, this is because the tests results are unchanged when that item is changed to be non-defective.

In the noisy setting, *complete* masking is no longer the dominant error event, but we can use a similar idea: Informally, an optimal decoder will fail when the following events hold simultaneously (see Corollary 1 in Appendix A for the formal version):

- (i) There exists a defective item j that has *relatively few tests* in which it is the only defective,

³A simpler approach based on Fano's inequality (e.g., see [33]) could also be used, but would only give the weaker statement $\mathbb{P}(\text{err}) \not\rightarrow 0$ rather than $\mathbb{P}(\text{err}) \rightarrow 1$.

and not too few of those tests are flipped by the noise.

- (ii) There exists a non-defective item j' appearing in *sufficiently many tests that don't contain any defective items*, and not too few of those tests are flipped by the noise.

The idea is that when both of these occur with suitable notions of “few” and “many”, the set $(S \setminus \{j\}) \cup \{j'\}$ will be favored to S itself. We proceed with further details towards making this intuition rigorous, and we will see that the formal version of conditions (i) and (ii) above lead to the terms f_1 and f_2 in the final results. This part of the analysis will be detailed in Appendix C (Bernoulli design) and Appendix E (near-constant weight design).

Maximum-likelihood decoding. Under a uniform prior on S , it is well known that the optimal decoding rule is *maximum-likelihood decoding*. Under our symmetric noise model, this is equivalent to choosing an estimate \hat{S} such that as many tests as possible match what the noiseless outcome would be. Thus, referring back to the above outline, we seek to find $j \in S$ and $j' \notin S$ such that $(S \setminus \{j\}) \cup \{j'\}$ has more such tests than S itself.

Identifying a suitable defective item. Suppose without loss of generality that $S = \{1, \dots, k\}$. We fix parameters $C > 0$ and $\zeta \in (0, 1)$ to be optimized at the end of our analysis (thus appearing as optimization parameters in our theorems), and consider the following “bad event”:⁴ *A given defective item has $\frac{Cnve^{-\nu}}{k}$ tests in which it is the only defective, and a fraction ζ of those tests are flipped.* We show that if n is below the stated threshold depending on f_1 , then some $j \in \{1, \dots, k\}$ satisfies this with high probability.

For the Bernoulli design, the number of tests where an item is the only defective follows a multinomial distribution, which is tricky to study directly. However, by restricting attention to $j \in \{1, \dots, k^\xi\}$ for ξ close to one, we are able to use a (rigorous) Poisson approximation [7] that is simpler to analyze due to having independence across the k^ξ items. For the near-constant design, some additional thought is needed, and we adopt an idea from [17], outlined as follows:

- Interpret the placements of items into tests as edges in a bipartite graph (or more precisely a multi-graph, since the same item can be placed in the same test twice).
- Establish that among the $k\Delta$ edges connecting defective items to tests, roughly $e^{-\nu}k\Delta$ of them are to tests containing exactly one defective item.
- Use a symmetry argument to show that for a given defective j , the number of tests in which it is the unique defective roughly follows a Hypergeometric distribution, $\text{Hg}(k\Delta, e^{-\nu}k\Delta, \Delta)$.

We need to take this idea further by considering the *conditional* distribution for defective item j given the placements of defective items $1, \dots, j-1$, but the conditioning turns out to have a minimal effect unless $j = \Theta(k)$. To avoid such cases, we use the same trick as the Bernoulli design, and only run up to $j = k^\xi$ for $\xi < 1$ very close to one.

We omit any further details here, but re-iterate that this part of the analysis leads to the f_1 term in the final bound.

⁴This parametrization of C turns out to be more notationally convenient.

Identifying a suitable non-defective item. For both designs, the test placements from one item to the next are independent, and also independent of the noise. As a result, we can condition on the defective placements and the noise, subject to the above “bad event” with parameters (C, ζ) occurring for some $j \in S$. Even after this conditioning, each non-defective item is placed into tests independently of the others, which is convenient for the analysis.

Consider any fixed $j' \in \{k+1, \dots, p\}$. When comparing the likelihoods of $(S \setminus \{j\}) \cup \{j'\}$ and S , there are only two types of tests that ultimately matter:

1. Tests including j but no item from $S \setminus \{j\}$;
2. Tests containing no item from S at all.

This is because all other tests contribute the same amount to the likelihood even after removing j and adding j' (due to the “OR” operation in (1)). We are interested in counting how many tests of the above kind j' is placed in, and moreover, how many of those tests are negative vs. positive (both are possible due to the noise). This comes down to the analysis of various binomial distributions, though it is complicated by delicate dependencies. We show that tests of the second kind above are dominant, whereas in this part of the analysis, the tests of the first kind above only contribute to lower-order asymptotic terms.

We omit any further details here, but re-iterate that this part of the analysis leads to the f_2 term in the final bound.

3 Outline of Achievability Proofs

In this section, we outline the proofs of our achievability results (i.e., upper bounds on the required number of tests), which come with several additional challenges compared to the converse results.

The analysis will be split into error events in which the final estimate \hat{S} has “low overlap” with S or “high overlap”. The main novelty is in the high overlap part, whereas for the low overlap part we will build on the information-theoretic framework of [39], which we now proceed to introduce.

3.1 Information-Theoretic Tools

Let $\mathbf{X} \in \{0, 1\}^{n \times p}$ denote the test matrix, $\mathbf{Y} \in \{0, 1\}^n$ the test results, and $S \subseteq [p]$ the defective set. Since both test designs are symmetric with respect to re-ordering items, we may focus on a specific choice of $S = s$ in our achievability analysis, say $s = [k] := \{1, \dots, k\}$. Let \mathbf{X}_s denote the resulting $n \times k$ sub-matrix of \mathbf{X} , and similarly when s is replaced by any other subset of $[p] := \{1, \dots, p\}$.

An error occurs when some $\bar{s} \neq s$ is favored by the decoder (to be defined below), and the contribution to the error probability can differ significantly depending on the amount of overlap between \bar{s} and s . For instance, there are only relatively few sets \bar{s} with $|s \setminus \bar{s}| = 1$, but the individual probability of favoring such \bar{s} is relatively high due to the low overlap. To capture this, we consider partitioning s into $(s_{\text{dif}}, s_{\text{eq}})$ with $s_{\text{dif}} \neq \emptyset$, and we denote $\ell = |s_{\text{dif}}| \in \{1, \dots, k\}$ (so

that $|s_{\text{eq}}| = k - \ell$). Intuitively, s_{eq} represents where an incorrect estimate overlaps with s , and s_{dif} represents the differing part.

For each such s_{dif} , we introduce the quantity

$$i^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}) := \log \frac{\mathbb{P}(\mathbf{Y} | \mathbf{X}_{s_{\text{dif}}}, \mathbf{X}_{s_{\text{eq}}})}{\mathbb{P}(\mathbf{Y} | \mathbf{X}_{s_{\text{eq}}})}, \quad (18)$$

which is the log-likelihood ratio of \mathbf{Y} given the full test sub-matrix \mathbf{X}_s vs. the smaller sub-matrix $\mathbf{X}_{s_{\text{eq}}}$ alone. Following the extensive literature on similar techniques for channel coding [25], we refer to this quantity as the *information density*. Notice that the average of (18) with respect to (\mathbf{X}, \mathbf{Y}) is the following conditional mutual information:

$$I_\ell^n := I(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}), \quad (19)$$

which depends on $(s_{\text{dif}}, s_{\text{eq}})$ only through $\ell := |s_{\text{dif}}|$ by the symmetry of the test designs.

The information-theoretic threshold decoder introduced in [39] can be described as follows: Fix the constants $\{\gamma_\ell\}_{\ell=1}^k$, and search for a set s of cardinality k such that

$$i^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}) \geq \gamma_{|s_{\text{dif}}|}, \quad \forall (s_{\text{dif}}, s_{\text{eq}}) \text{ such that } |s_{\text{dif}}| \neq 0. \quad (20)$$

If multiple such s exist, or if none exist, then an error is declared. This decoder is inspired by analogous thresholding techniques from the channel coding literature [23, 25], with the rough idea being that the numerator in (18) (i.e., the overall likelihood) tends to be much higher than the denominator for the correct set s , whereas for an incorrect s that has overlap s_{eq} with the correct one, such behavior is highly unlikely.

Limitation of existing approach. While the analysis of the decoder (20) in [39] leads to optimal thresholds in the noiseless setting, we found that even a sharpened analysis under their framework leads to a suboptimal result in the noisy setting. In more detail, their analysis leads to three terms: (i) a mutual information based term for $\ell = k$; (ii) a mutual information based term for $\ell = 1$; and (iii) a term capturing the concentration behavior of the information density. We found that a tight concentration analysis can lead to improvements in the third of these.⁵ However, there still exists a parameter (called δ_2 in [39]) that trades off the second and third terms and leaves the result significantly suboptimal, and we were unable to identify any promising route to avoiding this limitation when using the framework of [39].

3.2 A Hybrid Decoding Rule

In view of the above limitations, we introduce a hybrid decoding rule that allows us to use a novel maximum-likelihood analysis in the high-overlap regime, while still relying on the techniques of [39]

⁵Namely, in the denominator of the final result we get $\nu e^{-\nu}(1 - e^{-D(1/2\|\rho\|)})$ for the Bernoulli design, and $\nu^2 - \nu \log(e^\nu - 1 + e^{-D(1/2\|\rho\|)})$ for the near-constant weight design. We state these without proof because they would significantly lengthen the paper but still give a suboptimal final result.

in the low-overlap regime. The decoder is as follows: *Search for a set s of cardinality k such that both of the following are true:*

(i) *It holds that*

$$\mathbb{P}(\mathbf{Y}|\mathbf{X}_s) > \mathbb{P}(\mathbf{Y}|\mathbf{X}_{s'}), \quad \forall s' \text{ such that } 1 \leq |s \setminus s'| \leq \frac{k}{\log k}, \quad (21)$$

where we implicitly also constrain s' to have cardinality k .

(ii) *It holds (for suitably chosen $\{\gamma_\ell\}_{\frac{k}{\log k} < \ell \leq k}$) that*

$$i^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y}|\mathbf{X}_{s_{\text{eq}}}) \geq \gamma_{|s_{\text{dif}}|}, \quad \forall (s_{\text{dif}}, s_{\text{eq}}) \text{ such that } |s_{\text{dif}}| > \frac{k}{\log k}, \quad (22)$$

where we implicitly also constrain $(s_{\text{dif}}, s_{\text{eq}})$ to be a disjoint partition of s .

If no unique s exists satisfying both of these conditions, then an error is declared. Since we are using ℓ to represent the size of the set difference between the true defective set and an incorrect estimate (i.e., $\ell = |S \setminus \hat{S}| = |\hat{S} \setminus S|$), we will refer to the above cases as the low- ℓ (high overlap) and high- ℓ (low overlap) regimes respectively.

3.3 Analysis of the Low Overlap (High ℓ) Regime

For the Bernoulli design, the analysis of the threshold decoder (22) for $\ell > \frac{k}{\log k}$ will be taken directly from [39], and no change is needed. However, most of the analysis in [39] relied heavily on the test matrix having independent rows, so it is not applicable to the near-constant weight design. Thus, we need to analyze the information density and mutual information to fill in these gaps, with the main steps being as follows:

- We show that the mutual information I_ℓ^n has the same asymptotic behavior as that of the Bernoulli design;
- We establish a suitable concentration bound for the information density $i^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y}|\mathbf{X}_{s_{\text{eq}}})$.

Due to the lack of independence across tests, both of these require more effort than their counterpart for the Bernoulli design. On the other hand, for the concentration bound we are in the fortunate position of not needing precise constant factors, as those are only needed in the low- ℓ regime which we will handle using different methods.

This part of the analysis will be detailed in Appendix H (common analysis for both designs) and Appendix J (proofs of additional technical lemmas for the near-constant weight design).

3.4 Analysis of the High Overlap (Low ℓ) Regime

For the low- ℓ regime, we follow a similar structure to the converse bound, but we now need to consider all $\ell = 1, \dots, \frac{k}{\log k}$ instead of only $\ell = 1$. Specifically, we again consider (optimal) maximum-likelihood decoding, and accordingly, we seek to show that S is preferred to any other set of the

form $(S \setminus \mathcal{J}) \cup \mathcal{J}'$ with $\mathcal{J} \subset S$ and $\mathcal{J}' \subset \{1, \dots, p\} \setminus S$ both having cardinality $1 \leq \ell \leq \frac{k}{\log k}$. Recall that “being preferred” is equivalent to having more tests that match what the noiseless outcome would have been.

Generalizing the idea from the converse proof, we consider the event that a given defective subset $\mathcal{J} \subset S$ of size ℓ has $\frac{Cnve^{-\nu\ell}}{k}$ tests where at least one of its items is included but none from $S \setminus \mathcal{J}$ are included, and a fraction ζ of those tests are flipped. As well as generalizing beyond $\ell = 1$, we now also need to simultaneously consider *all* possible choices of (C, ζ) that can occur.

We let $k_{\ell, C, \zeta}$ denote the (random) number of size- ℓ subsets of S such that the above event holds with parameters (C, ζ) , and the first step is to rule out many (ℓ, C, ζ) triplets by showing that $k_{\ell, C, \zeta} = 0$ with high probability. This is done by directly studying $\mathbb{E}[k_{\ell, C, \zeta}]$ under the respective design, and then applying Markov’s inequality and a union bound. This part of the analysis leads to the term f_1 in the final bound.

For the (ℓ, C, ζ) triplets for which we cannot guarantee $k_{\ell, C, \zeta} = 0$, we need to consider the placements of non-defectives. We again follow a similar argument to the converse, but with more general $\ell \geq 1$ and all relevant (C, ζ) pairs being considered simultaneously. In particular, moving from $\ell = 1$ to $\ell > 1$ adds significant additional technical challenges, but conceptually we still follow similar steps as in the converse analysis. The term corresponding to $\ell = 1$ turns out to be the dominant one, and this part of the analysis leads to the term f_2 in the final bound.

This part of the analysis will be detailed in Appendix D (Bernoulli design) and Appendix F (near-constant weight design).

4 Conclusion

We have derived exact asymptotic thresholds on the number of tests required for probabilistic group testing under binary symmetry noise, under both the Bernoulli design and the near-constant weight design. We believe that our results bring the degree of understanding of this noisy group testing model significantly closer to the noiseless setting, but two major open problems remain:

- It is currently unclear whether the exact threshold for the near-constant weight design (with optimized ν) is also asymptotically optimal among *all* non-adaptive designs.
- The upper bounds were only derived using computationally intractable information-theoretic decoding rules, and establishing the same thresholds using polynomial-time algorithms (possibly with the test design slightly modified, e.g., via spatial coupling techniques [17]) would be of significant interest.

Noiseless counterparts for both of these points were established in [17], so their techniques may serve as the natural starting point for studying the same questions in the noisy setting.

Appendix

A Roadmap of the Proofs

In all four of our main results, the threshold consists of a maximum of two terms, the second of which itself consists of an optimization of parameters (C, ζ) . For both the achievability and converse results, these two terms are obtained using very different methods. Forward-references to the relevant subsequent appendices will be given below.

Converse results. For the converse analysis, the “max” operation comes from simply proving two different converse bounds and taking the stronger of the two. While we do not explicitly split the analysis into $\ell = 1, \dots, k$ for the converse, we still refer to the first term as the “high- ℓ ” analysis, and the second term as the “low- ℓ ” analysis. This is because:

- The starting point for the first term is a lower bound on the error probability from [39] that depends on ℓ but we specialize to $\ell = k$, which amounts to considering error events where the incorrect estimate is completely disjoint from the true defective set.
- For the second term, we consider error events under which an estimate of the form $\hat{S} = (S \setminus \{j\}) \cup \{j'\}$ is favored over S by the decoder, thus corresponding to $\ell = 1$ since we interpret $\ell = |S \setminus \hat{S}| = |\hat{S} \setminus S|$.

Achievability results. For the achievability results, we recall the hybrid decoder introduced in (21)–(22) in Section 3.2, which for a given candidate set s , does two different checks that correspond to considering high- ℓ and low- ℓ errors separately:

- The first check (corresponding to low ℓ) is simply whether s has a higher likelihood than all s' within distance $\frac{k}{\log k}$, so our analysis comes down to that of maximum-likelihood decoding.
- The second check (corresponding to high ℓ) is based on thresholding the information density (introduced in (18)), and we study its conditions for success by following the framework of [39] (but with significant gaps to fill in the case of the near-constant weight design).

We claim that in order for our decoder to succeed, the following conditions are sufficient:

1. The true defective set (say $s = \{1, \dots, k\}$ without loss of generality) satisfies (21)–(22);
2. For any other set \tilde{s} of cardinality k with $|s \setminus \tilde{s}| > \frac{k}{\log k}$, it holds that $\mathbb{P}^n(\mathbf{X}_{\tilde{s} \setminus s}; \mathbf{Y} | \mathbf{X}_{\tilde{s} \cap s}) < \gamma_\ell$, where $\ell = |\tilde{s} \setminus s| = |s \setminus \tilde{s}|$.

To see that these conditions are sufficient, first consider some incorrect estimate \tilde{s} with $|s \setminus \tilde{s}| \leq \frac{k}{\log k}$. Such \tilde{s} is ruled out by the low- ℓ part of our decoder (i.e., (21)), because the correct s is within distance $\frac{k}{\log k}$ and has a higher likelihood due to condition 1 above. On the other hand, for any incorrect estimate \tilde{s} with $|s \setminus \tilde{s}| > \frac{k}{\log k}$, condition 2 above immediately implies that the high- ℓ condition of

the decoder (i.e., (22)) is not satisfied for \tilde{s} . Thus, we conclude that the true defective set s satisfies (21)–(22) but none of the incorrect sets \tilde{s} do so, as desired.

In the subsequent appendices, we will write $\mathbb{P}(\text{err})$ to denote the failure probabilities associated with the above two sufficient conditions, where it will be clear from context whether we mean condition 1 or condition 2. We will show that attaining $\mathbb{P}(\text{err}) \rightarrow 0$ for condition 1 leads to the first term in each achievability result’s requirement on n , and attaining $\mathbb{P}(\text{err}) \rightarrow 0$ for condition 2 leads to the second term. The overall requirement on n is then the stricter of these two requirements.

A.1 Overview of the Appendices

We outline the structure of the remaining appendices as follows:

- In Appendix B, we present necessary and sufficient conditions for maximum-likelihood decoding to fail, which will be used in all four of our low- ℓ proofs.
- In Appendix C, we establish the low- ℓ converse result for the Bernoulli design.
- In Appendix D, we establish the low- ℓ achievability result for the Bernoulli design.
- In Appendix E, we establish the low- ℓ converse result for the near-constant weight design.
- In Appendix F, we establish the low- ℓ achievability result for the near-constant weight design.
- In Appendix G, we state and prove various technical lemmas that are used throughout Appendices E and F.
- In Appendix H, we establish the high- ℓ achievability result for both designs (with the Bernoulli design coming from [39] with only minor changes).
- In Appendix I, we establish the high- ℓ converse result for both designs (with the Bernoulli design coming directly from [39]).
- In Appendix J, we prove two technical lemmas used for the high- ℓ results under the near-constant weight design.

Before proceeding, we also state a useful observation regarding the scaling of the number of tests n , and some useful technical lemmas.

A.2 Note on the Scaling of n

In our achievability results we seek to show that the error probability approaches zero when $n \geq (1 + \eta)n^*$, and in the converse results we seek to show that any algorithm has error probability approaching one when $n \leq (1 - \eta)n^*$, where n^* is a threshold scaling as $\Theta(k \log p)$. Recall that throughout the entire paper, we consider the regime $k = \Theta(p^\theta)$ with $\theta \in (0, 1)$, which implies that $\log k = \Theta(\log p)$. This implies that $n^* = \Theta(k \log p) = \Theta(k \log k)$ in all of our results.

In view of these observations, we will assume throughout the analysis that n itself also scales as $\Theta(k \log p) = \Theta(k \log k)$. For the achievability part, this is justified by the fact that only $n = \Theta(k \log p)$ or $n = \omega(k \log p)$ are valid choices, and the former clearly provides a stronger result (because of requiring fewer tests). Similar, for the converse part, only $n = \Theta(k \log p)$ or $n = o(k \log p)$ are valid, and a converse for the former immediately implies a converse for the latter anyway (since the decoder could always choose to ignore some of the test results).

A.3 Useful Technical Lemmas

The following (anti)-concentration bounds for binomial random variables will be used frequently in our analysis. Recall the notation $D(a\|b) = a \log \frac{a}{b} + (1-a) \log \frac{1-a}{1-b}$ for binary KL divergence.

Lemma 1. ((Anti-)Concentration of Binomial Random Variables, e.g., [9, Sec. 4.7], [8]) *For $X \sim \text{Bin}(N, q)$, we have the following:*

- (Chernoff bound) *If $k \leq Nq$, then we have*

$$\mathbb{P}(X \leq k) \leq \exp\left(-N \cdot D\left(\frac{k}{N}\|q\right)\right), \quad (23)$$

which implies

$$\mathbb{P}(X \leq k) \leq \exp\left(-Nq\left(\frac{k}{Nq} \log \frac{k}{Nq} + \frac{k}{Nq} - 1\right)\right). \quad (24)$$

These bounds also remain true when replacing $\mathbb{P}(X \leq k)$ with $\mathbb{P}(X \geq k)$ for $k \geq Nq$.

- (Anti-concentration) *For any $k \in \{1, \dots, N-1\}$, we have*

$$\mathbb{P}(X = k) \geq \frac{1}{2\sqrt{2k(1 - \frac{k}{N})}} \exp\left(-N \cdot D\left(\frac{k}{N}\|q\right)\right), \quad (25)$$

Moreover, for $k \in \{0, 1, \dots, N\}$, we have

$$\mathbb{P}(X = k) \geq \frac{1}{\sqrt{2N}} \exp\left(-N \cdot D\left(\frac{k}{N}\|q\right)\right). \quad (26)$$

We also state the following related lemma regarding binomial coefficients.

Lemma 2. (Bounds on Binomial Coefficients, e.g., [9, Sec. 4.7]) *Given positive integers N and k with $k \in \{1, \dots, N-1\}$, we have*

$$\frac{\sqrt{\pi}}{2} \cdot \frac{\exp(NH_2(k/N))}{\sqrt{2\pi k(1 - k/N)}} \leq \binom{N}{k} \leq \frac{\exp(NH_2(k/N))}{\sqrt{2\pi k(1 - k/N)}}. \quad (27)$$

Moreover, for $k \in \{0, 1, \dots, N\}$, we have

$$\frac{\exp(NH_2(k/N))}{\sqrt{2N}} \leq \binom{N}{k} \leq \exp(NH_2(k/N)). \quad (28)$$

A.4 Table of Notation

The main recurring notation used throughout the appendices is summarized in Table 1.

B Necessary and Sufficient Conditions for Maximum-Likelihood Decoding

As noted in Section 2, the optimal decoding rule for minimizing the overall error probability (under a uniform prior on the defective set S) is maximum-likelihood decoding:

$$\hat{S} = \arg \max_{s' \in \mathcal{S}_k} \mathcal{L}(s'), \text{ where } \mathcal{L}(S) := \mathbb{P}(\mathbf{Y}|S, \mathbf{X}), \quad (29)$$

with \mathcal{S}_k being the set of all $\binom{p}{k}$ subsets of $[p]$ of size k . Our converse analysis will characterize conditions under which MLE has $P_e \rightarrow 1$, which implies the same for any decoder.

In our achievability analysis, in view of the two sufficient conditions identified in Appendix A, we will consider a *restricted* MLE decoder such that the argmax in (29) is restricted to sets within distance $\frac{k}{\log k}$ of the true one. Note that this does not imply that our decoder (introduced in Section 3.2) actually has knowledge of the true defective set; rather, restricted MLE is a hypothetical decoder whose success coincides with one of the sufficient conditions identified in Appendix A. Thus, when the true defective set is s , we are interested in the following:

$$\hat{S}' = \arg \max_{s' \in \mathcal{S}_k : |s \setminus s'| \leq \frac{k}{\log k}} \mathcal{L}(s'). \quad (30)$$

In fact, it will suffice to also consider this restricted problem in our converse analysis, observing that if $\hat{S}' \neq s$ in (30) then we clearly also have $\hat{S} \neq s$ in (29).

Under some defective set $S \subset [p]$ with $|S| = k$, we say a test is correct if its result is consistent with the noiseless setting, and for the given (\mathbf{X}, \mathbf{Y}) we define $N_{\mathbf{X}, \mathbf{Y}}(S)$ as the number of correct tests. Observe that

$$\mathcal{L}(S) = \rho^{n - N_{\mathbf{X}, \mathbf{Y}}(S)} (1 - \rho)^{N_{\mathbf{X}, \mathbf{Y}}(S)}, \quad (31)$$

so due to the fact that $\rho \in (0, \frac{1}{2})$, the MLE decoder finds \hat{S} with the most correct tests, i.e., $N_{\mathbf{X}, \mathbf{Y}}(\hat{S}) \geq N_{\mathbf{X}, \mathbf{Y}}(S)$ holds for any $S \in \mathcal{S}_k$ (or similarly for \hat{S}' in the restricted version).

We introduce some notation before proceeding. Let $\nu > 0$ be the parameter associated with the Bernoulli designs or near-constant weight designs. We say that the pair $(C, \zeta) \in [0, \infty) \times [0, 1]$ is *feasible* with respect to some given $\ell \in [1, \frac{k}{\log k}]$ if $\frac{C\nu e^{-\nu\ell}}{k}$ and $\frac{\zeta C\nu e^{-\nu\ell}}{k}$ are integers; when $C = 0$, we only view $(C, \zeta) = (0, 0)$ as feasible (rather than all $\zeta \in [0, 1]$). Given $\mathcal{J} \subset s$ with $|\mathcal{J}| = \ell$ for some specific $\ell \in [1, \frac{k}{\log k}]$, we define the following useful quantities:

- $\mathcal{M}_{\mathcal{J}} \subset [n]$ indexes the tests that include some defective from \mathcal{J} but no defective from $s \setminus \mathcal{J}$;
- $\mathcal{M}_{\mathcal{J}0}$ indexes the negative tests in $\mathcal{M}_{\mathcal{J}}$ (i.e., the tests in $\mathcal{M}_{\mathcal{J}}$ that are flipped by noise);

Table 1: List of recurring notation.

Introduced in the main text	
p, k, n	Number of items, defectives, and tests
ρ, ν, θ	Noise parameter, design parameter, sparsity parameter ($k = \Theta(p^\theta)$)
Δ	Number of placements in near-constant weight design ($\Delta = \frac{\nu n}{k}$)
S, s	Defective set (random variable, specific realization)
$\mathbf{X}, \mathbf{Y}, \mathbf{Z}$	Test matrix, test results, noise variables
$(s_{\text{dif}}, s_{\text{eq}})$	Partition of s into two disjoint sets
ℓ	Size of s_{dif} and/or size of $s \setminus s'$ for true s and estimated s'
C	Parameter associated with number of non-masked tests for defectives
ζ	Parameter associated with fraction of non-masked tests that are flipped
f_1	Function associated with some defective satisfying a “bad” event
f_2, g	Functions associated with some non-defective satisfying a “bad” event
d, d^*	Parameter associated with g , its optimal value
i^n, I_ℓ^n	Information density, conditional mutual information
j, \mathcal{J}	Generic defective item / size- ℓ set of defective items
j', \mathcal{J}'	Generic non-defective item / size- ℓ set of non-defective items
Introduced in the appendices	
\mathcal{L}	Likelihood function
$N_{\mathbf{X}, \mathbf{Y}}(S)$	Number of tests that are correct under S
$\mathcal{M}_{\mathcal{J}}$	Tests that contain defectives from \mathcal{J} but not other defectives
$\mathcal{M}_{\mathcal{J}0}, \mathcal{M}_{\mathcal{J}1}$	Negative (resp. positive) tests in $\mathcal{M}_{\mathcal{J}}$
\mathcal{N}_0	Tests that contain no defectives
$\mathcal{N}_{00}, \mathcal{N}_{01}$	Negative (resp. positive) tests in \mathcal{N}_0
$M_{(\cdot)}, N_{(\cdot)}$	Cardinalities of $\mathcal{M}_{(\cdot)}$ and $\mathcal{N}_{(\cdot)}$
$\mathcal{K}_{\ell, C, \zeta}, k_{\ell, C, \zeta}$	Set of size- ℓ subsets of s satisfying (C, ζ) conditions, its cardinality
$G_{\mathcal{J}, \mathcal{J}', 1}, G_{\mathcal{J}, \mathcal{J}', 2}$	Tests in $\mathcal{N}_{01} \cup \mathcal{M}_{\mathcal{J}1}$ (resp. $\mathcal{N}_{00} \cup \mathcal{M}_{\mathcal{J}0}$) containing an item from \mathcal{J}
$\mathcal{M}'_j, \mathcal{M}'_j, \mathcal{K}'_{C, \zeta}, k'_{C, \zeta}$	Variants that exclude multiple identical placements (near-const. design)
\mathbf{T}_s	Unordered multi-set indicating test placements (near-const. design)
$\mathcal{M}_{\mathcal{J}, \mathcal{J}'}, \mathcal{N}_{0, \mathcal{J}'}$	Tests in $\mathcal{M}_{\mathcal{J}}$ (resp. \mathcal{N}_0) containing an item from \mathcal{J}'
$\mathcal{A}_1, \mathcal{A}_2, \text{etc.}$	Used for high-probability events
A	Equal to $C(1 - 2\zeta)$, captures the dependence of g on (C, ζ)
\hat{f}_2	Variant of f_2 equaling 0 outside a certain range
ξ	Generic parameter taken close to 1
ψ_ℓ	Function representing concentration bound for i^n
α	Limiting value of $\frac{\ell}{k}$
n_1, n_2	Number of tests with (resp. without) an item from s_{eq}
$\mathbf{Y}_1, \mathbf{Y}_2$	Results of tests with (resp. without) an item from s_{eq}
M	Number of tests with an item from s_{dif} but not s_{eq}
V	Number of positive tests in \mathbf{Y}_2
\mathbf{P}	Probabilities implicitly conditioned on $\mathbf{X}_{s_{\text{eq}}}$

- $\mathcal{M}_{\mathcal{J}1}$ indexes the positive tests in $\mathcal{M}_{\mathcal{J}}$;
- We let their cardinalities be $M_{\mathcal{J}} := |\mathcal{M}_{\mathcal{J}}|$, $M_{\mathcal{J}0} := |\mathcal{M}_{\mathcal{J}0}|$ and $M_{\mathcal{J}1} := |\mathcal{M}_{\mathcal{J}1}|$.

Similarly,

- $\mathcal{N}_0 \subset [n]$ indexes the tests that include no defectives;
- \mathcal{N}_{00} indexes the negative tests in \mathcal{N}_0 ;
- \mathcal{N}_{01} indexes the positive tests in \mathcal{N}_0 (i.e., the tests in \mathcal{N}_0 that are flipped by noise);
- We denote their cardinalities by $|\mathcal{N}_0| = N_0$, $|\mathcal{N}_{00}| = N_{00}$ and $|\mathcal{N}_{01}| = N_{01}$.

In the following lemma, we provide deterministic conditions for the failure of the restricted MLE decoder.

Lemma 3. (Conditions for the Failure of MLE) *Let $s \subset [p]$ with cardinality k be the actual defective set, and consider the restricted MLE decoder in (30). Given $\ell \in [1, \frac{k}{\log k}]$, for a feasible pair $(C, \zeta) \in [0, \infty) \times [0, 1]$ such that $\frac{Cnve^{-\nu}\ell}{k}$ and $\frac{\zeta \cdot Cnve^{-\nu}\ell}{k}$ are integers, we define*

$$\mathcal{K}_{\ell, C, \zeta} = \left\{ \mathcal{J} \subset s, |\mathcal{J}| = \ell : M_{\mathcal{J}} = \frac{Cnve^{-\nu}\ell}{k}, M_{\mathcal{J}0} = \frac{\zeta \cdot Cnve^{-\nu}\ell}{k} \right\} \quad (32)$$

and let $k_{\ell, C, \zeta} := |\mathcal{K}_{\ell, C, \zeta}|$. Given $\mathcal{J} \in \mathcal{K}_{\ell, C, \zeta}$ and $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$, let $G_{\mathcal{J}, \mathcal{J}', 1}$ be the number of tests in $\mathcal{N}_{01} \cup \mathcal{M}_{\mathcal{J}1}$ that contain some item from \mathcal{J}' , and $G_{\mathcal{J}, \mathcal{J}', 2}$ be the number of tests in $\mathcal{N}_{00} \cup \mathcal{M}_{\mathcal{J}0}$ that contain some item from \mathcal{J}' . Then we have the following two statements:

(a) (Sufficient condition for failure) *If for some feasible pair $(C, \zeta) \in [0, \infty) \times [0, 1]$, $\mathcal{K}_{\ell, C, \zeta}$ is non-empty, and there exist some $\mathcal{J} \in \mathcal{K}_{\ell, C, \zeta}$ and some $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$ such that*

$$G_{\mathcal{J}, \mathcal{J}', 1} - G_{\mathcal{J}, \mathcal{J}', 2} > (1 - 2\zeta) \frac{Cnve^{-\nu}\ell}{k}, \quad (33)$$

then $\mathcal{L}(s') > \mathcal{L}(s)$ with $s' := (s \setminus \mathcal{J}) \cup \mathcal{J}'$, which implies the failure of the restricted MLE decoder (30).

(b) (Necessary condition for failure) *If the restricted MLE decoder (30) fails (i.e., it returns some $s' \neq s$ with $|s \setminus s'| \in [1, \frac{k}{\log k}]$), then for some feasible pair $(C, \zeta) \in [0, \infty) \times [0, 1]$ with respect to $\ell := |s \setminus \widehat{S}|$, $\mathcal{K}_{\ell, C, \zeta}$ is non-empty, and there exist $\mathcal{J} \in \mathcal{K}_{\ell, C, \zeta}$ and some $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$ such that*

$$G_{\mathcal{J}, \mathcal{J}', 1} - G_{\mathcal{J}, \mathcal{J}', 2} \geq (1 - 2\zeta) \frac{Cnve^{-\nu}\ell}{k}. \quad (34)$$

Proof. Recall that we let $N_{\mathbf{X}, \mathbf{Y}}(S)$ be the number of correct tests under the defective set s , which enters the likelihood via (31). We separately prove the statements (a) and (b) by analyzing the number of correct tests.

Proof of (a): For $\mathcal{J} \in \mathcal{K}_{\ell, C, \zeta}$ and $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$, we note that the length- k $s' := (s \setminus \mathcal{J}) \cup \mathcal{J}'$ satisfies $|s \setminus s'| = \ell$, and $\mathcal{L}(s') > \mathcal{L}(s)$ evidently implies the failure of MLE. Thus we only need to prove $\mathcal{L}(s') > \mathcal{L}(s)$, and it suffices to show $N_{\mathbf{X}, \mathbf{Y}}(s') > N_{\mathbf{X}, \mathbf{Y}}(s)$. To this end, we compare the number of correct tests under s and s' as follows:

- We start with the true defective set s . By removing \mathcal{J} from the defective set, the correct tests in $\mathcal{M}_{\mathcal{J}1}$ become incorrect, while the incorrect tests in $\mathcal{M}_{\mathcal{J}0}$ become correct. Thus, there is a loss of $|\mathcal{M}_{\mathcal{J}1}| - |\mathcal{M}_{\mathcal{J}0}| = (1 - \zeta)M_{\mathcal{J}} - \zeta M_{\mathcal{J}} = (1 - 2\zeta)M_{\mathcal{J}} = (1 - 2\zeta)\frac{Cnve^{-\nu\ell}}{k}$ correct tests.
- Next, we add \mathcal{J}' to the current defective set $s \setminus \mathcal{J}$ to arrive at $s' = (s \setminus \mathcal{J}) \cup \mathcal{J}'$. Then, the incorrect tests in $\mathcal{N}_{01} \cup \mathcal{M}_{\mathcal{J}1}$ become correct if they contain some item from \mathcal{J}' , while the correct tests in $\mathcal{N}_{00} \cup \mathcal{M}_{\mathcal{J}0}$ become incorrect if they contain some item from \mathcal{J}' . This leads to a gain of $G_{\mathcal{J}, \mathcal{J}', 1} - G_{\mathcal{J}, \mathcal{J}', 2}$ correct tests.

Therefore, by (33) we have

$$N_{\mathbf{X}, \mathbf{Y}}(s') - N_{\mathbf{X}, \mathbf{Y}}(s) = G_{\mathcal{J}, \mathcal{J}', 1} - G_{\mathcal{J}, \mathcal{J}', 2} - (1 - 2\zeta)\frac{Cnve^{-\nu\ell}}{k} > 0,$$

which leads to $\mathcal{L}(s') > \mathcal{L}(s)$ as claimed.

Proof of (b): If restricted MLE fails and returns a set s' with $|s \setminus s'| = \ell \in [1, \frac{k}{\log k}]$, then we note from the definition (30) that $\mathcal{L}(s') \geq \mathcal{L}(s)$, and thus $N_{\mathbf{X}, \mathbf{Y}}(s') \geq N_{\mathbf{X}, \mathbf{Y}}(s)$. Based on s' , we construct the feasible (C, ζ) for which the conclusion of (b) holds. Specifically, we let $\mathcal{J} = s \setminus s'$ with $|\mathcal{J}| = \ell$ and recall the definitions of $\mathcal{M}_{\mathcal{J}}, \mathcal{M}_{\mathcal{J}0}, \mathcal{M}_{\mathcal{J}1}$ and their cardinalities $M_{\mathcal{J}}, M_{\mathcal{J}0}, M_{\mathcal{J}1}$; then, we set $(C, \zeta) = (\frac{M_{\mathcal{J}}}{k^{-1}nve^{-\nu\ell}}, \frac{M_{\mathcal{J}0}}{M_{\mathcal{J}}})$ if $M_{\mathcal{J}} > 0$, or $(C, \zeta) = (0, 0)$ if $M_{\mathcal{J}} = 0$. By construction $(C, \zeta) \in [0, \infty) \times [0, 1]$ is feasible, and by (32) it is easy to see that $\mathcal{J} \in \mathcal{K}_{\ell, C, \zeta}$. Moreover, we let $\mathcal{J}' = s' \setminus s \subset [p] \setminus s$ satisfying $|\mathcal{J}'| = \ell$, and note that by the definition of $(\mathcal{J}, \mathcal{J}')$ we have $s' = (s \setminus \mathcal{J}) \cup \mathcal{J}'$. It remains to show that (34) holds for the $(\mathcal{J}, \mathcal{J}')$ we are considering. To this end, we compare the number of correct tests under s and s' exactly the same as the above two dot points in the proof of (a): Starting with the true defective set s , removing \mathcal{J} from the defective set leads to a loss of $(1 - 2\zeta)\frac{Cnve^{-\nu\ell}}{k}$ correct tests, while adding \mathcal{J}' to the current $s \setminus \mathcal{J}$ leads to a gain of $G_{\mathcal{J}, \mathcal{J}', 1} - G_{\mathcal{J}, \mathcal{J}', 2}$ correct tests. Overall, we obtain $N_{\mathbf{X}, \mathbf{Y}}(s') - N_{\mathbf{X}, \mathbf{Y}}(s) = G_{\mathcal{J}, \mathcal{J}', 1} - G_{\mathcal{J}, \mathcal{J}', 2} - (1 - 2\zeta)\frac{Cnve^{-\nu\ell}}{k}$, and thus $N_{\mathbf{X}, \mathbf{Y}}(s') \geq N_{\mathbf{X}, \mathbf{Y}}(s)$ yields (34) as claimed. \square

In the achievability analysis, we seek to establish the threshold for n above which the MLE fails with $o(1)$ probability. Since 3(b) provides a necessary condition for the failure of MLE, we can instead bound the probability of the events stated in Lemma 3(b).

Regarding the converse part, while we stated Lemma 3(a) for all $\ell \in [1, \frac{k}{\log k}]$ for consistency with part (b), it will turn out to be sufficient to restrict attention to $\ell = 1$ with a single feasible $(C, \zeta) \in (0, \infty) \times (0, 1)$. We thus specialize Lemma 3(a) to $\ell = 1$ with a simpler set of notation. Specifically, if $\mathcal{J} = \{j\}$ for some $j \in s$, then we will simply write $\mathcal{M}_{\mathcal{J}}, \mathcal{M}_{\mathcal{J}1}, \mathcal{M}_{\mathcal{J}0}$ respectively as $\mathcal{M}_j, \mathcal{M}_{j1}, \mathcal{M}_{j0}$, and their cardinalities are denoted by M_j, M_{j1}, M_{j0} , respectively. When further given $\mathcal{J}' = \{j'\}$

for some $j' \in [p] \setminus s$, we write $G_{\mathcal{J}, \mathcal{J}', 1}, G_{\mathcal{J}, \mathcal{J}', 2}, \mathcal{K}_{\ell, C, \zeta}, k_{\ell, C, \zeta}$ in Lemma 3 as $G_{j, j', 1}, G_{j, j', 2}, \mathcal{K}_{C, \zeta}, k_{C, \zeta}$, respectively. Then, Lemma 8(a) specializes to the following.

Corollary 1. (Simplified Sufficient Condition for the Failure of MLE) *Let $s \subset [p]$ with cardinality k be the actual defective set, and consider the MLE decoder and $\ell = 1$. For a feasible pair $(C, \zeta) \in (0, \infty) \times (0, 1)$ such that $\frac{Cnve^{-\nu}}{k}$ and $\frac{\zeta \cdot Cnve^{-\nu}}{k}$ are integers, we define*

$$\mathcal{K}_{C, \zeta} = \left\{ j \in s : M_j = \frac{Cnve^{-\nu}}{k}, M_{j0} = \frac{\zeta \cdot Cnve^{-\nu}}{k} \right\} \quad (35)$$

and let $k_{C, \zeta} := |\mathcal{K}_{C, \zeta}|$. Given $j \in \mathcal{K}_{C, \zeta}$ and $j' \in [p] \setminus s$, we let $G_{j, j', 1}$ be the number of tests in $\mathcal{N}_{01} \cup \mathcal{M}_{j1}$ that contain item j' , and $G_{j, j', 2}$ be the number of tests in $\mathcal{N}_{00} \cup \mathcal{M}_{j0}$ that contain item j' . Then if for some feasible $(C, \zeta) \in (0, \infty) \times (0, 1)$ we have that $\mathcal{K}_{C, \zeta}$ is non-empty, and that there exist some $j \in \mathcal{K}_{C, \zeta}$ and some $j' \in [p] \setminus s$ such that

$$G_{j, j', 1} - G_{j, j', 2} > (1 - 2\zeta) \frac{Cnve^{-\nu}}{k}, \quad (36)$$

then $\mathcal{L}(s') > \mathcal{L}(s)$ with $s' = (s \setminus \{j\}) \cup \{j'\}$, which implies the failure of MLE (29).

C Low- ℓ Converse Analysis for Bernoulli Designs

Let $\mathcal{S}_k = \{S' \subset [p] : |S'| = k\}$ and recall that we consider the prior $S \sim \text{Unif}(\mathcal{S}_k)$. As noted in Appendix B, it suffices to prove the converse bound for the optimal decoder, that is, maximum likelihood estimation (MLE) given in (29). By symmetry with respect to re-ordering items, we can simply assume that the actual defective set is $s = [k]$, i.e., items $1, \dots, k$ are defective, and items $k+1, \dots, p$ are non-defective.

Reduction to two conditions

For $j \in s$ and $j' \in [p] \setminus s$ we will use the notation $\mathcal{M}_j, \mathcal{M}_{j0}, \mathcal{M}_{j1}, \mathcal{N}_0, \mathcal{N}_{00}, \mathcal{N}_{01}$ and the corresponding cardinalities from Corollary 1. Note that these index sets and their cardinalities are deterministic given \mathbf{X}_s and \mathbf{Y} . We consider $\ell = 1$, and say that $(C, \zeta) \in (0, \infty) \times (0, 1)$ is *feasible* if $\frac{Cnve^{-\nu}}{k}$ and $\frac{\zeta \cdot Cnve^{-\nu}}{k}$ are integers. Given feasible (C, ζ) we will consider $\mathcal{K}_{C, \zeta} = \{j \in s : M_j = \frac{Cnve^{-\nu}}{k}, M_{j0} = \frac{\zeta \cdot Cnve^{-\nu}}{k}\}$ with cardinality $k_{C, \zeta} = |\mathcal{K}_{C, \zeta}|$ as in (35). For given $(j, j') \in \mathcal{K}_{C, \zeta} \times ([p] \setminus s)$, recall that $G_{j, j', 1}$ denotes the number of tests in $\mathcal{N}_{01} \cup \mathcal{M}_{j1}$ that contain the non-defective item j' , and $G_{j, j', 2}$ denotes number of tests in $\mathcal{N}_{00} \cup \mathcal{M}_{j0}$ that contain the non-defective item j' ; see Corollary 1.

We first invoke Corollary 1 to identify the sufficient condition for the failure of MLE. In particular, MLE fails if for some feasible $(C, \zeta) \in (0, \infty) \times (0, 1)$, the following two conditions **(C1)** and **(C2)** simultaneously hold:

- **(C1)** $k_{C, \zeta} \geq 1$ (i.e., $\mathcal{K}_{C, \zeta} \neq \emptyset$);

- **(C2)** There exist $j \in \mathcal{K}_{C,\zeta}$ and $k+1 \leq j' \leq p$ such that

$$G_{j,j',1} - G_{j,j',2} > (1-2\zeta) \frac{Cn\nu e^{-\nu}}{k}. \quad (37)$$

In the rest of the proof, we identify explicit conditions in the form of sufficient conditions on n for ensuring **(C1)** and **(C2)**. We proceed to study the two conditions under a fixed feasible pair (C, ζ) .

Condition for (C1) under the given (C, ζ)

We first derive a condition for ensuring $k_{C,\zeta} \geq 1$ with $1 - o(1)$ probability. Fix a constant $\xi < 1$ (to be chosen close to 1 later) and suppose k^ξ is an integer for convenience. For a given $j \in s$ and any test (say, with index i), we define the probability of test i containing only one defective j as

$$P_1 = \mathbb{P}(i \in \mathcal{M}_j) = \frac{\nu}{k} \left(1 - \frac{\nu}{k}\right)^{k-1} = \frac{\nu e^{-\nu}(1 + o(1))}{k}. \quad (38)$$

Due to the i.i.d. nature of the Bernoulli design, $\mathbf{M}_\xi := [M_1, M_2, \dots, M_{k^\xi}]$ is determined by a multinomial distribution with n trials in which each M_i has probability P_1 (and another variable, say M_0 , captures the remaining probability of $1 - k^\xi P_1$). This appears to be difficult to analyze directly; to overcome this, we let $\hat{\mathbf{M}}_\xi := [\hat{M}_1, \hat{M}_2, \dots, \hat{M}_{k^\xi}]$ have independent components following $\text{Poi}(nP_1)$, and use the following lemma (stated using generic notation).

Lemma 4. (Poisson Approximation of Multinomial, [7, Thm. 1]) *Consider a multinomial random vector $\mathbf{S} = [S_0, S_1, \dots, S_k] \sim \text{Mn}(N; [p_0, p_1, \dots, p_k])$ (for some $p_j \geq 0$ satisfying $\sum_{j=0}^k p_j = 1$), its sub-vector $\mathbf{S}' = [S_1, \dots, S_k]$, and a Poisson vector $\mathbf{T} = [T_1, T_2, \dots, T_k]$ with independent Poisson components $T_i \sim \text{Poi}(Np_j), j \in [k]$. If $\sum_{j=1}^k p_j > 0$, then as $N \rightarrow \infty$ we have*

$$\|\mathbf{S}' - \mathbf{T}\|_{\text{TV}} = \left(\sum_{j=1}^k p_j\right) \left(\frac{1}{\sqrt{2\pi e}} + O\left(\frac{1}{\sqrt{N \sum_{j=1}^k p_j}}\right)\right). \quad (39)$$

By Lemma 4 and the fact that $n, k \rightarrow \infty$ as $p \rightarrow \infty$, we have

$$\begin{aligned} d_{\text{TV}}(\mathbf{M}_\zeta, \hat{\mathbf{M}}_\zeta) &= k^\xi P_1 \left(\frac{1}{\sqrt{2\pi e}} + O\left(\frac{1}{\sqrt{nk^\xi P_1}}\right)\right) \\ &= O\left(\frac{1}{k^{1-\xi}} + \sqrt{\frac{1}{nk^{1-\xi}}}\right) = o(1). \end{aligned} \quad (40)$$

In addition, we have

$$k_{C,\zeta} = |\mathcal{K}_{C,\zeta}| \geq |\mathcal{K}_{C,\zeta} \cap [k^\xi]| = \sum_{j=1}^{k^\xi} \mathbb{1}\left(M_j = \frac{Cn\nu e^{-\nu}}{k}, M_{j0} = \frac{\zeta Cn\nu e^{-\nu}}{k}\right). \quad (41)$$

By the definition of $d_{\text{TV}}(\cdot, \cdot)$, we can further replace M_j by the more convenient independent Poisson

variables $\{\hat{M}_j : j \in [k^\xi]\}$ with $o(1)$ difference in probability:

$$\mathbb{P}(k_{C,\zeta} = 0) \leq \mathbb{P}\left(\sum_{j=1}^{k^\xi} \mathbb{1}\left(M_j = \frac{Cn\nu e^{-\nu}}{k}, M_{j0} = \frac{\zeta Cn\nu e^{-\nu}}{k}\right) = 0\right) \quad (42)$$

$$= \mathbb{P}\left(\sum_{j=1}^{k^\xi} \mathbb{1}\left(\hat{M}_j = \frac{Cn\nu e^{-\nu}}{k}, M_{j0} = \frac{\zeta Cn\nu e^{-\nu}}{k}\right) = 0\right) + o(1) \quad (43)$$

$$= \left[1 - \mathbb{P}\left(\hat{M}_j = \frac{Cn\nu e^{-\nu}}{k}, M_{j0} = \frac{\zeta Cn\nu e^{-\nu}}{k}\right)\right]^{k^\xi} + o(1) \quad (44)$$

$$= \left[1 - \mathbb{P}\left(\hat{M}_j = \frac{Cn\nu e^{-\nu}}{k}\right) \mathbb{P}\left(\text{Bin}\left(\frac{Cn\nu e^{-\nu}}{k}, \rho\right) = \frac{\zeta Cn\nu e^{-\nu}}{k}\right)\right]^{k^\xi} + o(1), \quad (45)$$

where (44) holds because the \hat{M}_j are independent, and the $\frac{Cn\nu e^{-\nu}}{k}$ tests in \mathcal{M}_j are disjoint for different $j \in [k^\xi]$ (by definition of \mathcal{M}_j), so the k^ξ summands in (43) are independent; then, (45) holds because \hat{M}_j is only related to the randomness of \mathbf{X}_s , while given $M_j = \frac{Cn\nu e^{-\nu}}{k}$, $M_{j0} \sim \text{Bin}(\frac{Cn\nu e^{-\nu}}{k}, \rho)$ is only related to the randomness of \mathbf{Z} . Since $\hat{M}_j \sim \text{Poi}(nP_1)$, we have

$$\mathbb{P}\left(\hat{M}_j = \frac{Cn\nu e^{-\nu}}{k}\right) = e^{-nP_1} \frac{(nP_1)^{\frac{Cn\nu e^{-\nu}}{k}}}{\left(\frac{Cn\nu e^{-\nu}}{k}\right)!} \quad (46)$$

$$= \frac{\exp\left(-\frac{n\nu e^{-\nu}(1+o(1))}{k}\right)}{(C(1+o(1)))^{\frac{Cn\nu e^{-\nu}}{k}}} \cdot \frac{\left(\frac{Cn\nu e^{-\nu}}{k}\right)^{\frac{Cn\nu e^{-\nu}}{k}}}{\left(\frac{Cn\nu e^{-\nu}}{k}\right)!} \quad (47)$$

$$\geq \exp\left(-\frac{n\nu e^{-\nu}}{k}(C \log C + 1 + o(1))\right) \cdot \exp\left(\frac{Cn\nu e^{-\nu}}{k}(1 + o(1))\right) \quad (48)$$

$$= \exp\left(-\frac{n\nu e^{-\nu}}{k}(C \log C - C + 1 + o(1))\right), \quad (49)$$

where we substitute (38) in (46) and use Stirling's approximation (specifically, we use $\frac{w^w}{w!} \geq \frac{e^{w-1}}{w} = e^{w(1+o(1))}$ for $w \rightarrow \infty$ in (48) with $w = \frac{Cn\nu e^{-\nu}}{k}$). Moreover, by anti-concentration ((26) in Lemma 1), we have

$$\mathbb{P}\left(\text{Bin}\left(\frac{Cn\nu e^{-\nu}}{k}, \rho\right) = \frac{\zeta Cn\nu e^{-\nu}}{k}\right) \geq \frac{1}{\sqrt{\frac{2Cn\nu e^{-\nu}}{k}}} \exp\left(-\frac{Cn\nu e^{-\nu}}{k} D(\zeta \parallel \rho)\right) \quad (50)$$

$$= \exp\left(-\frac{n\nu e^{-\nu}}{k}(C \cdot D(\zeta \parallel \rho) + o(1))\right), \quad (51)$$

where (51) holds because $\frac{n}{k} = \Theta(\log k) \rightarrow \infty$ allows us to incorporate the leading factor into the $o(1)$ term in the exponent; analogous simplifications will be used in the subsequent analysis. Combining

the bounds in (49) and (51), we obtain

$$\mathbb{P}\left(\hat{M}_j = \frac{Cn\nu e^{-\nu}}{k}\right) \mathbb{P}\left(\text{Bin}\left(\frac{Cn\nu e^{-\nu}}{k}, \rho\right) = \frac{\zeta Cn\nu e^{-\nu}}{k}\right) \quad (52)$$

$$\geq \exp\left(-\frac{n\nu e^{-\nu}}{k}[C \log C - C + C \cdot D(\zeta\|\rho) + 1 + o(1)]\right) := P_2. \quad (53)$$

Substituting this into (45), we obtain

$$\mathbb{P}(k_{C,\zeta} = 0) \leq (1 - P_2)^{k^\xi} + o(1) \quad (54)$$

$$\leq \exp(-k^\xi P_2) + o(1). \quad (55)$$

Thus, to guarantee $\mathbb{P}(k_{C,\zeta} = 0) = o(1)$, or equivalently $\mathbb{P}(k_{C,\zeta} > 0) = 1 - o(1)$, it suffices to have $k^\xi P_2 \rightarrow \infty$. Defining the shorthand

$$f_1(C, \zeta, \rho) := C \log C - C + C \cdot D(\zeta\|\rho) + 1, \quad (56)$$

and noting that $\eta_1 \log \frac{p}{k} \rightarrow \infty$ for any given $\eta_1 > 0$, the condition $k^\xi P_2 \rightarrow \infty$ can be ensured by

$$\xi \log k - \frac{n\nu e^{-\nu}}{k}(f_1(C, \zeta, \rho) + o(1)) \geq \eta_1 \log \frac{p}{k}. \quad (57)$$

Moreover, by $k = \Theta(p^\theta)$ we have $\log k \sim \frac{\theta}{1-\theta} \log \frac{p}{k}$. Substituting this into (57), letting ξ be arbitrarily close to 1 and rearranging, we find that **(C1)** holds with $1 - o(1)$ probability if

$$n \leq \frac{(1 - \eta_1) \frac{\theta}{1-\theta} k \log \frac{p}{k}}{\nu e^{-\nu}(f_1(C, \zeta, \rho) + o(1))} \quad (58)$$

holds for some $\eta_1 > 0$ possibly different from the one in (57).

Condition for (C2) under the given (C, ζ)

Given $k_{C,\zeta} \geq 1$ and fixing an arbitrary index $j \in \mathcal{K}_{C,\zeta}$, in this part we identify a sufficient condition for **(C2)**. Most of our analysis will be conditioned on $(\mathbf{X}_s, \mathbf{Y})$ and only based on the randomness of $\mathbf{X}_{[p]\setminus s}$, but we first deduce the high-probability behaviour of the quantities $N_{01} = |\mathcal{N}_{01}|$ and $N_{00} = |\mathcal{N}_{00}|$ depending on randomness of $(\mathbf{X}_s, \mathbf{Y})$. Specifically, \mathcal{N}_{01} (\mathcal{N}_{00} , resp.) consists of the positive (negative, resp.) tests containing no defective, and since the probability of a specific test containing no defective is given by $(1 - \frac{\nu}{k})^k = e^{-\nu}(1 + o(1))$, we have $N_{01} \sim \text{Bin}(n, \rho e^{-\nu}(1 + o(1)))$ and $N_{00} \sim \text{Bin}(n, (1 - \rho)e^{-\nu}(1 + o(1)))$. Then, by Hoeffding's inequality, we have that

$$\mathcal{A}_1 := \{N_{01} = \rho n e^{-\nu}(1 + o(1)), N_{00} = (1 - \rho)n e^{-\nu}(1 + o(1))\} \quad (59)$$

holds with $1 - o(1)$ probability. We will subsequently suppose that we are on the event \mathcal{A}_1 , which is justified by a simple union bound.

For the specific $j \in \mathcal{K}_{C,\zeta}$ we have $M_{j0} = \frac{\zeta \cdot Cnve^{-\nu}}{k} = o(n)$ and $M_{j1} = \frac{(1-\zeta) \cdot Cnve^{-\nu}}{k} = o(n)$, and thus they are negligible compared to N_{01}, N_{00} given by \mathcal{A}_1 in (59):

$$|\mathcal{N}_{01} \cup \mathcal{M}_{j1}| = N_{01} + M_{j1} = \rho ne^{-\nu}(1 + o(1)), \quad (60)$$

$$|\mathcal{N}_{00} \cup \mathcal{M}_{j0}| = N_{00} + M_{j0} = (1 - \rho)ne^{-\nu}(1 + o(1)). \quad (61)$$

Therefore, for a specific $k + 1 \leq j' \leq n$, conditioning on $(\mathbf{X}_s, \mathbf{Y})$, by the randomness of the j' -th column of \mathbf{X} (i.e., the placement of the j' -th item), we have

$$G_{j,j',1} \sim \text{Bin}\left(\rho ne^{-\nu}(1 + o(1)), \frac{\nu}{k}\right), \quad (62)$$

$$G_{j,j',2} \sim \text{Bin}\left((1 - \rho)ne^{-\nu}(1 + o(1)), \frac{\nu}{k}\right). \quad (63)$$

In order to establish a lower bound on the probability of (37), we introduce a parameter d (to be chosen later) satisfying

$$d > \max\left\{\frac{C(1 - 2\zeta)}{\rho}, 0\right\} \quad (64)$$

and consider the event $G_{j,j',1} = \frac{d\rho ve^{-\nu}n}{k}$ (assumed to be positive integer). We then use (25) from Lemma 1 to obtain

$$\mathbb{P}\left(G_{j,j',1} - G_{j,j',2} > (1 - 2\zeta)\frac{Cnve^{-\nu}}{k}\right) \quad (65)$$

$$\geq \mathbb{P}\left(G_{j,j',1} = \frac{d\rho ve^{-\nu}n}{k}\right) \mathbb{P}\left(G_{j,j',2} < \frac{nve^{-\nu}}{k}(d\rho - (1 - 2\zeta)C)\right) \quad (66)$$

$$\geq \Theta\left(\frac{1}{\sqrt{n/k}}\right) \exp\left(-\rho ne^{-\nu}(1 + o(1)) \cdot D\left(\frac{d(1 + o(1))\nu}{k} \parallel \frac{\nu}{k}\right)\right) \quad (67)$$

$$\times \Theta\left(\frac{1}{\sqrt{n/k}}\right) \exp\left(-(1 - \rho)ne^{-\nu}(1 + o(1)) \cdot D\left(\frac{(d\rho - (1 - 2\zeta)C)(1 + o(1))\nu}{1 - \rho} \parallel \frac{\nu}{k}\right)\right) \quad (68)$$

$$= \exp\left(-\frac{nve^{-\nu}}{k}[g(C, \zeta, d, \rho) + o(1)]\right), \quad (69)$$

where in (69) we define $g(C, \zeta, d, \rho)$ that depends on (C, ζ) only through $A := C(1 - 2\zeta)$ as follows:

$$g(C, \zeta, d, \rho) = \rho d \log d + (\rho d - A) \log\left(\frac{\rho d - A}{1 - \rho}\right) + 1 - 2\rho d + A, \quad (70)$$

and (69) itself follows from $D(\frac{a\nu}{k} \parallel \frac{\nu}{k}) = \frac{\nu}{k}(a \log a - a + 1 + o(1))$ and writing the pre-factors $\Theta((\frac{n}{k})^{-1/2})$ as $\exp(\frac{o(1)n}{k})$. Differentiating $g(C, \zeta, d, \rho)$ with respect to d , we obtain

$$\frac{\partial g}{\partial d} = \rho \log\left(\frac{d(\rho d - A)}{1 - \rho}\right), \quad (71)$$

which we observe is zero if and only if $\rho d^2 - Ad - (1 - \rho) = 0$. Thus, setting $\frac{\partial g}{\partial d} = 0$, we readily

obtain that

$$\min_{d > \max\{0, \frac{A}{\rho}\}} g(C, \zeta, d, \rho) = g(C, \zeta, d^*, \rho) := f_2(C, \zeta, \rho), \quad (72)$$

$$\text{where } d^* = \frac{A + \sqrt{A^2 + 4\rho(1 - \rho)}}{2\rho}. \quad (73)$$

Therefore, we let $d = d^* + o(1)$ (here the $o(1)$ term only serves to ensure that $\frac{d\rho\nu e^{-\nu}n}{k}$ is a positive integer) to get from (69) the strongest bound:

$$\mathbb{P}\left(G_{j,j',1} - G_{j,j',2} > (1 - 2\zeta)\frac{Cn\nu e^{-\nu}}{k}\right) \geq \exp\left(-\frac{n\nu e^{-\nu}}{k}(f_2(C, \zeta, \rho) + o(1))\right). \quad (74)$$

Recall that to ensure the failure of MLE, with a chosen $j \in \mathcal{K}_{C,\zeta}$ at hand, the condition (37) should hold for some $k + 1 \leq j' \leq p$. The placements of the different non-defective items are independent, and by implicitly conditioning on $(\mathbf{X}_s, \mathbf{Y})$ we have

$$\mathbb{P}\left(\exists k + 1 \leq j \leq p, \text{ s.t. } G_{j,j',1} - G_{j,j',2} > (1 - 2\zeta)\frac{Cn\nu e^{-\nu}}{k}\right) \quad (75)$$

$$= 1 - \mathbb{P}\left(\forall k + 1 \leq j \leq p, G_{j,j',1} - G_{j,j',2} \leq (1 - 2\zeta)\frac{Cn\nu e^{-\nu}}{k}\right) \quad (76)$$

$$= 1 - \left[\mathbb{P}\left(G_{j,j',1} - G_{j,j',2} \leq (1 - 2\zeta)\frac{Cn\nu e^{-\nu}}{k}\right)\right]^{p-k} \quad (77)$$

$$\geq 1 - \left[1 - \exp\left(-\frac{n\nu e^{-\nu}}{k}(f_2(C, \zeta, \rho) + o(1))\right)\right]^{p-k}. \quad (78)$$

Moreover, using $1 - a \leq e^{-a}$ we have

$$\left[1 - \exp\left(-\frac{n\nu e^{-\nu}}{k}(f_2(C, \zeta, \rho) + o(1))\right)\right]^{p-k} \leq \exp\left(-(p-k)\exp\left(-\frac{n\nu e^{-\nu}}{k}(f_2(C, \zeta, \rho) + o(1))\right)\right). \quad (79)$$

Thus, to ensure that (78) behaves as $1 - o(1)$, it suffices to have $\log(p-k) - \frac{n\nu e^{-\nu}}{k}(f_2(C, \zeta, \rho) + o(1)) \rightarrow \infty$. Substituting $\log(p-k) \sim \log p \sim \frac{1}{1-\theta} \log \frac{p}{k}$, we obtain that if

$$\frac{1}{1-\theta} \log \frac{p}{k} - \frac{n\nu e^{-\nu}}{k}(f_2(C, \zeta, \rho) + o(1)) \geq \eta_2 \log \frac{p}{k} \quad (80)$$

holds for some $\eta_2 > 0$, then **(C2)** holds with $1 - o(1)$ probability. Note that the above condition can also be written as

$$n \leq \frac{(1 - \eta_2)\frac{1}{1-\theta}k \log \frac{p}{k}}{\nu e^{-\nu}(f_2(C, \zeta, \rho) + o(1))} \quad (81)$$

for some $\eta_2 > 0$ possibly different from the one in (80).

Wrapping up

By merging η_1 , η_2 and $o(1)$ into a factor of $1 - \eta$, the statement “(58) and (81) simultaneously hold” can be written as

$$n \leq (1 - \eta) \frac{k \log \frac{p}{k}}{(1 - \theta)\nu e^{-\nu}} \frac{1}{\max\{\frac{1}{\theta}f_1(C, \zeta, \rho), f_2(C, \zeta, \rho)\}} \quad (82)$$

for some $\eta > 0$. Under this condition, the optimal decoder MLE fails with $1 - o(1)$ probability. Further optimizing over (C, ζ) to render the tightest converse bound, we arrive at the threshold

$$\frac{k \log \frac{p}{k}}{(1 - \theta)\nu e^{-\nu}} \frac{1}{\min_{C>0, \zeta \in (0,1)} \max\{\frac{1}{\theta}f_1(C, \zeta, \rho), f_2(C, \zeta, \rho)\}}, \quad (83)$$

where $f_1(C, \zeta, \rho)$ is given in (56), and $f_2(C, \zeta, \rho)$ is defined in (72) and (70). This establishes the second term in (4).

D Low- ℓ Achievability Analysis for Bernoulli Designs

Recall that in this part of the analysis, we restrict our attention to $\ell \in [1, \frac{k}{\log k}]$. Our strategy is to show that there is $o(1)$ probability of the restricted MLE decoder (30) failing. This is significantly more involved than the converse analysis because we need to consider the failure events associated with *all* $\ell \in [1, \frac{k}{\log k}]$ and the corresponding feasible (C, ζ) pairs, and finally bound the overall failure probability. Due to the symmetry of the design with respect to re-ordering items, it suffices to consider the fixed defective set $s = [k]$, meaning that items $1, 2, \dots, k$ are defective and items $k + 1, k + 2, \dots, p$ are non-defective.

Reduction to two conditions

For $\mathcal{J} \subset s$ with $|\mathcal{J}| = \ell$ we will use the notations $\mathcal{M}_{\mathcal{J}}, \mathcal{M}_{\mathcal{J}^c}, \mathcal{N}_0, \mathcal{N}_{00}, \mathcal{N}_{01}$ and their corresponding cardinalities (e.g., $M_{\mathcal{J}} = |\mathcal{M}_{\mathcal{J}}|$, $N_0 = |\mathcal{N}_0|$) from Lemma 3. Conditioned on \mathbf{X}_s and \mathbf{Y} , these index sets and their cardinalities are deterministic. Given $\ell \in [1, \frac{k}{\log k}]$, we say $(C, \zeta) \in [0, \infty) \times [0, 1]$ is *feasible* if $\frac{Cn\nu e^{-\nu}\ell}{k}$ and $\frac{\zeta \cdot Cn\nu e^{-\nu}\ell}{k}$ are integers, subject to the restriction that $(C, \zeta) = (0, 0)$ is the only feasible pair with $C = 0$. Given feasible (C, ζ) , we define

$$\mathcal{K}_{\ell, C, \zeta} = \left\{ \mathcal{J} \subset s, |\mathcal{J}| = \ell : M_{\mathcal{J}} = \frac{Cn\nu e^{-\nu}\ell}{k}, M_{\mathcal{J}^c} = \frac{\zeta \cdot Cn\nu e^{-\nu}\ell}{k} \right\}$$

as in (32). For given $(\mathcal{J}, \mathcal{J}') \in \mathcal{K}_{\ell, C, \zeta} \times ([p] \setminus s)$ with $|\mathcal{J}'| = \ell$, recall that $G_{\mathcal{J}, \mathcal{J}', 1}$ denotes the number of tests in $\mathcal{N}_{01} \cup \mathcal{M}_{\mathcal{J}^c}$ that contain some item from \mathcal{J}' , and $G_{\mathcal{J}, \mathcal{J}', 2}$ denotes the number of tests in $\mathcal{N}_{00} \cup \mathcal{M}_{\mathcal{J}}$ that contain some item from \mathcal{J}' . All of these definitions are consistent with Lemma 3.

To bound the probability that restricted MLE fails, we first invoke Lemma 3(b) to obtain a necessary condition for the failure of restricted MLE. In particular, Lemma 3(b) implies the following: If restricted MLE fails and returns some s' satisfying $|s \setminus s'| \in [1, \frac{k}{\log k}]$, then there exist

some feasible $(C, \zeta) \in [0, \infty) \times [0, 1]$ with respect to the specific $\ell := |s \setminus s'|$, such that **(C1)** and **(C2)** below simultaneously hold:

- **(C1)** $k_{\ell, C, \zeta} \geq 1$ (i.e., $\mathcal{K}_{\ell, C, \zeta} \neq \emptyset$);
- **(C2)** There exist $\mathcal{J} \in \mathcal{K}_{\ell, C, \zeta}$ and some $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$ such that

$$G_{\mathcal{J}, \mathcal{J}', 1} - G_{\mathcal{J}, \mathcal{J}', 2} \geq (1 - 2\zeta) \frac{Cn\nu e^{-\nu\ell}}{k}. \quad (84)$$

Therefore, it remains to identify the thresholds for n above which there is $o(1)$ probability of **(C1)** and **(C2)** simultaneously holding. In the following, we will consider a fixed $\ell \in [1, \frac{k}{\log k}]$, and we will later apply a union bound to cover all ℓ .

The case of $(C, \zeta) \in (C_0, \infty) \times [0, 1]$ under a given ℓ

In general, we can have integer-valued $\frac{Cn\nu e^{-\nu\ell}}{k}$ with arbitrarily large C . In this part, we first address the cases of $C > C_0$ for some sufficiently large absolute constant C_0 , so that we can simply concentrate on $C = O(1)$ later. For specific $\mathcal{J} \subset s$ with $|\mathcal{J}| = \ell$, we let the probability of a test belonging to $\mathcal{M}_{\mathcal{J}}$ be

$$P_1 := \left(1 - \left(1 - \frac{\nu}{k}\right)^\ell\right) \left(1 - \frac{\nu}{k}\right)^{k-\ell} = \frac{\ell\nu e^{-\nu}(1 + o(1))}{k}, \quad (85)$$

where the last equality follows from $\ell = o(k)$. We thus have $M_{\mathcal{J}} \sim \text{Bin}(n, P_1)$, and under the given ℓ we can bound the probability

$$\mathbb{P}\left(\textbf{(C1) and (C2) hold for some feasible } (C, \zeta) \in (C_0, \infty) \times [0, 1]\right) \quad (86)$$

$$\leq \mathbb{P}\left(\mathcal{K}_{\ell, C, \zeta} \neq \emptyset \text{ holds for some feasible } (C, \zeta) \in (C_0, \infty) \times [0, 1]\right) \quad (87)$$

$$\leq \mathbb{P}\left(\exists \mathcal{J} \subset s \text{ with } |\mathcal{J}| = \ell, \text{ s.t. } M_{\mathcal{J}} > \frac{C_0 n \nu e^{-\nu\ell}}{k}\right) \quad (88)$$

$$\leq \binom{k}{\ell} \mathbb{P}\left(\text{Bin}\left(n, \frac{\ell\nu e^{-\nu}(1 + o(1))}{k}\right) > \frac{C_0 n \nu e^{-\nu\ell}}{k}\right) \quad (89)$$

$$\leq \exp\left((1 + o(1))\ell \log k - \frac{\ell n}{k} \nu e^{-\nu}(C_0 \log C_0 - C_0 + 1 + o(1))\right) \quad (90)$$

$$\leq k^{-5}, \quad (91)$$

where in (87) we get an upper bound by only considering **(C1)** and not **(C2)**, in (88) we further relax $\mathcal{K}_{\ell, C, \zeta} \neq \emptyset$ to the first condition of $\mathcal{K}_{\ell, C, \zeta}$ (namely $M_{\mathcal{J}} = \frac{Cn\nu e^{-\nu\ell}}{k}$; see (32)), in (89) we apply a union bound over all $\mathcal{J} \subset s$ with $|\mathcal{J}| = \ell$, in (90) we use the Chernoff bound (see (24) in Lemma 1), and (91) holds by letting $C_0 = \Theta(1)$ be sufficiently large (recall that we have $\frac{n}{k} = \Theta(\log k)$). Therefore, for a given $\ell \in [1, \frac{k}{\log k}]$, **(C1)** and **(C2)** simultaneously hold for some feasible $(C, \zeta) \in (C_0, \infty) \times [0, 1]$ with probability no more than k^{-5} .

Condition for (C1) under a given (ℓ, C, ζ)

In this part, we identify the more explicit condition for **(C1)** under the given ℓ and a specific feasible (C, ζ) with $C \leq C_0$. We seek to bound $k_{\ell, C, \zeta}$. Our strategy is to first calculate $\mathbb{E}k_{\ell, C, \zeta}$ and then deduce the high-probability behaviour of $k_{\ell, C, \zeta}$ via Markov's inequality. Since $k_{\ell, C, \zeta} = \sum_{\mathcal{J} \subset s, |\mathcal{J}|=\ell} \mathbb{1}(M_{\mathcal{J}} = \frac{Cn\nu e^{-\nu}\ell}{k}, M_{\mathcal{J}^c} = \zeta M_{\mathcal{J}})$, we have

$$\mathbb{E}k_{\ell, C, \zeta} = \binom{k}{\ell} \mathbb{P} \left(\text{for fixed } \mathcal{J} \subset s \text{ with } |\mathcal{J}| = \ell, M_{\mathcal{J}} = \frac{Cn\nu e^{-\nu}\ell}{k}, M_{\mathcal{J}^c} = \zeta M_{\mathcal{J}} \right) \quad (92)$$

$$= \binom{k}{\ell} \mathbb{P} \left(\text{Bin}(n, P_1) = \frac{Cn\nu e^{-\nu}\ell}{k} \right) \mathbb{P} \left(\text{Bin}\left(\frac{Cn\nu e^{-\nu}\ell}{k}, \rho\right) = \frac{\zeta \cdot Cn\nu e^{-\nu}\ell}{k} \right) \quad (93)$$

$$\leq \binom{k}{\ell} \exp \left(-n \cdot D\left(\frac{C\nu e^{-\nu}\ell}{k} \parallel \frac{\nu e^{-\nu}\ell(1+o(1))}{k}\right) - \frac{Cn\nu e^{-\nu}\ell}{k} D(\zeta \parallel \rho) \right) \quad (94)$$

$$= \binom{k}{\ell} \exp \left(-\frac{n\nu e^{-\nu}\ell}{k} \left[f_1(C, \zeta, \rho) + o(1) \right] \right), \quad (95)$$

where in (94) we use the Chernoff bound (see (23) in Lemma 1), and in (95) we define

$$f_1(C, \zeta, \rho) := C \log C - C + C \cdot D(\zeta \parallel \rho) + 1, \quad (96)$$

and then use $D\left(\frac{C\nu e^{-\nu}\ell}{k} \parallel \frac{\nu e^{-\nu}\ell(1+o(1))}{k}\right) = \frac{\nu e^{-\nu}\ell}{k} (C \log C - C + 1 + o(1))$. We note that $f_1(C, \zeta, \rho)$ coincides with (56) in the proof of the converse bound for Bernoulli designs.

We now apply Markov's inequality to bound $k_{\ell, C, \zeta}$. We divide this step into two cases, namely $1 \leq \ell \leq \log k$ and $\log k < \ell \leq \frac{k}{\log k}$.

The case of $1 \leq \ell \leq \log k$: For specific (ℓ, C, ζ) , Markov's inequality gives

$$\mathbb{P} \left(k_{\ell, C, \zeta} \geq (\ell \log k)^5 \mathbb{E}[k_{\ell, C, \zeta}] \right) \leq (\ell \log k)^{-5}. \quad (97)$$

Using the bound on $\mathbb{E}k_{\ell, C, \zeta}$ from (95), we know that with probability at least $1 - (\ell \log k)^{-5}$, it holds that

$$k_{\ell, C, \zeta} < (\ell \log k)^5 \mathbb{E}k_{\ell, C, \zeta} \leq (\ell \log k)^5 \binom{k}{\ell} \exp \left(-\frac{n\nu e^{-\nu}\ell}{k} \left[f_1(C, \zeta, \rho) + o(1) \right] \right) \quad (98)$$

$$\leq \exp \left((1 + o(1)) \ell \log \frac{k}{\ell} - \frac{n\nu e^{-\nu}\ell}{k} \left[f_1(C, \zeta, \rho) + o(1) \right] \right). \quad (99)$$

The case of $\log k < \ell \leq \frac{k}{\log k}$: For specific (ℓ, C, ζ) , Markov's inequality gives

$$\mathbb{P} \left(k_{\ell, C, \zeta} \geq k^5 \mathbb{E}[k_{\ell, C, \zeta}] \right) \leq k^{-5}. \quad (100)$$

Combining with (95), we know that with probability at least $1 - k^{-5}$, it holds that

$$k_{\ell, C, \zeta} < k^5 \mathbb{E} k_{\ell, C, \zeta} \leq k^5 \binom{k}{\ell} \exp \left(- \frac{n\nu e^{-\nu} \ell}{k} \left[f_1(C, \zeta, \rho) + o(1) \right] \right) \quad (101)$$

$$\leq \exp \left((1 + o(1)) \ell \log \frac{k}{\ell} - \frac{n\nu e^{-\nu} \ell}{k} \left[f_1(C, \zeta, \rho) + o(1) \right] \right). \quad (102)$$

Combining the above two cases, we obtain the following conclusion for given (ℓ, C, ζ) :

$$\frac{n\nu e^{-\nu} \ell}{k} \left[f_1(C, \zeta, \rho) + o(1) \right] \geq (1 + \eta_1) \ell \log \frac{k}{\ell} \text{ for some } \eta_1 > 0 \quad (103)$$

$$\implies \mathbb{P}(k_{\ell, C, \zeta} \geq 1) \leq \hat{P}_\ell, \text{ where } \hat{P}_\ell = \begin{cases} (\ell \log k)^{-5}, & \text{if } 1 \leq \ell \leq \log k \\ k^{-5}, & \text{if } \log k < \ell \leq \frac{k}{\log k} \end{cases}, \quad (104)$$

where we implicitly assume that k is sufficiently large. Therefore, if (103) holds, then **(C1)** holds for some given (ℓ, C, ζ) with probability at most \hat{P}_ℓ .

Condition for **(C2)** under given (ℓ, C, ζ)

We again consider a given ℓ and a specific feasible pair of (C, ζ) with $C \leq C_0$, and now switch to analyzing **(C2)**. For $\mathcal{J}' \subset [p]$ with $|\mathcal{J}'| = \ell$, the probability of a test containing some item from \mathcal{J}' is:

$$P_2 := 1 - \left(1 - \frac{\nu}{k}\right)^\ell = \frac{\nu \ell (1 + o(1))}{k}. \quad (105)$$

Given (ℓ, C, ζ) , **(C2)** states that (84) holds for some $\mathcal{J} \in \mathcal{K}_{\ell, C, \zeta}$ and some $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$. In (84), recall that $G_{\mathcal{J}, \mathcal{J}', 1}$ ($G_{\mathcal{J}, \mathcal{J}', 2}$, resp.) is the number of tests in $\mathcal{N}_{01} \cup \mathcal{M}_{\mathcal{J}1}$ ($\mathcal{N}_{00} \cup \mathcal{M}_{\mathcal{J}0}$, resp.) that contain some item from \mathcal{J}' , and it will be convenient to further decompose these into

$$G_{\mathcal{J}, \mathcal{J}', 1} = \tilde{G}_{\mathcal{J}', 1} + U_{\mathcal{J}, \mathcal{J}', 1}, \quad G_{\mathcal{J}, \mathcal{J}', 2} = \tilde{G}_{\mathcal{J}', 2} + U_{\mathcal{J}, \mathcal{J}', 2}, \quad (106)$$

where we define $\tilde{G}_{\mathcal{J}', 1}$ ($\tilde{G}_{\mathcal{J}', 2}$, resp.) as the number of tests in \mathcal{N}_{01} (\mathcal{N}_{00} , resp.) that contain some item from \mathcal{J}' (thus $\tilde{G}_{\mathcal{J}', 1}$ and $\tilde{G}_{\mathcal{J}', 2}$ have no dependence on \mathcal{J} , as also reflected by our notation), and $U_{\mathcal{J}, \mathcal{J}', 1}$ ($U_{\mathcal{J}, \mathcal{J}', 2}$, resp.) as the number of tests in $\mathcal{M}_{\mathcal{J}1}$ ($\mathcal{M}_{\mathcal{J}0}$, resp.) that contain some item from \mathcal{J}' . Given $(\mathbf{X}_s, \mathbf{Y})$ and for specific \mathcal{J} and \mathcal{J}' , it follows that

$$\tilde{G}_{\mathcal{J}', 1} \sim \text{Bin}(N_{01}, P_2), \quad \tilde{G}_{\mathcal{J}', 2} \sim \text{Bin}(N_{00}, P_2) \quad (107)$$

$$U_{\mathcal{J}, \mathcal{J}', 1} \sim \text{Bin}(M_{\mathcal{J}1}, P_2), \quad U_{\mathcal{J}, \mathcal{J}', 2} \sim \text{Bin}(M_{\mathcal{J}0}, P_2) \quad (108)$$

are independent, as the tests comprising $\mathcal{N}_{01}, \mathcal{N}_{00}, \mathcal{M}_{\mathcal{J}0}, \mathcal{M}_{\mathcal{J}1}$ are disjoint. Under these definitions (84) can be written as

$$\underbrace{\tilde{G}_{\mathcal{J}', 1} - \tilde{G}_{\mathcal{J}', 2}}_{:= \tilde{G}_{\mathcal{J}'}} + U_{\mathcal{J}, \mathcal{J}', 1} - U_{\mathcal{J}, \mathcal{J}', 2} \geq (1 - 2\zeta) \frac{C n \nu e^{-\nu} \ell}{k}, \quad (109)$$

which further implies $\tilde{G}_{\mathcal{J}'} \geq (1 - 2\zeta) \frac{Cn\nu e^{-\nu}\ell}{k} - U_{\mathcal{J},\mathcal{J}',1}$ since $U_{\mathcal{J},\mathcal{J}',2} \geq 0$. Thus, if **(C2)** holds, then we have $\tilde{G}_{\mathcal{J}'} \geq (1 - 2\zeta) \frac{Cn\nu e^{-\nu}\ell}{k} - U_{\mathcal{J},\mathcal{J}',1}$ for some $\mathcal{J} \in \mathcal{K}_{\ell,C,\zeta}$ and some $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$, which further implies

$$\max_{\substack{\mathcal{J}' \subset [p] \setminus s \\ |\mathcal{J}'| = \ell}} \tilde{G}_{\mathcal{J}'} \geq (1 - 2\zeta) \frac{Cn\nu e^{-\nu}\ell}{k} - \max_{\substack{\mathcal{J} \in \mathcal{K}_{\ell,C,\zeta} \\ \mathcal{J}' \subset [p] \setminus s, |\mathcal{J}'| = \ell}} U_{\mathcal{J},\mathcal{J}',1}. \quad (110)$$

Thus, to bound the probability of **(C2)**, it suffices to bound the probability of its necessary condition given in (110).

The effect of $U_{\mathcal{J},\mathcal{J}',1}$: For convenience, we write the last term in (110) using the shorthand $\max_{\mathcal{J},\mathcal{J}'} U_{\mathcal{J},\mathcal{J}',1}$ with the constraints left implicit. We first show that the term $\max_{\mathcal{J},\mathcal{J}'} U_{\mathcal{J},\mathcal{J}',1}$ in the right-hand side of (110) only has a minimal impact, in the sense that scales as $o(\frac{\ell n}{k}) = o(\ell \log k)$. In fact, by $U_{\mathcal{J},\mathcal{J}',1} \sim \text{Bin}(M_{\mathcal{J}1}, P_2)$ and $M_{\mathcal{J}1} = \frac{(1-\zeta)Cn\nu e^{-\nu}\ell}{k} \leq \frac{C_0 n \ell}{k}$ (recall that we are now considering $C \leq C_0 = O(1)$ and $\nu e^{-\nu} \leq 1$ holds trivially), for fixed $\mathcal{J}, \mathcal{J}'$ we have

$$\mathbb{P} \left(U_{\mathcal{J},\mathcal{J}',1} \geq \frac{\ell n}{k(\log \frac{k}{\ell})^{1/2}} \right) \quad (111)$$

$$\leq \mathbb{P} \left(\text{Bin} \left(\frac{C_0 n \ell}{k}, \frac{\nu \ell (1 + o(1))}{k} \right) \geq \frac{\ell n}{k(\log \frac{k}{\ell})^{1/2}} \right) \quad (112)$$

$$\leq \exp \left(-\frac{\ell n}{k} \left(\left[\log \frac{k}{\ell} \right]^{-1/2} \left[\log \frac{k}{C_0 \nu \ell (\log \frac{k}{\ell})^{1/2}} + 1 \right] - \frac{C_0 \nu \ell}{k} \right) \right) \quad (113)$$

$$= \exp \left(-\Omega \left(\sqrt{\log \frac{k}{\ell}} \right) \cdot \ell \log k \right), \quad (114)$$

where we use Chernoff bound (see (24) in Lemma 1) in (113), and (114) holds since $\frac{\ell n}{k} = \Theta(\ell \log k)$. By a union bound over at most $\binom{k}{\ell} \times \binom{p}{\ell}$ choices of $(\mathcal{J}, \mathcal{J}')$, we obtain that

$$\mathbb{P} \left(\max_{\mathcal{J},\mathcal{J}'} U_{\mathcal{J},\mathcal{J}',1} \geq \frac{\ell n}{k(\log \frac{k}{\ell})^{1/2}} \right) \leq \exp \left(\ell \log \frac{ek}{\ell} + \ell \log \frac{ep}{\ell} - \Omega \left(\sqrt{\log \frac{k}{\ell}} \right) \cdot \ell \log k \right) \leq k^{-10\ell}, \quad (115)$$

where the last inequality holds for large enough k because $\frac{k}{\ell} \rightarrow \infty$ and $k = \Theta(p^\theta)$ for some $\theta \in (0, 1)$. Combining with $\frac{\ell n}{k(\log \frac{k}{\ell})^{1/2}} = o(\frac{\ell n}{k})$, with probability at least $1 - k^{-10\ell}$ we have

$$\max_{\mathcal{J},\mathcal{J}'} U_{\mathcal{J},\mathcal{J}',1} = o\left(\frac{\ell n}{k}\right), \quad (116)$$

under which the sufficient condition for **(C2)** given in (110) can be further simplified to

$$\max_{\substack{\mathcal{J}' \subset [p] \setminus s \\ |\mathcal{J}'| = \ell}} \tilde{G}_{\mathcal{J}'} \geq (1 - 2\zeta - o(1)) \frac{Cn\nu e^{-\nu}\ell}{k}. \quad (117)$$

Decomposing the event (117): With respect to the randomness of \mathbf{X}_s and \mathbf{Y} , we have already shown that (59) holds with $1 - o(1)$ probability, and thus we can proceed on the events

$$\mathcal{A}_1 = \{N_{01} = \rho n e^{-\nu}(1 + o(1)), N_{00} = (1 - \rho)n e^{-\nu}(1 + o(1))\} \quad (118)$$

by a simple union bound. In the following, for a generic event \mathcal{E} , an upper bound on $\mathbb{P}(\mathcal{E})$ should be more precisely understood as a bound on $\mathbb{P}(\mathcal{E} \cap \mathcal{A}_1)$, but to avoid cumbersome notation we will only make this explicit in the final stage (see (147) below).

Given $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$, by conditioning on $(\mathbf{X}_s, \mathbf{Y})$, the randomness of $\mathbf{X}_{\mathcal{J}'}$ gives

$$\tilde{G}_{\mathcal{J}'} = \tilde{G}_{\mathcal{J}',1} - \tilde{G}_{\mathcal{J}',2} \sim \text{Bin}(\rho n e^{-\nu}(1 + o(1)), P_2) - \text{Bin}((1 - \rho)n e^{-\nu}(1 + o(1)), P_2), \quad (119)$$

where P_2 is the probability defined in (105). We can now introduce

$$d_0 := \frac{C(1 - 2\zeta) + 1 - \rho}{\rho} \quad (120)$$

and proceed as follows:

$$\mathbb{P}\left(\tilde{G}_{\mathcal{J}'} \geq (1 - 2\zeta - o(1)) \frac{C n \nu e^{-\nu} \ell}{k}\right) \leq \mathbb{P}\left(\tilde{G}_{\mathcal{J}',1} \geq \frac{d_0 \cdot \rho \nu \ell n e^{-\nu}}{k}\right) \quad (121)$$

$$+ \sum_{\substack{0 \leq d < d_0 \\ \frac{d \rho \nu e^{-\nu} \ell n}{k} \in \mathbb{Z}}} \mathbb{P}\left(\tilde{G}_{\mathcal{J}',1} = \frac{d \cdot \rho \nu \ell n e^{-\nu}}{k}\right) \mathbb{P}\left(\tilde{G}_{\mathcal{J}',2} \leq [\rho d - (1 - 2\zeta - o(1))C] \frac{n \nu e^{-\nu} \ell}{k}\right). \quad (122)$$

To bound the terms in (121) and (122), we consider the following two cases.

Case 1: $(1 - 2\zeta)C > 2\rho - 1$. In this case we have $d_0 > 1$.

Bounding the term in (121): Consider conditioning on $(\mathbf{X}_s, \mathbf{Y})$ and utilizing the randomness of $\mathbf{X}_{\mathcal{J}'}$, and recall from (119) and (105) that we have $\tilde{G}_{\mathcal{J}',1} \sim \text{Bin}(\rho n e^{-\nu}(1 + o(1)), \frac{\nu \ell (1 + o(1))}{k})$. Since $d_0 > 1$, we can apply Chernoff bound ((23) in Lemma 1) to bound the term in (121) as

$$\mathbb{P}\left(\tilde{G}_{\mathcal{J}',1} \geq \frac{d_0 \cdot \rho \nu \ell n e^{-\nu}}{k}\right) \leq \exp\left(-\rho n e^{-\nu}(1 + o(1)) D\left(\frac{d_0 \ell \nu}{k} \parallel \frac{\ell \nu}{k}\right)\right). \quad (123)$$

Bounding the term in (122): First observe that $\mathbb{P}(\tilde{G}_{\mathcal{J}',2} \leq [\rho d - (1 - 2\zeta - o(1))C] \frac{n \nu e^{-\nu} \ell}{k}) = 0$ if $d < \frac{C(1 - 2\zeta)}{\rho} - o(1)$, and so we can further restrict the summation over d in (122) to the range

$$\mathcal{D} := \left\{ \max\left\{0, \frac{C(1 - 2\zeta)}{\rho}\right\} \leq d < d_0 : \frac{d \rho \nu e^{-\nu} \ell n}{k} \in \mathbb{Z} \right\} \quad (124)$$

up to a $o(1)$ term that has no impact on our subsequent analysis. Since $d_0 = O(1)$, the term in (122) involves no more than $O(\frac{\ell n}{k})$ summands, so using the Chernoff bound (see (23) in Lemma 1)

we can bound it as

$$O\left(\frac{\ell n}{k}\right) \max_{d \in \mathcal{D}} \left\{ \exp\left(-\rho n e^{-\nu}(1+o(1)) \cdot D\left(\frac{d\nu\ell(1+o(1))}{k} \parallel \frac{\nu\ell(1+o(1))}{k}\right)\right) \right. \quad (125)$$

$$\left. \cdot \exp\left(- (1-\rho)n e^{-\nu}(1+o(1)) \cdot D\left(\frac{[\rho d - (1-2\zeta)C]\nu\ell(1+o(1))}{(1-\rho)k} \parallel \frac{\nu\ell(1+o(1))}{k}\right)\right)\right\} \quad (126)$$

$$= \exp\left(-n e^{-\nu}(1+o(1)) \min_{d \in \mathcal{D}} \left[\rho \cdot D\left(\frac{d\nu\ell}{k} \parallel \frac{\nu\ell}{k}\right) + (1-\rho)D\left(\frac{[\rho d - (1-2\zeta)C]\nu\ell}{(1-\rho)k} \parallel \frac{\nu\ell}{k}\right)\right]\right) \quad (127)$$

$$\leq \exp\left(-\frac{\ell n \nu e^{-\nu}}{k} \left[\min_{d \geq \max\{0, \frac{A}{\rho}\}} g(C, \zeta, d, \rho) + o(1)\right]\right), \quad (128)$$

where:

- the Chernoff bound in (126) is justified by $\tilde{G}_{\mathcal{J}', 2} \sim \text{Bin}((1-\rho)n e^{-\nu}(1+o(1)), \frac{\nu\ell(1+o(1))}{k})$ (note also that $d < d_0$ implies $\rho d - (1-2\zeta)C < 1 - \rho$);
- to obtain (128), we define $A = C(1-2\zeta)$ and relax the range of d from $d \in \mathcal{D}$ to $d \geq \max\{0, \frac{A}{\rho}\}$, and we introduce the shorthand

$$g(C, \zeta, d, \rho) = \rho d \log d + (\rho d - A) \log\left(\frac{\rho d - A}{1 - \rho}\right) + 1 - 2\rho d + A. \quad (129)$$

Then, (128) follows from (127), by using $D(\frac{a\nu\ell}{k} \parallel \frac{\nu\ell}{k}) = \frac{\nu\ell}{k}(a \log a - a + 1 + o(1))$, and noting that $g(C, \zeta, d, \rho)$ defined here coincides with (70) in the proof of converse bound for Bernoulli designs.

Thus, by the same reasoning as (71)–(73), we know that

$$\min_{d \geq \max\{0, \frac{A}{\rho}\}} g(C, \zeta, d, \rho) = g(C, \zeta, d^*, \rho) := f_2(C, \zeta, \rho), \quad (130)$$

with d^* being the same as in (73):

$$d^* = \frac{A + \sqrt{A^2 + 4\rho(1-\rho)}}{2\rho}. \quad (131)$$

Therefore, we obtain a bound on the term in (122) as

$$\exp\left(-\frac{\ell n \nu e^{-\nu}}{k} (f_2(C, \zeta, \rho) + o(1))\right), \quad (132)$$

with $f_2(C, \zeta, \rho)$ being the same as in the proof of Bernoulli design converse bound; see (72).

Comparing the two bounds (132) and (123): We return to (127)–(128), in which we used $D(\frac{a\nu\ell}{k} \parallel \frac{\nu\ell}{k}) = \frac{\nu\ell}{k}(a \log a - a + 1 + o(1))$ and established that

$$\rho \cdot D\left(\frac{d\nu\ell}{k} \parallel \frac{\nu\ell}{k}\right) + (1-\rho) \cdot D\left(\frac{[\rho d - (1-2\zeta)C]\nu\ell}{(1-\rho)k} \parallel \frac{\nu\ell}{k}\right) = \frac{\nu\ell}{k} [g(C, \zeta, d, \rho) + o(1)]. \quad (133)$$

This observation, along with $d_0 = \frac{C(1-2\zeta)+1-\rho}{\rho} > 1$ (recall that we are considering $C(1-2\zeta) > 2\rho - 1$), allows us to compare the relevant terms as follows:

$$\frac{\ell n \nu e^{-\nu}}{k} f_2(C, \zeta, \rho) = \frac{\ell n \nu e^{-\nu}}{k} \min_{d \geq \max\{0, \frac{A}{\rho}\}} g(C, \zeta, d, \rho) \quad (134)$$

$$\leq \frac{\ell n \nu e^{-\nu}}{k} g(C, \zeta, d_0, \rho) \quad (135)$$

$$= n e^{-\nu} \left(\rho D \left(\frac{d_0 \nu \ell}{k} \parallel \frac{\nu \ell}{k} \right) + (1 - \rho) D \left(\frac{[\rho d_0 - (1 - 2\zeta)C] \nu \ell}{(1 - \rho)k} \parallel \frac{\nu \ell}{k} \right) \right) - o \left(\frac{\ell n}{k} \right) \quad (136)$$

$$= \rho n e^{-\nu} D \left(\frac{d_0 \nu \ell}{k} \parallel \frac{\nu \ell}{k} \right) - o \left(\frac{\ell n}{k} \right) \quad (137)$$

$$= \rho n e^{-\nu} (1 + o(1)) D \left(\frac{d_0 \nu \ell}{k} \parallel \frac{\nu \ell}{k} \right), \quad (138)$$

where in (136) we substitute (133), and in (137) we substitute $d_0 = \frac{(1-2\zeta)C+1-\rho}{\rho}$ and observe that the second term in (136) vanishes. Therefore, the bound (132) dominates the one in (123), and it follows that

$$\mathbb{P} \left(\tilde{G}_{\mathcal{J}'} \geq (1 - 2\zeta - o(1)) \frac{C n \nu e^{-\nu} \ell}{k} \right) \leq \exp \left(- \frac{\ell n \nu e^{-\nu}}{k} (f_2(C, \zeta, \rho) + o(1)) \right). \quad (139)$$

Case 2: $(1 - 2\zeta)C \leq 2\rho - 1$. In this case, it suffices to apply the trivial bound

$$\mathbb{P} \left(\tilde{G}_{\mathcal{J}'} \geq (1 - 2\zeta - o(1)) \frac{C n \nu e^{-\nu} \ell}{k} \right) \leq 1, \quad (140)$$

which in turn can trivially be written as $\exp(-\frac{\ell n \nu e^{-\nu}}{k} o(1))$.

Combining Cases 1-2: Some algebra verifies that $f_2(C, \zeta, \rho) = 0$ when $A = (1 - 2\zeta)C = 2\rho - 1$: Substituting this into (131) gives $d^* = 1$, and further we have $f_2(C, \zeta, \rho) = g(C, \zeta, 1, \rho) = 0$ by using (129) and $A = 2\rho - 1$. Thus, we can define the continuous function

$$\hat{f}_2(C, \zeta, \rho) := \begin{cases} f_2(C, \zeta, \rho) & \text{if } (1 - 2\zeta)C \geq 2\rho - 1 \\ 0 & \text{if } (1 - 2\zeta)C < 2\rho - 1. \end{cases} \quad (141)$$

Combining these two cases, we arrive at

$$\mathbb{P} \left(\tilde{G}_{\mathcal{J}'} \geq (1 - 2\zeta - o(1)) \frac{C n \nu e^{-\nu} \ell}{k} \right) \leq \exp \left(- \frac{\ell n \nu e^{-\nu}}{k} (\hat{f}_2(C, \zeta, \rho) + o(1)) \right) \quad (142)$$

for any feasible (C, ζ) with $C \leq C_0$.

The condition for (C2): Now we can bound the probability of **(C2)** holding for some (ℓ, C, ζ)

as

$$\begin{aligned} & \mathbb{P}\left(\mathbf{(C2)} \text{ holds for } (\ell, C, \zeta)\right) \\ & \leq \mathbb{P}\left((117) \text{ holds for } (\ell, C, \zeta)\right) + \mathbb{P}\left((116) \text{ does not hold}\right) \end{aligned} \quad (143)$$

$$\leq \binom{p}{\ell} \cdot \mathbb{P}\left(\tilde{G}_{\mathcal{J}'} \geq (1 - 2\zeta - o(1)) \frac{C n \nu e^{-\nu} \ell}{k}\right) + k^{-10\ell} \quad (144)$$

$$\leq \exp\left((1 + o(1))\ell \log \frac{p}{\ell} - \frac{\ell n \nu e^{-\nu}}{k} (\hat{f}_2(C, \zeta, \rho) + o(1))\right) + k^{-10\ell}, \quad (145)$$

where (143) follows because under the condition (116) we have that (117) is a necessary condition for “**(C2)** holds for the given (ℓ, C, ζ) ”, and in (144) we apply a union bound to account for the maximum over \mathcal{J}' in (117), also recalling that (116) holds with probability at least $1 - k^{-10\ell}$.

Therefore, for given (ℓ, C, ζ) , we have identified the following condition for **(C2)** to hold (recall that our analysis on **(C2)** is on the event \mathcal{A}_1 (118), and in our conclusion below we make this explicit):

$$\frac{\ell n \nu e^{-\nu}}{k} [\hat{f}_2(C, \zeta, \rho) + o(1)] \geq (1 + \eta_2) \ell \log \frac{p}{\ell}, \text{ for some } \eta_2 > 0 \quad (146)$$

$$\begin{aligned} & \implies \mathbb{P}\left(\{\mathbf{(C2)} \text{ holds for } (\ell, C, \zeta)\} \cap \mathcal{A}_1\right) \\ & \leq \exp\left(-\frac{\eta_2}{2} \ell \log \frac{p}{\ell}\right) + k^{-10\ell} \leq \hat{P}_\ell := \begin{cases} (\ell \log k)^{-5}, & \text{if } 1 \leq \ell \leq \log k \\ k^{-5}, & \text{if } \log k < \ell \leq \frac{k}{\log k}, \end{cases} \end{aligned} \quad (147)$$

since $\exp(-\frac{\eta_2}{2} \ell \log \frac{p}{\ell}) + k^{-10\ell}$ is an upper bound on (145) when (146) holds. Recall also that $\mathbb{P}(\mathcal{A}_1) = 1 - o(1)$.

Establishing the threshold

We are now ready to establish the threshold for n above which restricted MLE has $o(1)$ probability of failing. For clarity, we pause to review our previous developments:

- For given ℓ , from (86)–(91), **(C1)** and **(C2)** hold for some feasible $(C, \zeta) \in (C_0, \infty) \times [0, 1]$ with probability at most k^{-5} .
- For given ℓ and any feasible (C, ζ) with $C \leq C_0$, from (103)–(104), if

$$n \geq (1 + \eta_1) \frac{k \log \frac{k}{\ell}}{\nu e^{-\nu} [f_1(C, \zeta, \rho) + o(1)]} \quad (148)$$

holds for some $\eta_1 > 0$, then **(C1)** holds for the given (ℓ, C, ζ) with probability at most \hat{P}_ℓ .

- For given ℓ and any feasible (C, ζ) with $C \leq C_0$, by (146)–(147), if

$$n \geq (1 + \eta_2) \frac{k \log \frac{p}{\ell}}{\nu e^{-\nu} [\hat{f}_2(C, \zeta, \rho) + o(1)]} \quad (149)$$

holds for some $\eta_2 > 0$,⁶ then $\mathbb{P}(\{(\mathbf{C2}) \text{ holds for the given } (\ell, C, \zeta)\} \cap \mathcal{A}_1) \leq \hat{P}_\ell$.

Bounding the failure probability for a fixed ℓ : For a fixed ℓ , we first consider the feasible (C, ζ) with $C \leq C_0$, and by $\frac{Cn\nu e^{-\nu}\ell}{k}, \frac{\zeta \cdot Cn\nu e^{-\nu}\ell}{k} \in \mathbb{Z}$ there are no more than $O(\frac{\ell n}{k}) \times O(\frac{\ell n}{k}) = O((\ell \log k)^2)$ possible choices of such (C, ζ) . If it holds for some $\eta_1 > 0$ that

$$n \geq (1 + \eta_1) \frac{k}{\nu e^{-\nu}} \max_{\substack{0 < C \leq C_0 \\ 0 < \zeta < 1}} \min \left\{ \frac{\log \frac{k}{\ell}}{f_1(C, \zeta, \rho)}, \frac{\log \frac{p}{\ell}}{\hat{f}_2(C, \zeta, \rho)} \right\}, \quad (150)$$

then by the last two dot points reviewed above, for any feasible (C, ζ) with $C \leq C_0$,

$$\mathbb{P}((\mathbf{C1}) \text{ and } (\mathbf{C2}) \text{ hold for } (C, \zeta) \cap \mathcal{A}_1 \text{ holds}) \leq 2\hat{P}_\ell.$$

Moreover, by a union bound over the $O((\ell \log k)^2)$ possibilities of (C, ζ) ,

$$\mathbb{P}((\mathbf{C1}) \text{ and } (\mathbf{C2}) \text{ hold for some feasible } (C, \zeta) \text{ with } C \leq C_0 \cap \mathcal{A}_1 \text{ holds}) \leq O((\ell \log k)^2) \hat{P}_\ell.$$

On the other hand, the first dot point reviewed above gives

$$\mathbb{P}((\mathbf{C1}) \text{ and } (\mathbf{C2}) \text{ hold for some feasible } (C, \zeta) \text{ with } C > C_0 \cap \mathcal{A}_1 \text{ holds}) \leq k^{-5}.$$

Overall, for a fixed $\ell \in [1, \frac{k}{\log k}]$, if (150) holds for some $\eta_1 > 0$, then for the restricted MLE output \hat{S}' (see (30)) we have

$$\begin{aligned} & \mathbb{P}(\{|s \setminus \hat{S}'| = \ell\} \cap \mathcal{A}_1) \\ & \leq \mathbb{P}((\mathbf{C1}) \text{ and } (\mathbf{C2}) \text{ hold for some feasible } (C, \zeta) \cap \mathcal{A}_1 \text{ holds}) \end{aligned} \quad (151)$$

$$\leq k^{-5} + O((\ell \log k)^2) \hat{P}_\ell, \quad (152)$$

where (151) holds because “ $(\mathbf{C1})$ and $(\mathbf{C2})$ hold for some feasible (C, ζ) (under the given ℓ)” is a necessary condition for “ $|s \setminus \hat{S}'| = \ell$ ” due to Lemma 3(b).

Bounding the overall failure probability for $\ell \in [1, \frac{k}{\log k}]$: We further take a union bound over $\ell \in [1, \frac{k}{\log k}]$. We assume that the condition (150) holds for all $\ell \in [1, \frac{k}{\log k}]$, i.e.,

$$n \geq (1 + \eta_1) \frac{k}{\nu e^{-\nu}} \max_{1 \leq \ell \leq \frac{k}{\log k}} \max_{\substack{0 < C \leq C_0 \\ 0 < \zeta < 1}} \min \left\{ \frac{\log \frac{k}{\ell}}{f_1(C, \zeta, \rho)}, \frac{\log \frac{p}{\ell}}{\hat{f}_2(C, \zeta, \rho)} \right\} \quad (153)$$

$$\iff n \geq \frac{(1 + \eta_1) k \log \frac{p}{k}}{(1 - \theta) \nu e^{-\nu}} \frac{1}{\min_{C \in (0, C_0), \zeta \in (0, 1)} \max\{\frac{1}{\theta} f_1(C, \zeta, \rho), \hat{f}_2(C, \zeta, \rho)\}}, \quad (154)$$

where in (154) we observe that the maximum over ℓ is attained at $\ell = 1$ and use $\log p \sim$

⁶Note that $\hat{f}_2(C, \zeta, \rho)$ in the denominator equals to 0 in some cases (see (141)), making the condition (149) vacuous. To show $o(1)$ failure probability for MLE it suffices to show $o(1)$ probability of conditions $(\mathbf{C1})$ and $(\mathbf{C2})$ *simultaneously* holding, so when (149) becomes vacuous, condition $(\mathbf{C1})$ becomes the relevant one.

$\frac{1}{1-\theta} \log \frac{p}{k}$, $\log k \sim \frac{\theta}{1-\theta} \log \frac{p}{k}$. Then, if (154) holds, by union bound we can bound the probability that the restricted MLE decoder fails as

$$\mathbb{P}(\text{err}) \leq \mathbb{P}(\mathcal{A}_1^c) + \sum_{\ell=1}^{k/\log k} \mathbb{P}\left(\{|s \setminus \hat{S}'| = \ell\} \cap \mathcal{A}_1\right) \quad (155)$$

$$\leq o(1) + \sum_{\ell=1}^{k/\log k} \left(k^{-5} + O((\ell \log k)^2) \hat{P}_\ell\right) \quad (156)$$

$$\leq o(1) + \left[\sum_{\ell=1}^{\log k} (\ell \log k)^{-3} + \sum_{\ell=\log k}^{k/\log k} k^{-3} \right] = o(1), \quad (157)$$

where (155) is due to a union bound, and we use (152) in (156) and then substitute \hat{P}_ℓ (given by (147)) in (157). In conclusion, if (154) holds for some $\eta > 0$, then the restricted MLE decoder fails to recover $s = [k]$ with $o(1)$ probability.

Simplifying $\hat{f}_2(C, \zeta, \rho)$ to $f_2(C, \zeta, \rho)$: Recall that $\hat{f}_2(C, \zeta, \rho)$ is given in (141). The aim of this step is to show that the minimum over (C, ζ) in (154) is not attained in the domain of $(1 - 2\zeta)C < 2\rho - 1$, and thus we can safely simplify the $\hat{f}_2(C, \zeta, \rho)$ to $f_2(C, \zeta, \rho)$ in the threshold (154). Consider (C, ζ) satisfying $(1 - 2\zeta)C \leq 2\rho - 1$, i.e.,

$$\zeta \geq \zeta' := \frac{1}{2} + \frac{1 - 2\rho}{2C} \geq \frac{1}{2}, \quad (158)$$

in which we have $\hat{f}_2(C, \zeta, \rho) = 0$ (see (141) and recall that $\hat{f}_2(C, \zeta, \rho)$ is continuous), and that $f_1(C, \zeta, \rho)$ is monotonically increasing with regard to ζ in $[\frac{1}{2}, 1)$ (see (96)). Therefore, we can compare the values of $f_1(C, \zeta, \rho)$ and $f_2(C, \zeta, \rho)$ over (C, ζ) and (C, ζ') as

$$f_1(C, \zeta, \rho) \geq f_1(C, \zeta', \rho), \quad \hat{f}_2(C, \zeta, \rho) = \hat{f}_2(C, \zeta', \rho) = 0, \quad (159)$$

yielding that

$$\max \left\{ \frac{1}{\theta} f_1(C, \zeta, \rho), \hat{f}_2(C, \zeta, \rho) \right\} \geq \max \left\{ \frac{1}{\theta} f_1(C, \zeta', \rho), \hat{f}_2(C, \zeta', \rho) \right\}. \quad (160)$$

Observe that (C, ζ') satisfies $(1 - 2\zeta')C \geq 2\rho - 1$, in which $\hat{f}_2(C, \zeta, \rho) = f_2(C, \zeta, \rho)$, thus we can

proceed as

$$\min_{C \in (0, C_0), \zeta \in (0, 1)} \max \left\{ \frac{1}{\theta} f_1(C, \zeta, \rho), \hat{f}_2(C, \zeta, \rho) \right\} \quad (161)$$

$$= \min_{\substack{C \in (0, C_0), \zeta \in (0, 1) \\ \zeta \leq \frac{1}{2} + \frac{1-2\rho}{2C}}} \max \left\{ \frac{1}{\theta} f_1(C, \zeta, \rho), \hat{f}_2(C, \zeta, \rho) \right\} \quad (162)$$

$$= \min_{\substack{C \in (0, C_0), \zeta \in (0, 1) \\ (1-2\zeta)C \geq 2\rho-1}} \max \left\{ \frac{1}{\theta} f_1(C, \zeta, \rho), f_2(C, \zeta, \rho) \right\} \quad (163)$$

$$\geq \min_{\substack{C > 0 \\ \zeta \in (0, 1)}} \max \left\{ \frac{1}{\theta} f_1(C, \zeta, \rho), f_2(C, \zeta, \rho) \right\}. \quad (164)$$

Therefore, to ensure (154), it suffices to have

$$n \geq \frac{(1+\eta)k \log \frac{p}{k}}{(1-\theta)\nu e^{-\nu}} \frac{1}{\min_{C>0, \zeta \in (0, 1)} \max \left\{ \frac{1}{\theta} f_1(C, \zeta, \rho), f_2(C, \zeta, \rho) \right\}} \quad (165)$$

for some $\eta > 0$. This establishes the second term in (4).

E Low- ℓ Converse Analysis for Near-Constant Weight Designs

In near-constant weight design, each item is independently and uniformly placed at $\Delta = \frac{\nu n}{k}$ tests with replacement. As with the Bernoulli design, it suffices to analyze the MLE decoder (29) that finds \hat{S} having the most correct test outcomes (i.e., outcomes that would be obtained if there were no noise). By the symmetry of the test design with respect to re-ordering items, we can again consider $s = [k]$ as the underlying defective set without loss of generality.

Notation

For $j \in s$ and $j' \in [p] \setminus s$ we will use the notation $\mathcal{M}_j, \mathcal{M}_{j0}, \mathcal{M}_{j1}, \mathcal{N}_0, \mathcal{N}_{00}, \mathcal{N}_{01}$ and the corresponding cardinalities (e.g., $N_0 = |\mathcal{N}_0|$, $M_j = |\mathcal{M}_j|$) from Corollary 1. These index sets and their cardinalities are known under given \mathbf{X}_s and \mathbf{Y} . We consider $\ell = 1$, and say that $(C, \zeta) \in (0, \infty) \times (0, 1)$ is feasible if $\frac{C\nu e^{-\nu}}{k}$ and $\frac{\zeta C\nu e^{-\nu}}{k}$ are integers. Given feasible (C, ζ) we will consider $\mathcal{K}_{C, \zeta} = \{j \in s : M_j = \frac{C\nu e^{-\nu}}{k}, M_{j0} = \frac{\zeta C\nu e^{-\nu}}{k}\}$ with cardinality $k_{C, \zeta} = |\mathcal{K}_{C, \zeta}|$ as in (35). For given $(j, j') \in \mathcal{K}_{C, \zeta} \times ([p] \setminus s)$, recall that $G_{j, j', 1}$ denotes the number of tests in $\mathcal{N}_{01} \cup \mathcal{M}_{j1}$ that contain the non-defective item j' , and $G_{j, j', 2}$ denotes number of tests in $\mathcal{N}_{00} \cup \mathcal{M}_{j0}$ that contain the non-defective item j' ; see Corollary 1.

We need some additional conventions for analyzing the near-constant design. For $j \in s$ and $j' \in [p] \setminus s$, we let $\mathcal{M}_{j, j'}$ index the tests in \mathcal{M}_j that contain item j' , and $\mathcal{N}_{0, j'}$ index the tests in \mathcal{N}_0 that contain item j' . Then we note the equivalent interpretations of $G_{j, j', 1}$ and $G_{j, j', 2}$: $G_{j, j', 1}$ represents the number of positive tests in $\mathcal{N}_{0, j'} \cup \mathcal{M}_{j, j'}$, and $G_{j, j', 2}$ represents the number of negative tests in $\mathcal{N}_{0, j'} \cup \mathcal{M}_{j, j'}$; this alternative formulation happens to be more convenient for analyzing the

near-constant weight design.

In addition, for $j \in s$, we let \mathcal{M}'_j index the tests where item j not only appears as the only defective but is also placed precisely once, and let its cardinality be $M'_j = |\mathcal{M}'_j|$. Compared to Bernoulli designs, this is a new ingredient for handling near-constant weight designs whose random placement is done with replacement. Evidently, it holds that $M'_j \leq M_j$, and $M'_j < M_j$ occurs when item j is placed in some test more than once but is still the only defective. This situation only occurs for a small number of defective items, as we will formalize via event \mathcal{A}_1 below.

Reduction to two conditions

Similarly to the converse analysis for Bernoulli design, by Corollary 1, MLE fails if for some feasible $(C, \zeta) \in (0, e^\nu) \times (0, 1)$, the following two conditions **(C1)** and **(C2)** simultaneously hold:

- **(C1)** $k_{C,\zeta} \geq 1$ (i.e., $\mathcal{K}_{C,\zeta} \neq \emptyset$);
- **(C2)** There exist $j \in \mathcal{K}_{C,\zeta}$ and $k+1 \leq j' \leq p$ such that

$$G_{j,j',1} - G_{j,j',2} > (1 - 2\zeta) \frac{Cn\nu e^{-\nu}}{k}. \quad (166)$$

Some useful events

Before proceeding, we first construct three high-probability events. By Lemma 5(a) in Appendix G, for some sufficiently large constant $C_0 > 0$, the event

$$\mathcal{A}_1 = \left\{ |\{j \in [k] : M'_j < M_j\}| \leq C_0 \log k \right\}. \quad (167)$$

holds with $1 - o(1)$ probability. In addition, recalling that M'_j is the number of tests in which item j is placed precisely once and is the only defective, we further define

$$\widetilde{M} := \sum_{j=1}^k M'_j, \quad (168)$$

which represents the number of tests connected to precisely one defective item. Lemma 5(b) in Appendix G gives that the event

$$\mathcal{A}_2 = \left\{ \widetilde{M} = (1 + o(1))e^{-\nu}k\Delta \right\} \quad (169)$$

holds with probability $1 - o(1)$.

Lastly, we deduce the high-probability behaviour of N_0 . For $\mathcal{M} \subset s$, letting $W^{(\mathcal{M})}$ denote the number of tests in which some item from \mathcal{M} is placed, Lemma 7 in Appendix G gives that $\mathbb{E}W^{(s)} = (1 - e^{-\nu} + o(1))n$, and that $W^{(s)} = \mathbb{E}W^{(s)} + \delta'n\sqrt{\frac{\ell}{k}}$ holds with probability at least $1 - 2\exp(-2\nu^{-1}\delta^2n)$ for any $\delta > 0$ and for some $\delta' < \delta$. Therefore, by letting $\delta = \Theta(\frac{1}{\sqrt{k}})$ we obtain

$W^{(s)} = (1 - e^{-\nu} + o(1))n$, which implies $N_0 = n - W^{(s)} = ne^{-\nu}(1 + o(1))$, holds with $1 - o(1)$ probability. Thus, the event

$$\mathcal{A}_3 := \{N_0 = ne^{-\nu}(1 + o(1))\} \quad (170)$$

holds with $1 - o(1)$ probability. Throughout the proof, we will often apply the conditions in $\mathcal{A}_1 \cap \mathcal{A}_2 \cap \mathcal{A}_3$, which is justified via simple union bound.

Condition for (C1)

Let $\xi \in (\frac{1}{2}, 1)$ be such that k^ξ is an integer. Given $s_0 \subset s = [k]$, we let \mathbf{T}_{s_0} be an *unordered multi-set* of length $|s_0|\Delta$ whose entries are in $[n]$, with the overall multi-set representing the $|s_0|\Delta$ placements from the items in s_0 in an unordered manner. If $s_0 = \{j\}$ for some $j \in [k]$, then we simply write $\mathbf{T}_{\{j\}} = \mathbf{T}_j$ for the Δ placements of item j .

Compared to M_j , it is more convenient to deal with M'_j . Therefore, instead of directly studying $\mathcal{K}_{C,\zeta}$, we let M'_{j0} be the number of negative tests (with results flipped by noise) in \mathcal{M}'_j and define

$$\mathcal{K}'_{C,\zeta} = \left\{j \in [k] : M'_j = \frac{Cn\nu e^{-\nu}}{k}, M'_{j0} = \frac{\zeta \cdot Cn\nu e^{-\nu}}{k}\right\}, \quad (171)$$

and let $k'_{C,\zeta} = |\mathcal{K}'_{C,\zeta}|$. To quantify the difference between $\mathcal{K}_{C,\zeta}$ and $\mathcal{K}'_{C,\zeta}$, we provide the following observation

$$\mathcal{K}'_{C,\zeta} \subset \mathcal{K}_{C,\zeta} \cup \{j \in [k] : M'_j < M_j\}. \quad (172)$$

Combining with the event \mathcal{A}_1 in (167), this implies

$$k'_{C,\zeta} \leq k_{C,\zeta} + C_0 \log k. \quad (173)$$

Given $M'_j = \frac{Cn\nu e^{-\nu}}{k}$, we have $M'_{j0} \sim \text{Bin}(\frac{Cn\nu e^{-\nu}}{k}, \rho)$ based on the randomness of \mathbf{Z} , so $M'_{j0} = \frac{\zeta \cdot Cn\nu e^{-\nu}}{k}$ can be easily analyzed. Thus, the major challenge lies in analyzing the event $M'_j = \frac{Cn\nu e^{-\nu}}{k}$, and we now proceed to study the distribution of M'_j .

The distribution of M'_j : We follow the idea in [16, Lemma 3.3] of conditioning on \mathbf{T}_s (i.e., $\mathbf{T}_{[k]}$) and using the symmetry of the test design to deduce that the conditional assignments of tests to items are uniform, i.e., every partition of \mathbf{T}_s into k sets of size- Δ is equally likely. In particular, note that \widetilde{M} defined in (168) is a known quantity from the given \mathbf{T}_s , and given \widetilde{M} , the number of such tests assigned to any given $j \in s$, which has been defined as M'_j , follows a hypergeometric distribution: There are Δ draws (without replacement) from $k\Delta$ objects (that represent the overall $k\Delta$ placements in \mathbf{T}_s), \widetilde{M} of which are of a special type (namely, the placement is to a test that is connected to precisely one defective item). For instance, regarding the allocation of these $k\Delta$ placements to item 1, we have $(M'_1 | \mathbf{T}_s) \sim \text{Hg}(k\Delta, \widetilde{M}, \Delta)$. To understand the behavior of the first k^ξ items simultaneously, we interpret the above-mentioned uniform allocation from \mathbf{T}_s as being done

sequentially:

- Given \mathbf{T}_s , we allocate a random size- Δ subset to item 1, which yields $(M'_1|\mathbf{T}_s) \sim \text{Hg}(k\Delta, \widetilde{M}, \Delta)$ as described above.
- Given \mathbf{T}_s and \mathbf{T}_1 (the placements for item 1), we allocate a random size- Δ subset of the *reduced multi-set* $\mathbf{T}_s \setminus \mathbf{T}_1$ to item 2, which gives $M'_2|(\mathbf{T}_s, \mathbf{T}_1) \sim \text{Hg}((k-1)\Delta, \widetilde{M} - M'_1, \Delta)$ (where M'_1 is determined via $\mathbf{T}_s, \mathbf{T}_1$).
- ...
- Given $\mathbf{T}_s, \mathbf{T}_1, \dots, \mathbf{T}_{k^\xi-1}$, the quantities $M'_1, M'_2, \dots, M'_{k^\xi-1}$ are known, and an analogous argument gives $M'_{k^\xi}|(\mathbf{T}_s, \{\mathbf{T}_j\}_{j=1}^{k^\xi-1}) \sim \text{Hg}((k - k^\xi + 1)\Delta, \widetilde{M} - \sum_{j=1}^{k^\xi-1} M'_j, \Delta)$.

Similarly to the analysis of the Bernoulli design, we only consider this procedure up to index k^ξ , since going all the way up to k would introduce non-negligible dependencies. For $\xi \in (\frac{1}{2}, 1)$ we have $k - j = (1 + o(1))k$ for any $0 \leq j \leq k^\xi$. Then, we utilize the event \mathcal{A}_2 in (169) that specifies $\widetilde{M} = (1 + o(1))e^{-\nu}k\Delta$, and write

$$(1 + o(1))e^{-\nu}k\Delta = \widetilde{M} \geq \widetilde{M} - \sum_{j=1}^i M'_j \geq \widetilde{M} - k^\xi\Delta = (1 + o(1))e^{-\nu}k\Delta \quad (174)$$

for any $0 \leq i \leq k^\xi$. Therefore, the following holds for any $1 \leq j \leq k^\xi$:

$$M'_j|(\mathbf{T}_s, \mathbf{T}_1, \dots, \mathbf{T}_{j-1}) \sim \text{Hg}((1 + o(1))k\Delta, (1 + o(1))e^{-\nu}k\Delta, \Delta). \quad (175)$$

The Chernoff bound and matching anti-concentration for the hypergeometric distribution (see Lemma 6 in Appendix G) gives that

$$\mathbb{P}\left(M'_j = \frac{Cn\nu e^{-\nu}}{k} | (\mathbf{T}_s, \mathbf{T}_1, \dots, \mathbf{T}_{j-1})\right) = \exp\left(-\frac{\nu n}{k} [D(Ce^{-\nu}\|e^{-\nu}) + o(1)]\right) \quad (176)$$

for any $1 \leq j \leq k^\xi$.

Bounding $\mathbb{P}(j \in \mathcal{K}'_{C,\zeta})$ from below: Under $M'_j = \frac{Cn\nu e^{-\nu}}{k}$, with the randomness of \mathbf{Z} we have $M'_{j0} \sim \text{Bin}(\frac{Cn e^{-\nu}\nu}{k}, \rho)$, which should equal $\frac{\zeta \cdot Cn\nu e^{-\nu}}{k}$ to ensure $j \in \mathcal{K}'_{C,\zeta}$. Specifically, for $1 \leq j \leq k^\xi$

we proceed as follows:

$$\mathbb{P}(j \in \mathcal{K}'_{C,\zeta} | (\mathbf{T}_s, \mathbf{T}_1, \dots, \mathbf{T}_{j-1})) \quad (177)$$

$$= \mathbb{P}\left(M'_j = \frac{Ce^{-\nu}\nu n}{k}, M'_{j0} = \frac{\zeta \cdot Cn\nu e^{-\nu}}{k} | (\mathbf{T}_s, \mathbf{T}_1, \dots, \mathbf{T}_{j-1})\right) \quad (178)$$

$$= \mathbb{P}\left(M'_j = \frac{Cn\nu e^{-\nu}}{k} | (\mathbf{T}_s, \mathbf{T}_1, \dots, \mathbf{T}_{j-1})\right) \cdot \mathbb{P}\left(\text{Bin}\left(\frac{C\nu e^{-\nu}n}{k}, \rho\right) = \frac{\zeta \cdot C\nu e^{-\nu}n}{k}\right) \quad (179)$$

$$\geq \exp\left(-\frac{\nu n e^{-\nu}}{k} \left[e^\nu D(Ce^{-\nu}\|e^{-\nu}) + C \cdot D(\zeta\|\rho) + o(1)\right]\right) \quad (180)$$

$$:= \exp\left(-\frac{\nu n e^{-\nu}}{k} [f_1(C, \zeta, \rho, \nu) + o(1)]\right) := P_{\text{in}}, \quad (181)$$

where in (180) we substitute (176) and apply (26) in Lemma 1 (with leading factor absorbed into the exponent), and in (181) we define

$$f_1(C, \zeta, \rho, \nu) = e^\nu D(Ce^{-\nu}\|e^{-\nu}) + C \cdot D(\zeta\|\rho). \quad (182)$$

Deriving the condition for (C1): Noting that

$$k'_{C,\zeta} = \sum_{j=1}^k \mathbb{1}(j \in \mathcal{K}'_{C,\zeta}) \geq \sum_{j=1}^{k^\xi} \mathbb{1}(j \in \mathcal{K}'_{C,\zeta}) := k'_{C,\zeta,\xi}, \quad (183)$$

and from (181) we have $(k'_{C,\zeta,\xi} | \mathbf{T}_s)$ stochastically dominates $\text{Bin}(k^\xi, P_{\text{in}})$, and thus

$$\mathbb{P}(k'_{C,\zeta,\xi} \leq z | \mathbf{T}_s) \leq \mathbb{P}(\text{Bin}(k^\xi, P_{\text{in}}) \leq z) \quad (184)$$

for any z . Combining this property with (173) and (183), we can proceed as

$$\mathbb{P}(k_{C,\zeta} = 0 | \mathbf{T}_s) \leq \mathbb{P}(k'_{C,\zeta} \leq C_0 \log k | \mathbf{T}_s) \quad (185)$$

$$\leq \mathbb{P}(k'_{C,\zeta,\xi} \leq C_0 \log k | \mathbf{T}_s) \quad (186)$$

$$= \mathbb{P}(\text{Bin}(k^\xi, P_{\text{in}}) \leq C_0 \log k), \quad (187)$$

and since this holds regardless of the conditioning variable, we obtain

$$\mathbb{P}(k_{C,\zeta} = 0) \leq \mathbb{P}(\text{Bin}(k^\xi, P_{\text{in}}) \leq C_0 \log k). \quad (188)$$

We now claim that to ensure that (C1) holds with $1 - o(1)$ probability, it suffices to ensure

$$k^\xi P_{\text{in}} = \exp\left(\xi \log k - \frac{\nu n e^{-\nu} [f_1(C, \zeta, \rho, \nu) + o(1)]}{k}\right) \geq \exp\left(\eta_1 \log \frac{p}{k}\right) \rightarrow \infty \quad (189)$$

for some $\eta_1 > 0$. Indeed, if (189) holds, then by the Chernoff bound (see (24) in Lemma 1), we have

$$\mathbb{P}\left(\text{Bin}(k^\xi, P_{\text{in}}) \leq C_0 \log k\right) \leq \exp\left(-k^\xi P_{\text{in}} \left[\frac{C_0 \log k}{k^\xi P_{\text{in}}} \log\left(\frac{C_0 \log k}{k^\xi P_{\text{in}}}\right) + \frac{C_0 \log k}{k^\xi P_{\text{in}}} - 1\right]\right) \quad (190)$$

$$\leq \exp\left(-\frac{1}{2}k^\xi P_{\text{in}}\right), \quad (191)$$

where the second line holds since $\frac{C_0 \log k}{k^\xi P_{\text{in}}} \leq \frac{C_0 \log k}{\exp(\eta_1 \log \frac{p}{k})} = o(1)$ for large enough k . The claim thus follows by combining with (188).

It remains to deduce the threshold from (189). By $k = \Theta(p^\theta)$ we have $\log k \sim \frac{\theta}{1-\theta} \log \frac{p}{k}$, and then by choosing ξ arbitrarily close to 1 we can ensure (189) by

$$n \leq (1 - \eta_1) \frac{\frac{\theta}{1-\theta} k \log \frac{p}{k}}{\nu e^{-\nu} f_1(C, \zeta, \rho, \nu)} \quad (192)$$

for some $\eta_1 > 0$ possibly different from the one in (189).

Condition for (C2)

The subsequent analysis is built on condition **(C1)** that ensures $k_{C,\zeta} \geq 1$, and we consider a single specific index $j \in \mathcal{K}_{C,\zeta}$. We will study the placement of each non-defective item j' for $k+1 \leq j' \leq p$. For the given $j \in \mathcal{K}_{C,\zeta}$ and a fixed $k+1 \leq j' \leq p$, we define $R_{j'} = |\mathcal{N}_{0,j'} \cup \mathcal{M}_{j,j'}|$ as the number of tests in $\mathcal{N}_0 \cup \mathcal{M}_j$ that contain item j' , and define $R'_{j'}$ as the number of placements of item j' that are in $\mathcal{N}_0 \cup \mathcal{M}_j$. It is evident that $R'_{j'} \geq R_{j'}$ always holds, and $R'_{j'} > R_{j'}$ when item j' is placed more than once into some test in $\mathcal{N}_0 \cup \mathcal{M}_j$. To formulate condition **(C2)**, we are primarily interested in $R_{j'}$, but it is more convenient to first study $R'_{j'}$. Note that under the event \mathcal{A}_3 in (170) which states

$$ne^{-\nu}(1 + o(1)) = N_0 \leq |\mathcal{N}_0 \cup \mathcal{M}_j| \leq N_0 + \ell\Delta = ne^{-\nu}(1 + o(1)), \quad (193)$$

a placement of item j' increments $R'_{j'}$ by 1 with probability $e^{-\nu}(1 + o(1))$, thus we have $R'_{j'} \sim \text{Bin}(\Delta, e^{-\nu}(1 + o(1)))$ with respect to the randomness in placing item j' into tests.

Bounding the difference between $R_{j'}$ and $R'_{j'}$: We are ultimately interested in $R_{j'}$, but we have characterized $R'_{j'}$ as an intermediate step. Accordingly, we proceed to quantify the difference between $R_{j'}$ and $R'_{j'}$. We envision that the Δ placements of item j' are done subsequently, and let $D_{j'}$ be the number of “collisions”, defined as

$$D_{j'} = \sum_{i=1}^{\Delta} \mathbb{1}(\text{the } i\text{-th placement coincides with the } i_1\text{-th placement for some } i_1 < i). \quad (194)$$

Then we have

$$R'_{j'} - D_{j'} \leq R_{j'} \leq R'_{j'}, \quad (195)$$

and we will show that $D_{j'}$ is asymptotically negligible. Specifically, observe that the i -th placement

increments $D_{j'}$ by 1 with probability less than $\frac{\Delta}{n}$, so for any $D_0 > 0$, a union bound (over the $\left(\frac{\Delta}{D_0}\right)$ possibilities of the D_0 placements that increment $D_{j'}$) gives

$$\mathbb{P}\left(D_{j'} \geq D_0\right) \leq \left(\frac{\Delta}{D_0}\right)\left(\frac{\Delta}{n}\right)^{D_0} \leq \left(\frac{e\Delta^2}{D_0 n}\right)^{D_0}. \quad (196)$$

Thus, we take $D_0 = \frac{10}{\theta}$ in (196) and apply a union bound over $k+1 \leq j' \leq p$, yielding

$$\mathbb{P}\left(\sup_{k+1 \leq j' \leq p} D_{j'} \geq \frac{10}{\theta}\right) \leq p \cdot \left(\frac{e\theta\Delta^2}{10n}\right)^{\frac{10}{\theta}} \quad (197)$$

$$= \exp\left(\log p + \frac{10}{\theta} \log\left(\frac{e\theta\nu^2 n}{10k}\right) - \frac{10}{\theta} \log k\right) \quad (198)$$

$$\leq k^{-5}, \quad (199)$$

where (199) holds when k is large enough because $p = \Theta(k^{1/\theta})$ and $n = \Theta(k \log k)$. In subsequent analyses, we implicitly suppose (via a union bound) that we are on the event

$$D_{j'} < \frac{10}{\theta}, \quad \forall k+1 \leq j' \leq p, \quad (200)$$

which holds with probability at least $1 - k^{-5}$.

Studying the condition for (166): Given $R_{j'}$, the randomness of the noise \mathbf{Z} gives

$$G_{j,j',1} \sim \text{Bin}(R_{j'}, \rho), \quad G_{j,j',2} = R_{j'} - G_{j,j',1}. \quad (201)$$

Thus, (166) can be expressed as $2G_{j,j',1} - R_{j'} > (1 - 2\zeta)\frac{Cn\nu e^{-\nu}}{k}$, or equivalently

$$G_{j,j',1} > \left(\frac{1}{2} - \zeta\right)\frac{Cn\nu e^{-\nu}}{k} + \frac{1}{2}R_{j'}. \quad (202)$$

By introducing a parameter d (to be chosen later) that satisfies

$$|C(1 - 2\zeta)| \leq d \leq e^\nu, \quad (203)$$

we consider the event $R'_{j'} = \frac{dnve^{-\nu}}{k}$ and proceed as follows:

$$\mathbb{P}\left(G_{j,j',1} > \left(\frac{1}{2} - \zeta\right) \frac{Cnve^{-\nu}}{k} + \frac{1}{2}R_{j'}\right) \quad (204)$$

$$\geq \mathbb{P}\left(R'_{j'} = \frac{dnve^{-\nu}}{k}\right) \mathbb{P}\left(\text{Bin}(R_{j'}, \rho) > \left(\frac{1}{2} - \zeta\right) \frac{Cnve^{-\nu}}{k} + \frac{1}{2}R_{j'} \mid R'_{j'} = \frac{dnve^{-\nu}}{k}\right) \quad (205)$$

$$\geq \mathbb{P}\left(\text{Bin}\left(\frac{\nu n}{k}, e^{-\nu} + o(1)\right) = \frac{dnve^{-\nu}}{k}\right) \cdot \mathbb{P}\left(\text{Bin}\left(\frac{(1+o(1))dnve^{-\nu}}{k}, \rho\right) > \frac{d+C(1-2\zeta)}{2} \frac{nve^{-\nu}}{k}\right) \quad (206)$$

$$\geq \exp\left(-\frac{\nu ne^{-\nu}}{k} \left[e^{\nu} \cdot D(de^{-\nu} \| e^{-\nu}) + d \cdot D\left(\frac{1}{2} + \frac{C(1-2\zeta)}{2d} \parallel \rho\right) + o(1)\right]\right) \quad (207)$$

$$:= \exp\left(-\frac{\nu ne^{-\nu}}{k} [g(C, \zeta, d, \rho, \nu) + o(1)]\right), \quad (208)$$

where in (206) we use $R'_{j'} \sim \text{Bin}(\Delta, e^{-\nu} + o(1))$ along with the event (200) and $R'_{j'} = \frac{dnve^{-\nu}}{k}$, from which (195) gives $R_{j'} = \frac{dnve^{-\nu}(1+o(1))}{k}$ (since $\frac{n}{k} \rightarrow \infty$ and $\frac{10}{\theta}$ is a constant); then in (207) we apply (26) in Lemma 1 to the two probability terms (with leading factors absorbed into the exponent), and in (208) we introduce the following shorthand:

$$g(C, \zeta, d, \rho, \nu) = e^{\nu} \cdot D(de^{-\nu} \| e^{-\nu}) + d \cdot D\left(\frac{1}{2} + \frac{C(1-2\zeta)}{2d} \parallel \rho\right). \quad (209)$$

This depends on (C, ζ) only through $A := C(1-2\zeta)$, and can be expanded as

$$g(C, \zeta, d, \rho, \nu) = d \log d + (e^{\nu} - d) \log \frac{1 - de^{-\nu}}{1 - e^{-\nu}} + \frac{d+A}{2} \log \frac{d+A}{2d\rho} + \frac{d-A}{2} \log \frac{d-A}{2d(1-\rho)}. \quad (210)$$

Optimizing d : Our next step is to specify d that minimizes $g(C, \zeta, d, \rho, \nu)$ to render the tightest lower bound (208). Note that range of d is given by $|A| \leq d \leq e^{\nu}$ in (203), and we define

$$d^* = \arg \min g(C, \zeta, d, \rho, \nu), \quad \text{subject to } |A| \leq d \leq e^{\nu}. \quad (211)$$

To pinpoint d^* , we differentiate $g(C, \zeta, d, \rho, \nu)$ with respect to d to obtain

$$\frac{\partial g}{\partial d} = \log\left(\frac{\sqrt{d^2 - A^2}}{2\sqrt{\rho(1-\rho)}}\right) - \log\left(\frac{e^{\nu} - d}{e^{\nu} - 1}\right). \quad (212)$$

Note that $\frac{\partial g}{\partial d}$ has the same sign as $\frac{d^2 - A^2}{4\rho(1-\rho)} - \frac{(e^{\nu} - d)^2}{(e^{\nu} - 1)^2}$, which expands as

$$g^{(1)}(d) := \left(\frac{1}{4\rho(1-\rho)} - \frac{1}{(e^{\nu} - 1)^2}\right)d^2 + \frac{2e^{\nu}}{(e^{\nu} - 1)^2}d - \left(\frac{A^2}{4\rho(1-\rho)} + \frac{e^{2\nu}}{(e^{\nu} - 1)^2}\right), \quad (213)$$

where we only regard d as the variable. We note $g^{(1)}(0) < 0$ and perform some algebra to find that

$$g^{(1)}(|A|) = -\left(\frac{e^\nu - A}{e^\nu - 1}\right)^2 \leq 0, \quad g^{(1)}(e^\nu) = \frac{e^{2\nu} - A^2}{4\rho(1-\rho)} \geq 0. \quad (214)$$

Based on these observations, we determine d^* by considering the following cases:

- When $4\rho(1-\rho) < (e^\nu - 1)^2$, $g^{(1)}(d)$ has a unique zero in $d > 0$. By (214) this zero falls in the range of feasible d (i.e., $[|A|, e^\nu]$), so it is precisely the desired d^* , yielding

$$d^* = \frac{-e^\nu + \sqrt{e^{2\nu} + \left(\frac{(e^\nu-1)^2}{4\rho(1-\rho)} - 1\right)\left(\frac{A^2(e^\nu-1)^2}{4\rho(1-\rho)} + e^{2\nu}\right)}}{\frac{(e^\nu-1)^2}{4\rho(1-\rho)} - 1}, \quad \text{if } 4\rho(1-\rho) < (e^\nu - 1)^2. \quad (215)$$

- When $4\rho(1-\rho) = (e^\nu - 1)^2$, using (214) it is easy to see that d^* is the unique zero of $g^{(1)}(d)$, and thus

$$d^* = \frac{A^2 + e^{2\nu}}{2e^\nu}, \quad \text{if } 4\rho(1-\rho) = (e^\nu - 1)^2. \quad (216)$$

- When $4\rho(1-\rho) > (e^\nu - 1)^2$, we can verify that the two zeros of $g^{(1)}(d)$ both fall in $(0, \infty)$. Combining with (214), we find that d^* equals the smaller zero of $g^{(1)}(d)$:

$$d^* = \frac{-e^\nu + \sqrt{e^{2\nu} + \left(\frac{(e^\nu-1)^2}{4\rho(1-\rho)} - 1\right)\left(\frac{A^2(e^\nu-1)^2}{4\rho(1-\rho)} + e^{2\nu}\right)}}{\frac{(e^\nu-1)^2}{4\rho(1-\rho)} - 1}, \quad \text{if } 4\rho(1-\rho) > (e^\nu - 1)^2. \quad (217)$$

Therefore, d^* defined in (211) is given by

$$d^* = \begin{cases} (215), & \text{when } 4\rho(1-\rho) \neq (e^\nu - 1)^2 \\ \frac{A^2 + e^{2\nu}}{2e^\nu}, & \text{when } 4\rho(1-\rho) = (e^\nu - 1)^2 \end{cases}. \quad (218)$$

A simple limiting argument verifies that d^* is continuous with respect to (ρ, ν) : When $a := \frac{(e^\nu-1)^2}{4\rho(1-\rho)} \rightarrow 1$, we have

$$\lim_{a \rightarrow 1} \frac{-e^\nu + \sqrt{e^{2\nu} + (a-1)(A^2a + e^{2\nu})}}{a-1} \quad (219)$$

$$= \lim_{a \rightarrow 1} \frac{(a-1)(A^2a + e^{2\nu})}{(a-1)(\sqrt{e^{2\nu} + (a-1)(A^2a + e^{2\nu})} + e^\nu)} = \frac{A^2 + e^{2\nu}}{2e^\nu}. \quad (220)$$

In summary, we can introduce the shorthand

$$f_2(C, \zeta, \rho, \nu) = g(C, \zeta, d^*, \rho, \nu) \quad (221)$$

and arrive at the following conclusion: For fixed $j \in \mathcal{K}_{C,\zeta}$ and $j' \in [p] \setminus s$ we have

$$\mathbb{P}\left(G_{j,j',1} > \left(\frac{1}{2} - \zeta\right) \frac{Cn\nu e^{-\nu}}{k} + \frac{1}{2}R_{j'}\right) \geq \exp\left(-\frac{\nu n e^{-\nu}}{k}[f_2(C, \zeta, \rho, \nu) + o(1)]\right) \quad (222)$$

Recall that **(C2)** requires that (202) holds for some $j' \in [p] \setminus s$, and the placements of the $p - k$ non-defective items are independent. Therefore, by arguments analogous to (75)–(81), we obtain the threshold

$$n \leq \frac{(1 - \eta_2) \frac{1}{1-\theta} k \log \frac{p}{k}}{\nu e^{-\nu} (f_2(C, \zeta, \rho, \nu) + o(1))} \quad (223)$$

for some $\eta_2 > 0$, which suffices for ensuring **(C2)** holds with $1 - o(1)$ probability.

Wrapping up

To ensure the failure of MLE, we need **(C1)** and **(C2)** simultaneously hold for some feasible (C, ζ) . By the above analyses, for given (C, ζ) we have **(C1)** and **(C2)** with $1 - o(1)$ probability if both (192) and (223) hold, and by merging η_1, η_2 into a single η , this can be written as

$$n \leq \frac{(1 - \eta) k \log \frac{p}{k}}{(1 - \theta) \nu e^{-\nu}} \frac{1}{\max\{\frac{1}{\theta} f_1(C, \zeta, \rho, \nu), f_2(C, \zeta, \rho, \nu)\}} \quad (224)$$

for some $\eta > 0$. Optimizing over (C, ζ) establishes the second term in (11).

F Low- ℓ Achievability Analysis for Near-Constant Weight Designs

Recall that in this part of the analysis we only need to consider $\ell \in [1, \frac{k}{\log k}]$. We will establish the threshold for n above which the restricted MLE decoder (29) has $o(1)$ probability of failing. As with the Bernoulli design, this is significantly more challenging than the converse. We suppose that the true defective set is $s = [k]$ without loss of generality.

Notation

We first recap some notation that we also used when studying the Bernoulli design. For $\mathcal{J} \subset s$ with $|\mathcal{J}| = \ell$ we will use the notation $\mathcal{M}_{\mathcal{J}}, \mathcal{M}_{\mathcal{J}0}, \mathcal{M}_{\mathcal{J}1}, \mathcal{N}_0, \mathcal{N}_{00}, \mathcal{N}_{01}$ and their corresponding cardinalities from Lemma 3. Note that conditioned on \mathbf{X}_s and \mathbf{Y} , these index sets and their cardinalities are deterministic. For $\ell \in [1, \frac{k}{\log k}]$, we say that $(C, \zeta) \in [0, \infty) \times [0, 1]$ is *feasible* if $\frac{Cn\nu e^{-\nu}\ell}{k}$ and $\frac{\zeta \cdot Cn\nu e^{-\nu}\ell}{k}$ are integers, subject to the restriction that $(C, \zeta) = (0, 0)$ is the only feasible pair with $C = 0$. For feasible (C, ζ) , we define the set

$$\mathcal{K}_{\ell, C, \zeta} = \left\{ \mathcal{J} \subset s, |\mathcal{J}| = \ell : M_{\mathcal{J}} = \frac{Cn\nu e^{-\nu}\ell}{k}, M_{\mathcal{J}0} = \frac{\zeta \cdot Cn\nu e^{-\nu}\ell}{k} \right\}$$

with cardinality $k_{\ell, C, \zeta} = |\mathcal{K}_{\ell, C, \zeta}|$, as defined in (32). For given $\mathcal{J} \in \mathcal{K}_{\ell, C, \zeta}$ and $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$, recall that $G_{\mathcal{J}, \mathcal{J}', 1}$ denotes the number of tests in $\mathcal{N}_{01} \cup \mathcal{M}_{\mathcal{J}1}$ that contain some item from

\mathcal{J}' , and $G_{\mathcal{J},\mathcal{J}',2}$ denotes the number of tests in $\mathcal{N}_{00} \cup \mathcal{M}_{\mathcal{J}0}$ that contain some item from \mathcal{J}' .

We also need some additional notation when studying the near-constant weight design. For $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$, we let $\mathcal{M}_{\mathcal{J},\mathcal{J}'}$ index the tests in $\mathcal{M}_{\mathcal{J}}$ that contain some item from \mathcal{J}' , and $\mathcal{N}_{0,\mathcal{J}'}$ index the tests in \mathcal{N}_0 that contain some item in \mathcal{J}' . With these notations, we can equivalently interpret $G_{\mathcal{J},\mathcal{J}',1}$ as the number of positive tests in $\mathcal{M}_{\mathcal{J},\mathcal{J}'} \cup \mathcal{N}_{0,\mathcal{J}'}$, and $G_{\mathcal{J},\mathcal{J}',2}$ as the number of negative tests in $\mathcal{M}_{\mathcal{J},\mathcal{J}'} \cup \mathcal{N}_{0,\mathcal{J}'}$. Furthermore, we note that

$$G_{\mathcal{J},\mathcal{J}',2} = M_{\mathcal{J},\mathcal{J}'} + N_{0,\mathcal{J}'} - G_{\mathcal{J},\mathcal{J}',1}. \quad (225)$$

Reduction to two conditions

We seek to bound the probability that restricted MLE fails, considering each $\ell \in [1, \frac{k}{\log k}]$ separately. As before, we observe the following via Lemma 3(b): If restricted MLE fails and returns some s' with $|s \setminus s'| \in [1, \frac{k}{\log k}]$, then for $\ell := |s \setminus s'|$, there exists some feasible $(C, \zeta) \in [0, \infty) \times [0, 1]$ with respect to the specific ℓ such that **(C1)** and **(C2)** below simultaneously hold:

- **(C1)** $k_{\ell,C,\zeta} \geq 1$ (i.e., $\mathcal{K}_{\ell,C,\zeta} \neq \emptyset$);
- **(C2)** There exist $\mathcal{J} \in \mathcal{K}_{\ell,C,\zeta}$ and some $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$ such that

$$G_{\mathcal{J},\mathcal{J}',1} - G_{\mathcal{J},\mathcal{J}',2} \geq (1 - 2\zeta) \frac{Cn\nu e^{-\nu}\ell}{k}. \quad (226)$$

By substituting (225), this can be equivalently formulated as

$$G_{\mathcal{J},\mathcal{J}',1} \geq \left(\frac{1}{2} - \zeta\right) \frac{Cn\nu e^{-\nu}\ell}{k} + \frac{1}{2}(N_{0,\mathcal{J}'} + M_{\mathcal{J},\mathcal{J}'}). \quad (227)$$

In order to show that the overall failure probability is $o(1)$, we will first show that the probability of both **(C1)** and **(C2)** holding for some (ℓ, C, ζ) is suitably small. Observe that for any length- ℓ $\mathcal{J} \subset [p]$, we have $M_{\mathcal{J}} \leq \ell\Delta = \frac{\ell\nu n}{k}$, which implies that $k_{\ell,C,\zeta} = 0$ holds for any $C > e^\nu$. Thus, **(C1)** does not hold when $C > e^\nu$, so from now on we concentrate on a specific $\ell \in [1, \frac{k}{\log k}]$ and a given feasible $(C, \zeta) \in [0, e^\nu] \times [0, 1]$.

Condition for (C1) under the given (ℓ, C, ζ)

We first bound $\mathbb{E}k_{\ell,C,\zeta}$ and then deduce the high-probability behaviour of $k_{\ell,C,\zeta}$ via Markov's inequality. Since $k_{\ell,C,\zeta} = \sum_{\mathcal{J} \subset s, |\mathcal{J}|=\ell} \mathbb{1}(M_{\mathcal{J}} = \frac{Cn\nu e^{-\nu}\ell}{k}, M_{\mathcal{J}0} = \zeta M_{\mathcal{J}})$, we have

$$\mathbb{E}k_{\ell,C,\zeta} = \binom{k}{\ell} \mathbb{P} \left(\text{for fixed } \mathcal{J} \subset s \text{ with } |\mathcal{J}| = \ell, M_{\mathcal{J}} = \frac{Cn\nu e^{-\nu}\ell}{k}, M_{\mathcal{J}0} = \zeta M_{\mathcal{J}} \right) \quad (228)$$

$$= \binom{k}{\ell} \mathbb{P} \left(M_{\mathcal{J}} = \frac{Cn\nu e^{-\nu}\ell}{k} \right) \mathbb{P} \left(\text{Bin} \left(\frac{Cn\nu e^{-\nu}\ell}{k}, \rho \right) = \frac{\zeta \cdot Cn\nu e^{-\nu}\ell}{k} \right) \quad (229)$$

since \mathbf{X}_s (which determines $M_{\mathcal{J}}$) and \mathbf{Z} (which determines $M_{\mathcal{J}0}$ under given $M_{\mathcal{J}}$) are independent.

Bounding $\mathbb{P}(M_{\mathcal{J}} = \frac{Cnve^{-\nu}\ell}{k})$: We make use of Lemma 7 in Appendix G, in which $\mathcal{W}^{(\mathcal{J})}$ is defined to index the tests that contain some item from \mathcal{J} , and its cardinality is defined as $W^{(\mathcal{J})} := |\mathcal{W}^{(\mathcal{J})}|$. We thus have that $i \in \mathcal{M}_{\mathcal{J}}$ holds if and only if $i \in \mathcal{W}^{(\mathcal{J})} \setminus \mathcal{W}^{(s \setminus \mathcal{J})}$.

Recall also the following notation introduced above (171): Given $s_0 \subset s = [k]$, we let \mathbf{T}_{s_0} be an unordered multi-set of length $|s_0|\Delta$ whose entries are in $[n]$, with the overall multi-set representing the $|s_0|\Delta$ placements from the items in s_0 in an unordered manner. It follows that $\mathcal{W}^{(\mathcal{J})}$ and $\mathcal{W}^{(s \setminus \mathcal{J})}$ are determined by the randomness of $\mathbf{T}_{\mathcal{J}}$ and $\mathbf{T}_{s \setminus \mathcal{J}}$, respectively. Given \mathcal{J} satisfying $|\mathcal{J}| = \ell$ with $\frac{\ell}{k} = o(1)$, (299) and (301) in Lemma 7 give that for any $\delta > 0$ we have $W^{(s \setminus \mathcal{J})} = (1 - e^{-\nu} + \delta' \sqrt{\frac{\ell}{k}} + o(1))n$ for some $|\delta'| < \delta$, with probability at least $1 - 2\exp(-2\nu^{-1}\delta^2 n)$. Therefore, we set $\delta = \sqrt{C_* \frac{\nu\ell}{k}}$ (thus $\delta \sqrt{\frac{\ell}{k}} = o(1)$) for some sufficiently large absolute constant C_* to obtain that the event $W^{(s \setminus \mathcal{J})} = (1 - e^{-\nu} + o(1))n$, which implies $n - W^{(s \setminus \mathcal{J})} = ne^{-\nu}(1 + o(1))$, holds with probability at least $1 - 2\exp(-\frac{C_*\ell n}{k})$. Thus, the event

$$\mathcal{A}_{\mathcal{J}} = \left\{ n - W^{(s \setminus \mathcal{J})} = ne^{-\nu}(1 + o(1)) \right\} \quad (230)$$

holds with probability at least $1 - 2\exp(-\frac{C_*\ell n}{k})$, where C_* is a given constant that can be made arbitrarily large. Therefore, we can proceed as

$$\mathbb{P}\left(M_{\mathcal{J}} = \frac{Cnve^{-\nu}\ell}{k}\right) \leq \mathbb{P}\left(M_{\mathcal{J}} = \frac{Cnve^{-\nu}\ell}{k} \middle| \mathcal{A}_{\mathcal{J}}\right) + \mathbb{P}(\mathcal{A}_{\mathcal{J}}^c) \leq \tilde{P}_1 + 2\exp\left(-\frac{C_*\ell n}{k}\right), \quad (231)$$

where we introduce the shorthand $\tilde{P}_1 := \mathbb{P}(M_{\mathcal{J}} = \frac{Cnve^{-\nu}\ell}{k} | \mathcal{A}_{\mathcal{J}})$.

To bound \tilde{P}_1 , we divide the n tests into two parts $\mathcal{W}^{(s \setminus \mathcal{J})}$ and $[n] \setminus \mathcal{W}^{(s \setminus \mathcal{J})}$. Then, conditioning on $\mathbf{T}_{s \setminus \mathcal{J}}$ and using the randomness of $\mathbf{T}_{\mathcal{J}}$, we can identify $M_{\mathcal{J}}$ with the number of tests in $[n] \setminus \mathcal{W}^{(s \setminus \mathcal{J})}$ that contain item from \mathcal{J} . With this perspective, Lemma 8 in Appendix G states that

$$\tilde{P}_1 \leq \exp\left(-\frac{\ell\nu n}{k} \left[D(Ce^{-\nu} \| e^{-\nu}(1 + o(1))) + o(1) \right]\right) \quad (232)$$

$$= \exp\left(-\frac{\ell\nu n}{k} [D(Ce^{-\nu} \| e^{-\nu}) + o(1)]\right). \quad (233)$$

Note that $D(Ce^{-\nu} \| e^{-\nu})$, as a function of $C \in (0, e^{-\nu}]$, is uniformly bounded from above. Hence, we can choose C_* large enough so that the bound on \tilde{P}_1 in (233) dominates $2\exp(-\frac{C_*\ell n}{k})$ in the right-hand side of (231), and thus obtain

$$\mathbb{P}\left(M_{\mathcal{J}} = \frac{Cnve^{-\nu}\ell}{k}\right) \leq \exp\left(-\frac{\ell\nu n}{k} [D(Ce^{-\nu} \| e^{-\nu}) + o(1)]\right). \quad (234)$$

Bounding $\mathbb{P}(\text{Bin}(\frac{Cnve^{-\nu}\ell}{k}, \rho) = \frac{\zeta \cdot Cnve^{-\nu}\ell}{k})$ and combining: We apply the Chernoff bound (see (23) in Lemma 1) to obtain

$$\mathbb{P}\left(\text{Bin}\left(\frac{Cnve^{-\nu}\ell}{k}, \rho\right) = \frac{\zeta \cdot Cnve^{-\nu}\ell}{k}\right) \leq \exp\left(-\frac{Cnve^{-\nu}\ell}{k} \cdot D(\zeta \| \rho)\right). \quad (235)$$

Substituting (234) and (235) into (229) yields

$$\mathbb{E}k_{\ell,C,\zeta} \leq \binom{k}{\ell} \exp \left(- \frac{\ell \nu e^{-\nu}}{k} \underbrace{\left[e^{\nu} D(Ce^{-\nu} \| e^{-\nu}) + C \cdot D(\zeta \| \rho) + o(1) \right]}_{:=f_1(C,\zeta,\rho,\nu)} \right). \quad (236)$$

Note that $f_1(C, \zeta, \rho, \nu)$ here has appeared in the proof of converse bound; see (182).

We separately deal with the two cases $1 \leq \ell \leq \log k$ and $\log k < \ell \leq \frac{k}{\log k}$ using Markov's inequality. Specifically, following exactly the same arguments as the Bernoulli design (see (97)–(104)), we conclude the following for a given (ℓ, C, ζ) :

$$\frac{n \nu e^{-\nu} \ell}{k} [f_1(C, \zeta, \rho, \nu) + o(1)] \geq (1 + \eta_1) \ell \log \frac{k}{\ell}, \text{ for some } \eta_1 > 0 \quad (237)$$

$$\implies \mathbb{P}(k_{\ell,C,\zeta} \geq 1) \leq \hat{P}_\ell, \text{ where } \hat{P}_\ell = \begin{cases} (\ell \log k)^{-5}, & \text{if } 1 \leq \ell \leq \log k \\ k^{-5}, & \text{if } \log k < \ell \leq \frac{k}{\log k} \end{cases}. \quad (238)$$

Therefore, if (237) holds, then **(C1)** holds for some given (ℓ, C, ζ) with probability at most \hat{P}_ℓ .

Condition for **(C2)** under given (ℓ, C, ζ)

As in (170), with respect to the randomness of \mathbf{T}_s , the event

$$\mathcal{A}_1 = \{N_0 = (1 + o(1))e^{-\nu}n\} \quad (239)$$

holds with $1 - o(1)$ probability. In the following, for a generic event \mathcal{E} , an upper bound on $\mathbb{P}(\mathcal{E})$ in this part should be understood as a bound on $\mathbb{P}(\mathcal{E} \cap \mathcal{A}_1)$, but we will only make this explicit in the concluding stage; see (275) below.

We consider a fixed (ℓ, C, ζ) and seek to establish the condition for **(C2)**, i.e., (227) holding for some $\mathcal{J} \in \mathcal{K}_{\ell,C,\zeta}$ and some $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$. Note that $G_{\mathcal{J},\mathcal{J}',1}$ in (227) is the number of positive tests in $\mathcal{N}_{0,\mathcal{J}'} \cup \mathcal{M}_{\mathcal{J},\mathcal{J}'}$, and thus we can further decompose $G_{\mathcal{J},\mathcal{J}',1}$ into

$$G_{\mathcal{J},\mathcal{J}',1} = \tilde{G}_{\mathcal{J}',1} + U_{\mathcal{J},\mathcal{J}',1}, \quad (240)$$

where $\tilde{G}_{\mathcal{J}',1}$ is the number of positive tests in $\mathcal{N}_{0,\mathcal{J}'}$ (with no dependence on \mathcal{J} , as reflected by the notation), and $U_{\mathcal{J},\mathcal{J}',1}$ is the number of positive tests in $\mathcal{M}_{\mathcal{J},\mathcal{J}'}$. Substituting (240) into (227), we can restate **(C2)** as follows:

$$\tilde{G}_{\mathcal{J}',1} \geq \left(\frac{1}{2} - \zeta\right) \frac{C n \nu e^{-\nu} \ell}{k} + \frac{1}{2} N_{0,\mathcal{J}'} + \left(\frac{1}{2} M_{\mathcal{J},\mathcal{J}'} - U_{\mathcal{J},\mathcal{J}',1}\right) \quad (241)$$

for some $\mathcal{J} \in \mathcal{K}_{\ell,C,\zeta}$ and some $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$. Therefore, if **(C2)** holds, then the following

also necessarily holds:

$$\max_{\substack{\mathcal{J}' \subset [p] \setminus s \\ |\mathcal{J}'| = \ell}} \left(\tilde{G}_{\mathcal{J}',1} - \frac{1}{2} N_{0,\mathcal{J}'} \right) \geq \left(\frac{1}{2} - \zeta \right) \frac{Cnve^{-\nu}\ell}{k} - \max_{\substack{\mathcal{J} \in \mathcal{K}_{\ell,C,\zeta} \\ \mathcal{J}' \subset [p] \setminus s, |\mathcal{J}'| = \ell}} M_{\mathcal{J},\mathcal{J}'}, \quad (242)$$

where we express “there exist” (or “for some”) in **(C2)** via the “max” operation, and apply the trivial inequality $\frac{1}{2}M_{\mathcal{J},\mathcal{J}'} - U_{\mathcal{J},\mathcal{J}',1} \geq -M_{\mathcal{J},\mathcal{J}'}$.

The effect of $M_{\mathcal{J},\mathcal{J}'}$: We show that the term $\max_{\mathcal{J},\mathcal{J}'} M_{\mathcal{J},\mathcal{J}'}$ indeed has minimal effect. Given specific $(\mathcal{J}, \mathcal{J}')$, recall that $M_{\mathcal{J},\mathcal{J}'}$ is the number of tests in $\mathcal{M}_{\mathcal{J}}$ that contain some item from \mathcal{J}' ; we now further let $\widetilde{M}_{\mathcal{J},\mathcal{J}'}$ be the number of placements from the items in \mathcal{J}' that fall in the $\frac{Cnve^{-\nu}\ell}{k}$ tests in $\mathcal{M}_{\mathcal{J}}$. Then, note that we always have $M_{\mathcal{J},\mathcal{J}'} \leq \widetilde{M}_{\mathcal{J},\mathcal{J}'}$, and $M_{\mathcal{J},\mathcal{J}'} < \widetilde{M}_{\mathcal{J},\mathcal{J}'}$ happens if some test in $\mathcal{M}_{\mathcal{J}}$ receives more than one placement from the items in \mathcal{J}' . Therefore, we have

$$\max_{\mathcal{J},\mathcal{J}'} M_{\mathcal{J},\mathcal{J}'} \leq \max_{\mathcal{J},\mathcal{J}'} \widetilde{M}_{\mathcal{J},\mathcal{J}'}, \quad (243)$$

where $(\mathcal{J}, \mathcal{J}')$ are implicitly subject to the same constraints as the last term in (242), and similarly in the developments below. We consider fixed $\mathcal{J} \in \mathcal{K}_{\ell,C,\zeta}$ and $\mathcal{J}' \subset [p] \setminus s$ with $|\mathcal{J}'| = \ell$. Since $\mathcal{M}_{\mathcal{J}} = \frac{Cnve^{-\nu}\ell}{k}$ and the test placements are uniform in $[n]$, a placement from items in \mathcal{J}' falls in $\mathcal{M}_{\mathcal{J}}$ with probability $\frac{Cve^{-\nu}\ell}{k}$, thus the randomness of $\mathbf{T}_{\mathcal{J}'}$ gives $\widetilde{M}_{\mathcal{J},\mathcal{J}'} \sim \text{Bin}(\frac{\ell n}{k}, \frac{Cve^{-\nu}\ell}{k})$. Then we can proceed analogously to the calculations in (111)–(115). In particular, our assumption $C \leq e^\nu$ is equivalent to $Ce^{-\nu} \leq 1$, and thus for given $(\mathcal{J}, \mathcal{J}')$ we have

$$\mathbb{P}\left(\widetilde{M}_{\mathcal{J},\mathcal{J}'} \geq \frac{\ell n}{k(\log \frac{k}{\ell})^{1/2}}\right) \leq \mathbb{P}\left(\text{Bin}\left(\frac{\nu \ell n}{k}, \frac{\nu \ell}{k}\right) \geq \frac{\ell n}{k(\log \frac{k}{\ell})^{1/2}}\right) \quad (244)$$

$$\leq \exp\left(-\Omega\left(\sqrt{\log \frac{k}{\ell}} \cdot \ell \log k\right)\right), \quad (245)$$

where (245) follows from the same argument as the Bernoulli design (see (114)), with only the constant factors changing. Moreover, by repeating the union bound over $(\mathcal{J}, \mathcal{J}')$ as in (115), we obtain

$$\mathbb{P}\left(\max_{\mathcal{J},\mathcal{J}'} \widetilde{M}_{\mathcal{J},\mathcal{J}'} \geq \frac{\ell n}{k(\log \frac{k}{\ell})^{1/2}}\right) \leq k^{-10\ell} \quad (246)$$

for sufficiently large k , and thus it holds with probability at least $1 - k^{-10\ell}$ that

$$\max_{\mathcal{J},\mathcal{J}'} M_{\mathcal{J},\mathcal{J}'} \leq \frac{\ell n}{k(\log \frac{k}{\ell})^{1/2}} = o\left(\frac{\ell n}{k}\right). \quad (247)$$

On this high-probability event, the necessary condition for **(C2)** given in (242) can be written as

$$\max_{\substack{\mathcal{J}' \subset [p] \setminus s \\ |\mathcal{J}'| = \ell}} \left(\tilde{G}_{\mathcal{J}',1} - \frac{1}{2} N_{0,\mathcal{J}'} \right) \geq \left(\frac{1}{2} - \zeta - o(1) \right) \frac{Cnve^{-\nu}\ell}{k}. \quad (248)$$

Thus, to bound the probability of **(C2)**, it suffices to bound the probability of its necessary condition (248). We proceed to study (248) by first studying a fixed \mathcal{J}' and then applying a union bound to account for the maximum.

For given \mathcal{J}' , we define

$$d_0 := \frac{C(2\zeta - 1)}{1 - 2\rho} \quad (249)$$

and write

$$\mathbb{P}\left(\tilde{G}_{\mathcal{J}',1} - \frac{1}{2}N_{0,\mathcal{J}'} \geq \left(\frac{1}{2} - \zeta - o(1)\right) \frac{Cn\nu e^{-\nu}\ell}{k}\right) \leq \mathbb{P}\left(N_{0,\mathcal{J}'} \leq \frac{d_0\nu e^{-\nu}\ell n}{k}\right) \quad (250)$$

$$+ \sum_{\substack{d_0 < d \leq e^\nu \\ \frac{d\nu e^{-\nu}\ell n}{k} \in \mathbb{Z}}} \mathbb{P}\left(N_{0,\mathcal{J}'} = \frac{d\nu e^{-\nu}\ell n}{k}\right) \mathbb{P}\left(\tilde{G}_{\mathcal{J}',1} \geq \frac{d + C(1 - 2\zeta)}{2} \frac{\nu e^{-\nu}\ell n}{k} \middle| N_{0,\mathcal{J}'} = \frac{d\nu e^{-\nu}\ell n}{k}\right). \quad (251)$$

Case 1: $d_0 \in [0, 1]$ (i.e., $0 \leq C(2\zeta - 1) \leq 1 - 2\rho$). We will separately bound the probability terms in the right hand side of (250) and in (251).

Bounding the term in (250): We first bound the term in the right-hand side of (250). By definition, $N_{0,\mathcal{J}'}$ is the number of tests in \mathcal{N}_0 that contain some item from \mathcal{J}' , thus on the event of $N_0 = (1 + o(1))ne^{-\nu}$ (as explained in the text below (239)), Lemma 8 gives

$$\mathbb{P}\left(N_{0,\mathcal{J}'} = \frac{d\nu e^{-\nu}\ell n}{k}\right) \leq \exp\left(-\frac{\ell\nu n}{k} [D(de^{-\nu}\|e^{-\nu}) + o(1)]\right) \quad (252)$$

$$\leq \exp\left(-\frac{\ell\nu n}{k} [D(d_0e^{-\nu}\|e^{-\nu}) + o(1)]\right) \quad (253)$$

for any $d \leq d_0$ such that $\frac{d\nu e^{-\nu}\ell n}{k} \in \mathbb{Z}$, where (253) follows from $d_0 \leq 1$ and the monotonicity of $D(de^{-\nu}\|e^{-\nu})$ with respect to d . Therefore, the term in the right-hand side of (250) can be bounded as

$$\mathbb{P}\left(N_{0,\mathcal{J}'} \leq \frac{d_0\nu e^{-\nu}\ell n}{k}\right) = \sum_{0 \leq d \leq d_0: \frac{d\nu e^{-\nu}\ell n}{k} \in \mathbb{Z}} \mathbb{P}\left(N_{0,\mathcal{J}'} = \frac{d\nu e^{-\nu}\ell n}{k}\right) \quad (254)$$

$$\leq O\left(\frac{\ell n}{k}\right) \exp\left(-\frac{\ell\nu n}{k} [D(d_0e^{-\nu}\|e^{-\nu}) + o(1)]\right) \quad (255)$$

$$\leq \exp\left(-\frac{\ell\nu n}{k} [D(d_0e^{-\nu}\|e^{-\nu}) + o(1)]\right), \quad (256)$$

where (254) follows from $d_0 = O(1)$ (see (249)) and hence there are at most $O(\frac{\ell n}{k})$ summands, in (256) the leading factor $O(\frac{\ell n}{k})$ is absorbed into the $o(1)$ term in the exponent.

Bounding the term in (251): Analogously to (252), on the event of $N_0 = (1 + o(1))ne^{-\nu}$, Lemma 8 gives

$$\mathbb{P}\left(N_{0,\mathcal{J}'} = \frac{d\nu e^{-\nu}\ell n}{k}\right) \leq \exp\left(-\frac{\ell\nu n}{k} \cdot [D(de^{-\nu}\|e^{-\nu}) + o(1)]\right). \quad (257)$$

Then, conditioned on $N_{0,\mathcal{J}'} = \frac{d\nu e^{-\nu}\ell n}{k}$ we have $\tilde{G}_{\mathcal{J}',1} \sim \text{Bin}(\frac{d\nu e^{-\nu}\ell n}{k}, \rho)$ via the randomness of \mathbf{Z} , and thus the condition $d > d_0$ (which implies $\frac{d+C(1-2\zeta)}{2} > d\rho$) allows us to apply the Chernoff bound (see (23) in Lemma 1) to obtain

$$\mathbb{P}\left(\text{Bin}\left(\frac{d\nu e^{-\nu}\ell n}{k}, \rho\right) \geq \frac{d+C(1-2\zeta)}{2} \frac{\nu e^{-\nu}\ell n}{k}\right) \leq \exp\left(-\frac{d\nu e^{-\nu}\ell n}{k} \cdot D\left(\frac{1}{2} + \frac{C(1-2\zeta)}{2d} \parallel \rho\right)\right). \quad (258)$$

Combining the bounds in (257) and (258), without accounting for the summation over d , a single summand in (251) corresponding to a specific d can be bounded by

$$\exp\left(-\frac{\ell\nu n e^{-\nu}}{k} \underbrace{\left[e^\nu \cdot D(de^{-\nu} \parallel e^{-\nu}) + d \cdot D\left(\frac{1}{2} + \frac{C(1-2\zeta)}{2d} \parallel \rho\right) + o(1)\right]}_{:=g(C,\zeta,d,\rho,\nu)}\right). \quad (259)$$

Note that $g(C, \zeta, d, \rho, \nu)$ here has appeared in the proof of converse bound; see (209).

To further account for the summation over $\{d_0 < d \leq e^\nu : \frac{d\nu e^{-\nu}\ell n}{k} \in \mathbb{Z}\}$, which contains at most $O(\frac{\ell n}{k})$ elements, we bound all these summands by their common upper bound

$$\exp\left(-\frac{\ell\nu n e^{-\nu}}{k} [g(C, \zeta, d^*, \rho, \nu) + o(1)]\right), \quad (260)$$

where d^* is given by

$$d^* = \arg \min_d g(C, \zeta, d, \rho, \nu), \quad \text{subject to } |C(2\zeta - 1)| < d \leq e^\nu, \quad (261)$$

where we relax the range of d from $\frac{C(2\zeta-1)}{1-2\rho} = d_0 \leq d \leq e^\nu$ to $|C(2\zeta - 1)| \leq d \leq e^\nu$ so that it matches the corresponding development in the converse proof (see (211)). We again write $A := C(1 - 2\zeta)$, and note that d^* in (261) has been pinpointed in (218). Therefore, the term in (251) is bounded by

$$O\left(\frac{\ell n}{k}\right) \exp\left(-\frac{\ell\nu n e^{-\nu}}{k} [g(C, \zeta, d^*, \rho, \nu) + o(1)]\right) = \exp\left(-\frac{\ell\nu n e^{-\nu}}{k} \underbrace{[g(C, \zeta, d^*, \rho, \nu) + o(1)]}_{:=f_2(C,\zeta,\rho,\nu)}\right), \quad (262)$$

where $f_2(C, \zeta, \rho, \nu)$ coincides with the one in the proof of near-constant weight design converse bound; see (221) in Appendix E.

Comparing the bounds in (256) and (262): Observe that (261) and $C(2\zeta - 1) \leq d_0 \leq e^\nu$ give

$$f_2(C, \zeta, \rho, \nu) = g(C, \zeta, d^*, \rho, \nu) \leq g(C, \zeta, d_0, \rho, \nu) = e^\nu D(d_0 e^{-\nu} \parallel e^{-\nu}), \quad (263)$$

where the last equality follows from $D(\frac{1}{2} + \frac{C(1-2\zeta)}{2d} \parallel \rho) = 0$ when $d = d_0$. Thus, the bound in (262) dominates the one in (256), and by substituting into (250)–(251), we obtain

$$\mathbb{P}\left(\tilde{G}_{\mathcal{J}',1} - \frac{1}{2}N_{0,\mathcal{J}'} \geq \left(\frac{1}{2} - \zeta - o(1)\right) \frac{Cn\nu e^{-\nu}\ell}{k}\right) \leq \exp\left(-\frac{\ell\nu n e^{-\nu}}{k} [f_2(C, \zeta, \rho, \nu) + o(1)]\right) \quad (264)$$

for the case $C(2\zeta - 1) \leq 1 - 2\rho$.

Case 2: $d_0 < 0$ (i.e., $C(2\zeta - 1) < 0$). In this case, the term in (250) vanishes, and we focus on bounding the term in (251). As in Case 1, given $N_{0,\mathcal{J}'} = \frac{d\nu e^{-\nu}\ell n}{k}$ we have $\tilde{G}_{\mathcal{J}',1} \sim \text{Bin}(\frac{d\nu e^{-\nu}\ell n}{k}, \rho)$, and so we have the summand being 0 if $\frac{d+C(1-2\zeta)}{2} > d$, i.e., $d < C(1 - 2\zeta)$. Therefore, we can restrict the summation over d in (251) to

$$\mathcal{D} = \left\{ C(1 - 2\zeta) \leq d \leq e^\nu : \frac{d\nu e^{-\nu}\ell n}{k} \in \mathbb{Z} \right\}. \quad (265)$$

Then similarly to Case 1, we bound the $O(\frac{\ell n}{k})$ summands by their common upper bound as per (260) with d^* given in (261), and obtain the same bound as (264).

Case 3: $d_0 > 1$ (i.e., $C(2\zeta - 1) > 1 - 2\rho$). In this case, we simply apply the trivial bound

$$\mathbb{P}\left(\tilde{G}_{\mathcal{J}',1} - \frac{1}{2}N_{0,\mathcal{J}'} \geq \left(\frac{1}{2} - \zeta - o(1)\right) \frac{Cn\nu e^{-\nu}\ell}{k}\right) \leq 1, \quad (266)$$

which in turn trivially behaves as $\exp(-\frac{\ell\nu n}{k}o(1))$.

Combining the Cases 1-3: We recall that $f_2(C, \zeta, \rho, \nu)$ is given in (262) and define

$$\hat{f}_2(C, \zeta, \rho, \nu) = \begin{cases} f_2(C, \zeta, \rho, \nu) & \text{when } C(2\zeta - 1) \leq 1 - 2\rho \\ 0 & \text{when } C(2\zeta - 1) > 1 - 2\rho. \end{cases}, \quad (267)$$

Combining the three cases discussed above, we obtain

$$\mathbb{P}\left(\tilde{G}_{\mathcal{J}',1} - \frac{1}{2}N_{0,\mathcal{J}'} \geq \left(\frac{1}{2} - \zeta - o(1)\right) \frac{Cn\nu e^{-\nu}\ell}{k}\right) \leq \exp\left(-\frac{\ell\nu n e^{-\nu}}{k} [\hat{f}_2(C, \zeta, \rho, \nu) + o(1)]\right). \quad (268)$$

Additionally, we can verify that $f_2(C, \zeta, \rho, \nu) = 0$ when $C(2\zeta - 1) = 1 - 2\rho$: Using (218) and some simple algebra, we find that $d^* = 1$ holds when $A = C(1 - 2\zeta) = 2\rho - 1$; substituting this into $g(C, \zeta, d, \rho, \nu)$ in (259) yields

$$g(C, \zeta, 1, \rho, \nu) = D\left(\frac{1 + C(1 - 2\zeta)}{2} \parallel \rho\right) = 0. \quad (269)$$

Thus, $\hat{f}_2(C, \zeta, \rho, \nu)$ is continuous.

The condition for (C2): We now proceed as follows:

$$\mathbb{P}\left(\text{(C2) holds for } (\ell, C, \zeta)\right) \quad (270)$$

$$\leq \mathbb{P}\left((248) \text{ hold for } (\ell, C, \zeta)\right) + \mathbb{P}\left((247) \text{ does not hold}\right) \quad (271)$$

$$\leq \binom{p}{\ell} \exp\left(-\frac{\ell\nu n e^{-\nu}}{k} [\hat{f}_2(C, \zeta, \rho, \nu) + o(1)]\right) + k^{-10\ell} \quad (272)$$

$$\leq \exp\left((1 + o(1))\ell \log \frac{p}{\ell} - \frac{\ell\nu n e^{-\nu}}{k} [\hat{f}_2(C, \zeta, \rho, \nu) + o(1)]\right) + k^{-10\ell}, \quad (273)$$

where (271) holds because under the condition (247) we have that (248) is a necessary condition for “**(C2)** holds for (ℓ, C, ζ) ”, and in (272) we apply a union bound over no more than $\binom{p}{\ell}$ possibilities of \mathcal{J}' to account for the maximum in (248), and recall that (247) holds with probability at least $1 - k^{-10\ell}$. Therefore, for given (ℓ, C, ζ) we have identified the following condition for **(C2)** to hold (where we now make the role of \mathcal{A}_1 explicit):

$$\begin{aligned} & \frac{\ell \nu n e^{-\nu}}{k} \hat{f}_2(C, \zeta, \rho, \nu) \geq (1 + \eta_2) \ell \log \frac{p}{\ell} \quad \text{for some } \eta_2 > 0 \\ \implies & \mathbb{P}\left(\{(\mathbf{C2}) \text{ holds for } (\ell, C, \zeta)\} \cap \mathcal{A}_1\right) \end{aligned} \quad (274)$$

$$\leq \exp\left(-\frac{\eta_2}{2} \ell \log \frac{p}{\ell}\right) + k^{-10\ell} \leq \hat{P}_\ell := \begin{cases} (\ell \log k)^{-5}, & \text{if } 1 \leq \ell \leq \log k \\ k^{-5}, & \text{if } \log k < \ell \leq \frac{k}{\log k}, \end{cases} \quad (275)$$

since $\exp(-\frac{\eta_2}{2} \ell \log \frac{p}{\ell})$ is an upper bound on (273) when (274) holds.

Establishing the threshold

We are now in a position to derive the threshold for n above which restricted MLE has $o(1)$ probability of failing. We pause to review our previous developments:

- For given ℓ and any feasible (C, ζ) , by (237)–(238), if

$$n \geq \frac{(1 + \eta_1) k \log \frac{k}{\ell}}{\nu e^{-\nu} f_1(C, \zeta, \rho, \nu)} \quad (276)$$

holds for some $\eta_1 > 0$ with $f_1(C, \zeta, \rho, \nu)$ being defined in (236), then **(C1)** holds for the given (ℓ, C, ζ) with probability at most \hat{P}_ℓ .

- For given ℓ and any feasible (C, ζ) , by (274)–(275), if

$$n \geq \frac{(1 + \eta_2) k \log \frac{p}{\ell}}{\nu e^{-\nu} [\hat{f}_2(C, \zeta, \rho, \nu) + o(1)]} \quad (277)$$

holds for some $\eta_2 > 0$ with $\hat{f}_2(C, \zeta, \rho, \nu)$ being given in (267) and (262), then the probability of “**(C2)** holds for the given (ℓ, C, ζ) and \mathcal{A}_1 holds” is no more than \hat{P}_ℓ .

Bounding the failure probability for a fixed ℓ : For fixed $\ell \in [1, \frac{k}{\log k}]$, we first consider the feasible $(C, \zeta) \in [0, e^\nu] \times [0, 1]$ such that $\frac{C e^{-\nu} \ell \nu n}{k}, \frac{\zeta C e^{-\nu} \ell \nu n}{k}$ are integers, and note that there are at most $O((\frac{\ell n}{k})^2) = O((\ell \log k)^2)$ possibilities of feasible pairs of (C, ζ) . If it holds for some $\eta > 0$ that

$$n \geq (1 + \eta) \frac{k}{\nu e^{-\nu}} \max_{\substack{C \in (0, e^\nu) \\ \zeta \in (0, 1)}} \min \left\{ \frac{\log \frac{k}{\ell}}{f_1(C, \zeta, \rho, \nu)}, \frac{\log \frac{p}{\ell}}{\hat{f}_2(C, \zeta, \rho, \nu)} \right\} \quad (278)$$

then the two dot points reviewed above give that, for any feasible $(C, \zeta) \in [0, e^\nu] \times [0, 1]$ (with

respect to the given ℓ) we have

$$\mathbb{P}(\{(\mathbf{C1}) \text{ and } (\mathbf{C2}) \text{ simultaneously hold for } (\ell, C, \zeta)\} \cap \mathcal{A}_1) \leq 2\hat{P}_\ell.$$

Since it is unlikely for $(\mathbf{C1})$ to hold when $C > e^\nu$, for the restricted MLE decoder in (30) we obtain that

$$\begin{aligned} & \mathbb{P}\left(\{|s \setminus \hat{S}'| = \ell\} \cap \mathcal{A}_1\right) \\ & \leq \mathbb{P}\left(\{(\mathbf{C1}) \text{ and } (\mathbf{C2}) \text{ hold for some feasible } (C, \zeta) \in [0, e^\nu] \times [0, 1]\} \cap \mathcal{A}_1 \text{ holds}\right) \end{aligned} \quad (279)$$

$$\leq O((\ell \log k)^2) \hat{P}_\ell, \quad (280)$$

where (279) holds due to Lemma 3(b), and (280) follows from a union bound over all feasible (C, ζ) under the given $\ell \in [1, \frac{k}{\log k}]$.

Bounding the overall failure probability for $\ell \in [1, \frac{k}{\log k}]$: We further take a union bound over all $\ell \in [1, \frac{k}{\log k}]$. We assume that the condition (278) holds for all $\ell \in [1, \frac{k}{\log k}]$; by the same reasoning as (153)–(154), this assumption amounts to

$$n \geq \frac{(1 + \eta)k \log \frac{p}{k}}{(1 - \theta)\nu e^{-\nu}} \frac{1}{\min_{C \in (0, e^\nu), \zeta \in (0, 1)} \max\{\frac{1}{\theta} f_1(C, \zeta, \rho, \nu), \hat{f}_2(C, \zeta, \rho, \nu)\}} \quad (281)$$

for some $\eta > 0$. Then, if (281) holds, by union bound we can bound the overall failure probability as

$$\mathbb{P}(\text{err}) \leq \mathbb{P}(\mathcal{A}_1^c) + \sum_{\ell=1}^{k/\log k} \mathbb{P}\left(\{|s \setminus \hat{S}'| = \ell\} \cap \mathcal{A}_1\right) \quad (282)$$

$$\leq o(1) + \sum_{\ell=1}^{k/\log k} O((\ell \log k)^2) \hat{P}_\ell \quad (283)$$

$$\leq o(1) + \left[\sum_{\ell=1}^{\log k} (\ell \log k)^{-3} + \sum_{\ell=\log k}^{k/\log k} k^{-3} \right] = o(1), \quad (284)$$

where (282) follows from a union bound, and we use (280) in (283) and then substitute \hat{P}_ℓ (given by (275)) in (284). In conclusion, if (281) holds for some $\eta > 0$, then restricted MLE has $o(1)$ probability of failure.

Simplifying $\hat{f}_2(C, \zeta, \rho, \nu)$ to $f_2(C, \zeta, \rho, \nu)$: Recall that $\hat{f}_2(C, \zeta, \rho, \nu)$ is given in (267). In this step, we show that the minimum over (C, ζ) in (281) is not attained in the domain of $C(2\zeta - 1) > 1 - 2\rho$, thus we can safely simplify $\hat{f}_2(C, \zeta, \rho, \nu)$ to $f_2(C, \zeta, \rho, \nu)$ in the threshold (281). To this end, note that $C(2\zeta - 1) > 1 - 2\rho$ is equivalent to

$$\zeta > \zeta' := \frac{1}{2} + \frac{1 - 2\rho}{2C} \geq \frac{1}{2},$$

and we note that $\hat{f}_2(C, \zeta, \rho, \nu) = 0$ holds for any $\zeta \geq \zeta'$ (see (267) and recall that $\hat{f}_2(C, \zeta, \rho, \nu)$ is continuous), and that $f_1(C, \zeta, \rho, \nu)$ is monotonically increasing with respect to ζ in $[\frac{1}{2}, 1)$ (see (236)). Therefore, for any given (C, ζ) with $C(2\zeta - 1) > 1 - 2\rho$ we can also consider (C, ζ') that falls in the case of $C(2\zeta - 1) \leq 1 - 2\rho$, and we can compare the values of $f_1(C, \zeta, \rho, \nu)$ and $\hat{f}_2(C, \zeta, \rho, \nu)$ as

$$f_1(C, \zeta, \rho, \nu) \geq f_1(C, \zeta', \rho, \nu), \quad \hat{f}_2(C, \zeta', \rho, \nu) = \hat{f}_2(C, \zeta, \rho, \nu) = 0,$$

yielding that

$$\max \left\{ \frac{1}{\theta} f_1(C, \zeta, \rho, \nu), \hat{f}_2(C, \zeta, \rho, \nu) \right\} \geq \max \left\{ \frac{1}{\theta} f_1(C, \zeta', \rho, \nu), \hat{f}_2(C, \zeta', \rho, \nu) \right\}. \quad (285)$$

Thus, the minimum over (C, ζ) in (281) is attained at the case $C(2\zeta - 1) \leq 1 - 2\rho$, in which $\hat{f}_2(C, \zeta, \rho, \nu) = f_2(C, \zeta, \rho, \nu)$. Then similarly to (161)–(164) we have

$$\min_{C \in (0, e^\nu), \zeta \in (0, 1)} \max \left\{ \frac{1}{\theta} f_1(C, \zeta, \rho, \nu), \hat{f}_2(C, \zeta, \rho, \nu) \right\} \quad (286)$$

$$\geq \min_{C \in (0, e^\nu), \zeta \in (0, 1)} \max \left\{ \frac{1}{\theta} f_1(C, \zeta, \rho, \nu), f_2(C, \zeta, \rho, \nu) \right\}, \quad (287)$$

implying that to ensure (281) it suffices to have

$$n \geq \frac{(1 + \eta)k \log \frac{p}{k}}{(1 - \theta)\nu e^{-\nu}} \frac{1}{\min_{C \in (0, e^\nu), \zeta \in (0, 1)} \max \left\{ \frac{1}{\theta} f_1(C, \zeta, \rho, \nu), f_2(C, \zeta, \rho, \nu) \right\}}. \quad (288)$$

This establishes the second term in (11).

G Technical Lemmas for Near-Constant Weight Designs

Compared to the Bernoulli design, an additional challenge in the near-constant weight design is that a test may receive multiple placements from a single item. We present two high-probability events that prove useful in overcoming this difficulty.

Lemma 5. ([16, Prop. II.3] and [16, Eq. (40)]) *Under the near-constant design where each item is uniformly placed (with replacement) to $\Delta = \frac{\nu n}{k}$ tests, we have the following:*

(a) *With $1 - o(1)$ probability, the number of defective items appearing in any test more than once scales at most as $O(\log k)$.*

(b) *With probability at least $1 - p^{-7}$, the number of tests assigned to precisely one defective item (referred to as a degree-1 test in [16]) is given by $(1 + o(1))e^{-\nu}k\Delta$.*

As another useful technical result, we present the (anti-)concentration of the hypergeometric distribution. The Chernoff bound (upper bound) can be found in [26], and we expect that the matching lower bound is also known, but we provide a short proof for completeness.

Lemma 6. *If $U \sim \text{Hg}((1+o(1))k\Delta, (1+o(1))e^{-\nu}k\Delta, \Delta)$ with $k \rightarrow \infty$, $\Delta = \frac{\nu n}{k}$, and $n = \Theta(k \log k)$, then for a given $C \in (0, 1)$ such that $C\Delta$ is an integer, we have*

$$\mathbb{P}(U = C\Delta) = \exp(-\Delta[D(C\|e^{-\nu}) + o(1)]). \quad (289)$$

Proof. By letting $k_0 = (1+o(1))k$ and $k_\nu = (1+o(1))e^{-\nu}k$ we can write $U \sim \text{Hg}(k_0\Delta, k_\nu\Delta, \Delta)$, which implies that

$$\mathbb{P}(U = C\Delta) = \frac{\binom{k_\nu\Delta}{C\Delta} \binom{(k_0-k_\nu)\Delta}{(1-C)\Delta}}{\binom{k_0\Delta}{\Delta}}. \quad (290)$$

Given positive integers m and αm for some $\alpha \in (0, 1)$, we have $\frac{\sqrt{\pi}}{2}G \leq \binom{m}{\alpha m} \leq G$ where $G = \frac{\exp(mH_2(\alpha))}{\sqrt{2\pi m\alpha(1-\alpha)}}$ (see Lemma 2), which gives

$$\binom{m}{\alpha m} = \Theta(1) \frac{\exp(mH_2(\alpha))}{\sqrt{m\alpha(1-\alpha)}}.$$

Substituting this into (290), we obtain

$$\begin{aligned} \mathbb{P}(U = C\Delta) &= \Theta(1) \frac{\frac{\exp(k_\nu\Delta H_2(\frac{C}{k_\nu}))}{\sqrt{C\Delta(1-\frac{C}{k_\nu})}} \cdot \frac{\exp((k_0-k_\nu)\Delta H_2(\frac{1-C}{k_0-k_\nu}))}{\sqrt{(1-C)\Delta(1-\frac{1-C}{k_0-k_\nu})}}}{\frac{\exp(k_0\Delta H_2(\frac{1}{k_0}))}{\sqrt{\Delta(1-\frac{1}{k_0})}}} \\ &= \frac{\Theta(1)}{\sqrt{C(1-C)\Delta}} \exp\left(\underbrace{\Delta \left[k_\nu H_2\left(\frac{C}{k_\nu}\right) + (k_0 - k_\nu) H_2\left(\frac{1-C}{k_0-k_\nu}\right) - k_0 H_2\left(\frac{1}{k_0}\right) \right]}_{:=\mathcal{T}}\right), \end{aligned} \quad (291)$$

where (292) follows from $1 - \frac{C}{k_\nu}$, $1 - \frac{1-C}{k_0-k_\nu}$, and $1 - \frac{1}{k_0}$ all behaving as $\Theta(1)$. Now we seek to further simplify \mathcal{T} . Note that $\frac{C}{k_\nu}$, $\frac{1-C}{k_0-k_\nu}$, and $\frac{1}{k_0}$ all scale as $o(1)$, and it holds for any $t = o(1)$ that

$$H_2(t) = -t \log t - (1-t) \log(1-t) = -t(\log t - 1 + o(1)). \quad (293)$$

This yields

$$\mathcal{T} = k_\nu \left(-\frac{C}{k_\nu} \left[\log\left(\frac{C}{k_\nu}\right) - 1 + o(1) \right] \right) - (k_0 - k_\nu) \cdot \frac{1-C}{k_0-k_\nu} \cdot \left[\log\left(\frac{1-C}{k_0-k_\nu}\right) - 1 + o(1) \right] \quad (294)$$

$$+ k_0 \cdot \frac{1}{k_0} \cdot \left[\log\left(\frac{1}{k_0}\right) - 1 + o(1) \right] \quad (295)$$

$$= -C \log\left(\frac{C}{k_\nu}\right) - (1-C) \log\left(\frac{1-C}{k_0-k_\nu}\right) + \log\left(\frac{1}{k_0}\right) + o(1) \quad (296)$$

$$= -C \log\left(\frac{C}{e^{-\nu}}\right) - (1-C) \log\left(\frac{1-C}{1-e^{-\nu}}\right) + o(1) = -D(C\|e^{-\nu}) + o(1), \quad (297)$$

where in (297) we substitute $k_0 = (1+o(1))k$ and $k_\nu = (1+o(1))e^{-\nu}k$. To complete the proof,

we substitute (297) into (292), and then note that the leading factor $\frac{\Theta(1)}{\sqrt{C(1-C)\Delta}}$ with fixed $C \in [\Delta^{-1}, 1 - \Delta^{-1}]$ (this follows from $C \in (0, 1)$ and hence $1 \leq C\Delta \leq \Delta - 1$) can be written as $\exp(-o(\Delta))$ since $\Delta = \frac{\nu n}{k} = \Theta(\log k) \rightarrow \infty$. \square

Next, we present a useful concentration bound analogously to [29, Lem. 1]. We consider $\mathcal{J} \subset [p]$ with $|\mathcal{J}| = \ell$ and use $\mathcal{W}^{(\mathcal{J})}$ to index the tests that contain some item from \mathcal{J} , with cardinality given by $W^{(\mathcal{J})} := |\mathcal{W}^{(\mathcal{J})}|$.

Lemma 7. (Concentration of $W^{(\mathcal{J})}$) *Given $\mathcal{J} \subset [p]$ with $|\mathcal{J}| = \ell$, for any $\varepsilon > 0$ we have*

$$\mathbb{P}\left(|W^{(\mathcal{J})} - \mathbb{E}W^{(\mathcal{J})}| \geq \varepsilon\right) \leq 2 \exp\left(-\frac{2\varepsilon^2 k}{\nu \ell n}\right), \quad (298)$$

where the expectation satisfies

$$\mathbb{E}W^{(\mathcal{J})} = n \left(1 - \exp\left(-\frac{\ell \nu (1 + o(1))}{k}\right)\right) \quad (299)$$

$$= \begin{cases} (1 - e^{-\nu \alpha} + o(1))n, & \text{if } \frac{\ell}{k} \rightarrow \alpha \in (0, 1] \\ \frac{n \nu \ell}{k} (1 + o(1)), & \text{if } \frac{\ell}{k} \rightarrow 0. \end{cases} \quad (300)$$

Moreover, given any $\delta > 0$, with probability at least $1 - 2 \exp(-2\nu^{-1}\delta^2 n)$, we have

$$W^{(\mathcal{J})} = \mathbb{E}W^{(\mathcal{J})} + \delta' n \sqrt{\frac{\ell}{k}}, \text{ for some } |\delta'| < \delta. \quad (301)$$

Proof. We first prove (299). We can write $W^{(\mathcal{J})} = \sum_{i=1}^n \mathbb{1}(\text{test } i \text{ contains some item from } \mathcal{J})$, which leads to

$$\mathbb{E}W^{(\mathcal{J})} = n \cdot \mathbb{P}(\text{test } i \text{ contains some item from } \mathcal{J}) \quad (302)$$

$$= n \left[1 - \left(1 - \frac{1}{n}\right)^{\ell \Delta}\right] \quad (303)$$

$$= n \left[1 - \exp\left(\frac{\ell \nu n}{k} \log\left(1 - \frac{1}{n}\right)\right)\right] \quad (304)$$

$$= n \left[1 - \exp\left(-\frac{\ell \nu (1 + o(1))}{k}\right)\right], \quad (305)$$

where we use $\log(1 - \frac{1}{n}) = -\frac{1}{n}(1 + o(1))$ in (305). Thus, (299) follows, and to obtain (300) we substitute $\frac{\ell}{k} = \alpha + o(1)$ if $\frac{\ell}{k} \rightarrow \alpha \in (0, 1]$, or $1 - \exp(-\beta) = \beta(1 + o(1))$ as $\beta \rightarrow 0$ if $\frac{\ell}{k} = o(1)$.

Next, we establish (298). Because the $\ell \Delta$ placements of items in \mathcal{J} are independent and changing one placement changes the quantity $W^{(\mathcal{J})}$ by at most one, McDiarmid's inequality [34] gives

$$\mathbb{P}(|W^{(\mathcal{J})} - \mathbb{E}W^{(\mathcal{J})}| \geq \varepsilon) \leq 2 \exp\left(-\frac{2\varepsilon^2}{\ell \Delta}\right)$$

for any $\varepsilon > 0$, which yields (298) by substituting $\ell \Delta = \frac{\nu \ell n}{k}$. Finally, we show (301) by using (298).

To this end, we take $\varepsilon = \delta n \sqrt{\frac{\ell}{k}}$ in (298), and note that the event in (301) holds true if and only if $|W^{(\mathcal{J})} - \mathbb{E}W^{(\mathcal{J})}| < \delta n \sqrt{\frac{\ell}{k}}$. \square

Finally, the following lemma characterizes the probability mass function for the random variable indicating how many tests from a certain subset of $[n]$ contain items from a certain subset of $[p]$.

Lemma 8. *Suppose that the n tests are divided into two parts consisting of n_1 tests and n_2 tests, with the first part and the second part respectively indexed by $\mathcal{N}_1 \subset [n]$ and $\mathcal{N}_2 \subset [n]$, where $|\mathcal{N}_1| = n_1$, $|\mathcal{N}_2| = n_2$, and $n_1 + n_2 = n$. Given $\mathcal{J} \subset [p]$ with $|\mathcal{J}| = \ell$, consider the $\ell\Delta = \frac{\ell\nu n}{k}$ placements of the ℓ items in \mathcal{J} , and let M be the number of tests in \mathcal{N}_2 that contain some item from \mathcal{J} . Then for any $0 \leq m \leq \ell\Delta$, we have*

$$\mathbb{P}(M = m) \leq \exp \left(-\ell\Delta \left[D \left(\frac{m}{\ell\Delta} \parallel \frac{n_2}{n} \right) + \frac{\ell\Delta}{4n_1} \right] \right), \quad (306)$$

where the probability is with respect to the randomness in placing each item from \mathcal{J} in $\Delta = \frac{\nu n}{k}$ tests chosen uniformly at random with replacement.

Proof. Our argument builds on [29, Sec. IV-C], which characterizes a similar distribution (for different purposes) under the near-constant weight design. Our Lemma 8 generalizes the one therein in two ways. First, we consider the placement of ℓ items, while [29, Lemma 4] only concerns the case that $\ell = 1$. Second, our statement is more general since we leave \mathcal{N}_1 and \mathcal{N}_2 unspecified, while [29, Lemma 4] concerns a more specific choice. As was done in [29], we make use of the *Stirling numbers of the second kind*, defined as $\left\{ \begin{smallmatrix} n \\ k \end{smallmatrix} \right\} := \frac{1}{k!} \sum_{j=0}^k (-1)^{k-j} \binom{k}{j} j^n$, e.g., see [31, P. 204].

The probability of the event that exactly u placements fall in \mathcal{N}_1 is $\mathbb{P}(\text{Bin}(\ell\Delta, \frac{n_1}{n}) = u)$. Conditioning on this event, we need to calculate the probability that the remaining $\ell\Delta - u$ placements are at m distinct tests in \mathcal{N}_2 . Evidently, we require $\ell\Delta - u \geq m$, or equivalently $u \leq \ell\Delta - m$, in order to have $M = m$. Note that there are overall $n_2^{\ell\Delta - u}$ possibilities for performing $\ell\Delta - u$ placements in n_2 tests. By a standard counting argument, we can choose these m distinct tests with $\binom{n_2}{m}$ ways, then $\left\{ \begin{smallmatrix} \ell\Delta - u \\ m \end{smallmatrix} \right\}$ ways of arranging the $\ell\Delta - u$ placements into these m unlabelled bins (which represent the m tests) such that none of them are empty, and finally $m!$ different labellings of the bins. These findings give

$$\mathbb{P}(M = m) = \sum_{u=0}^{\ell\Delta - m} \mathbb{P}(\text{Bin}(\ell\Delta, \frac{n_1}{n}) = u) \cdot \mathbb{P}(\ell\Delta - u \text{ placements occupy } m \text{ tests in } \mathcal{N}_2) \quad (307)$$

$$= \sum_{u=0}^{\ell\Delta - m} \binom{\ell\Delta}{u} \left(\frac{n_1}{n} \right)^u \left(\frac{n_2}{n} \right)^{\ell\Delta - u} \binom{n_2}{m} \cdot \frac{\left\{ \begin{smallmatrix} \ell\Delta - u \\ m \end{smallmatrix} \right\} m!}{n_2^{\ell\Delta - u}} \quad (308)$$

$$= \frac{\binom{n_2}{m} m!}{n^{\ell\Delta}} \cdot \underbrace{\sum_{u=0}^{\ell\Delta - m} \binom{\ell\Delta}{u} n_1^u \left\{ \begin{smallmatrix} \ell\Delta - u \\ m \end{smallmatrix} \right\}}_{:=\Phi}. \quad (309)$$

We bound Φ as follows:

$$\Phi = \sum_{t=0}^{\ell\Delta-m} \binom{\ell\Delta}{m+t} n_1^{\ell\Delta-m-t} \left\{ \begin{matrix} m+t \\ m \end{matrix} \right\} \quad (310)$$

$$\leq \sum_{t=0}^{\ell\Delta-m} \binom{\ell\Delta}{m+t} \binom{t+m}{m} m^t n_1^{\ell\Delta-m-t} \quad (311)$$

$$= \binom{\ell\Delta}{m} n_1^{\ell\Delta-m} \sum_{t=0}^{\ell\Delta-m} \binom{\ell\Delta-m}{t} \left(\frac{m}{n_1}\right)^t \quad (312)$$

$$= \binom{\ell\Delta}{m} n_1^{\ell\Delta-m} \left(1 + \frac{m}{n_1}\right)^{\ell\Delta-m} \quad (313)$$

$$:= C \binom{\ell\Delta}{m} n_1^{\ell\Delta-m}, \quad (314)$$

where we use a change of the variable $t = \ell\Delta - m - u$ in (310), $\left\{ \begin{matrix} t+m \\ m \end{matrix} \right\} \leq \binom{t+m}{m} m^t$ [36] in (311), $\binom{\ell\Delta}{m+t} \binom{t+m}{m} = \binom{\ell\Delta}{m} \binom{\ell\Delta-m}{t}$ in (312), and $\sum_{t=0}^{\ell\Delta-m} \binom{\ell\Delta-m}{t} \left(\frac{m}{n_1}\right)^t = \left(1 + \frac{m}{n_1}\right)^{\ell\Delta-m}$ in (313); finally, in (314), $C := \left(1 + \frac{m}{n_1}\right)^{\ell\Delta-m}$ is bounded by

$$C = \exp\left([\ell\Delta - m] \log\left(1 + \frac{m}{n_1}\right)\right) \leq \exp\left(\frac{m[\ell\Delta - m]}{n_1}\right) \leq \exp\left(\frac{\ell^2 \Delta^2}{4n_1}\right). \quad (315)$$

Substituting (310) back into (307), we obtain

$$\mathbb{P}(M = m) \leq C \binom{\ell\Delta}{m} n_1^{\ell\Delta-m} \frac{\binom{n_2}{m} m!}{n^{\ell\Delta}} \quad (316)$$

$$\leq C \binom{\ell\Delta}{m} \frac{n_1^{\ell\Delta-m} n_2^m}{n^{\ell\Delta}} \quad (317)$$

$$= \exp\left(\frac{\ell^2 \Delta^2}{4n_1}\right) \mathbb{P}\left(\text{Bin}(\ell\Delta, \frac{n_2}{n}) = m\right) \quad (318)$$

$$\leq \exp\left(\frac{\ell^2 \Delta^2}{4n_1} - \ell\Delta \cdot D\left(\frac{m}{\ell\Delta} \parallel \frac{n_2}{n}\right)\right), \quad (319)$$

where we use $\binom{n_2}{m} m! \leq n_2^m$ in (317), (315) in (318), and the Chernoff bound (see (23) in Lemma 1) in (319). This completes the proof. \square

H High- ℓ Achievability Analysis for Both Designs

H.1 Initial Results

With the information-theoretic tools from Section 3.1 in place, we proceed to state an initial non-asymptotic bound of [39] that we use as a starting point. Recalling that we already split the analysis into $\ell \leq \frac{k}{\log k}$ and $\ell > \frac{k}{\log k}$ (and here we consider the latter), we define $\ell_{\min} = 1 + \lfloor \frac{k}{\log k} \rfloor$ for notational convenience.

From [39, Thm. 4],⁷ we know that by a suitable choice of $\{\gamma_\ell\}_{\ell=\ell_{\min}}^k$, the error probability is upper bounded as follows for any $\delta_1 > 0$:

$$\mathbb{P}(\text{err}) \leq \mathbb{P}\left[\bigcup_{(s_{\text{dif}}, s_{\text{eq}}) : |s_{\text{dif}}| \geq \ell_{\min}} \left\{ i^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}) \leq \log \binom{p-k}{|s_{\text{dif}}|} + \log \left(\frac{k}{\delta_1} \binom{k}{|s_{\text{dif}}|} \right) \right\} \right] + \delta_1. \quad (320)$$

As a result, for any $\delta_2 \in (0, 1)$, if the number of tests n is large enough to ensure that

$$\max_{\ell=\ell_{\min}, \dots, k} \frac{\log \binom{p-k}{\ell} + \log \left(\frac{k}{\delta_1} \binom{k}{\ell} \right)}{I_\ell^n (1 - \delta_2)} \leq 1, \quad (321)$$

and if each information density satisfies a (one-sided) concentration bound of the form

$$\mathbb{P}[i^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}) \leq (1 - \delta_2) I_\ell^n] \leq \psi_\ell(n, \delta_2), \quad (322)$$

for some functions $\{\psi_\ell\}_{\ell=1}^k$, then we have

$$\mathbb{P}(\text{err}) \leq \sum_{\ell=\ell_{\min}}^k \binom{k}{\ell} \psi_\ell(n, \delta_2) + \delta_1. \quad (323)$$

Note that while [39] focuses on the Bernoulli design, the analysis up to this point does not use any specific properties of that design except symmetry with respect to re-ordering items, so it remains valid for the near-constant weight design. The only minor difference above compared to [39] is that we use the full mutual information I_ℓ^n instead of simplifying it to nI_ℓ for some single-test counterpart I_ℓ (which can be done for the Bernoulli design but not in general).

We can take $\delta_1 \rightarrow 0$ sufficiently slowly to have an asymptotically negligible impact in (321), so the key step in ensuring $\mathbb{P}(\text{err}) \rightarrow 0$ is showing that the summation in (323) approaches 0. We will take δ_2 to be a constant arbitrarily close to zero, as was done in [39],⁸ and we will show that the condition (321) simplifies to a requirement on n corresponding to the first term in (4).

Naturally, to simplify (321) we need to characterize the behavior of I_ℓ^n . For the Bernoulli design, the following asymptotic characterization was given in [39].

Lemma 9. (Mutual Information for the Bernoulli Design, [39, Prop. 5]) *For the noisy group testing problem under Bernoulli design with i.i.d. Bernoulli($\frac{\nu}{k}$) entries, consider arbitrary sequences of $k \rightarrow \infty$ and $\ell \in \{1, \dots, k\}$ (both indexed by p). If $\frac{\ell}{k} = o(1)$, then*

$$I_\ell^n = \left(\frac{n\nu e^{-\nu} \ell}{k} (1 - 2\rho) \log \frac{1 - \rho}{\rho} \right) (1 + o(1)). \quad (324)$$

⁷More precisely, this is the adapted version for approximate recovery in [39, App. D], because as noted in Appendix A, error events for incorrect sets within distance $\frac{k}{\log k}$ of the true defective set are considered separately (in our analysis of the low- ℓ regime) so are not considered here.

⁸For low ℓ values it should instead be taken as a parameter to be optimized [39], but we are only using these techniques for the high- ℓ regime.

Moreover, if $\frac{\ell}{k} \rightarrow \alpha \in (0, 1]$, then

$$I_\ell^n = ne^{-(1-\alpha)\nu} (H_2(e^{-\alpha\nu} \star \rho) - H_2(\rho)) (1 + o(1)). \quad (325)$$

We will show that the same result holds for the near-constant weight design in Lemma 10 below.

H.2 Bernoulli Design

As we have mentioned earlier, the high- ℓ analysis of the Bernoulli design comes directly from [39] with only minor changes. We thus only provide a brief outline for the unfamiliar reader, citing the relevant results therein.

The following concentration term was shown to be valid (i.e., (322) holds) in [39, Prop. 3]:

$$\psi_\ell(n, \delta_2) = 2 \exp \left(- \frac{n(\delta_2 I_\ell)^2}{4(8 + \delta_2 I_\ell)} \right), \quad (326)$$

where $I_\ell = I_\ell^n/n$ is the mutual information associated with a single test. Under this choice, it is shown in [39, Prop. 7(ii)] that $\sum_{\ell > \frac{k}{\log k}} \binom{k}{\ell} \psi_\ell(n, \delta_2) \rightarrow 0$ as long as $n = \Omega(k \log \frac{p}{k})$ (regardless of the implied constant). Moreover, in [39, Sec. C.2] it is shown for $\nu = \log 2$ that the maximum in (321) is asymptotically attained by $\ell = k$, and that the limiting behavior exactly corresponds to the first term in (4).

A slight additional requirement in our work is that we need to handle general $\nu > 0$, rather than only $\nu = \log 2$, with the non-trivial step being to show that the maximum in (321) is still asymptotically attained by $\ell = k$.⁹ Fortunately, this follows from another well-known result in the literature [32, Thm. 3] (see also [45, Lemma 2]) that we now proceed to outline.

First note that for $\ell > \frac{k}{\log k}$ and $k = o(p)$, the second term in the numerator of (321) is asymptotically negligible compared to the first. We further have $\log \binom{p-k}{\ell} = (\ell \log \frac{p}{\ell})(1 + o(1)) = (\ell \log \frac{p}{k})(1 + o(1))$, so the maximization over ℓ simplifies to maximizing $\frac{\ell}{I_\ell^n}$, or equivalently minimizing $\frac{I_\ell^n}{\ell}$. The above-mentioned results in [32, Thm. 3] and [45, Lemma 2] directly state that $\ell = k$ minimizes $\frac{I_\ell^n}{\ell}$ (for a broader class of noise models that includes ours), as desired.

Having established that $\sum_{\ell > \frac{k}{\log k}} \binom{k}{\ell} \psi_\ell(n, \delta_2) \rightarrow 0$ and that $\ell = k$ asymptotically attains the maximum in (321), the desired threshold on n follows by writing the numerator therein as $(k \log \frac{p}{k})(1 + o(1))$ and the denominator as $n(H_2(e^{-\nu} \star \rho) - H_2(\rho))(1 + o(1)) \times (1 - \delta_2)$ (see (325) with $\alpha = 1$), and recalling that δ_2 can be arbitrarily small. This establishes the first term in (4).

H.3 Near-Constant Weight Design

For the near-constant weight design, we again adopt the choices of δ_1 and δ_2 described above for the Bernoulli design. Notice that the study of the condition (321) has no direct relation to the choice of ψ_ℓ . Thus, to simplify (321), we can apply the exact same analysis as the Bernoulli case as long

⁹The proof of $\sum_{\ell > \frac{k}{\log k}} \binom{k}{\ell} \psi_\ell(n, \delta_2) \rightarrow 0$ is not impacted by changing ν , since doing so only amounts to changing constant factors in the exponent in (326).

as we can show that the mutual information terms I_ℓ^n have the same asymptotic behavior; this is done in the following lemma, whose proof is given in Appendix J.2.

Lemma 10. (Mutual Information, Near-Constant Weight Design) *For the noisy group testing problem under the near-constant weight design with $\Delta = \frac{\nu n}{k}$, consider sequences of sparsity levels $k \rightarrow \infty$ and $\ell \in \{1, \dots, k\}$ (both indexed by p). If $\frac{\ell}{k} = o(1)$, then*

$$I_\ell^n = \left(\frac{n\nu e^{-\nu} \ell}{k} (1 - 2\rho) \log \frac{1 - \rho}{\rho} \right) (1 + o(1)). \quad (327)$$

Moreover, if $\frac{\ell}{k} \rightarrow \alpha \in (0, 1]$, then

$$I_\ell^n = ne^{-(1-\alpha)\nu} (H_2(e^{-\alpha\nu} \star \rho) - H_2(\rho)) (1 + o(1)). \quad (328)$$

From this lemma and the above discussion, we deduce that (321) again leads to the first term in (16). It only remains to show that $\sum_{\ell > \frac{k}{\log k}} \binom{k}{\ell} \psi_\ell(n, \delta_2) \rightarrow 0$.

While we cannot directly use (326), the proof of [39, Prop. 7(ii)] reveals that we do not precisely need (326), but rather, the term inside the exponent can be scaled by any finite constant factor. Since $I_\ell = O(1)$ and $\delta_2 = \Theta(1)$, the denominator in (326) is dominated by the constant term (to within a constant factor), and thus, what we seek is a tail bound of the form $e^{-\Theta(nI_\ell^2)} = e^{-\Theta((I_\ell^2)/n)}$. We provide this in the following lemma, whose proof is given in Appendix J.1.

Lemma 11. (Concentration Bound for $\frac{\ell}{k} \rightarrow \alpha$, Near-Constant Weight Design) *For the noisy group testing problem under near-constant design, consider sequences $k \rightarrow \infty$ and $\ell \geq 1$ (indexed by p) such that $\frac{\ell}{k} \rightarrow \alpha \in [0, 1]$. For any given δ smaller than a quantity depending on (α, ν, ρ) , the following holds for sufficiently large p and any $s_{\text{dif}} \subset S$ with cardinality ℓ :*

$$\mathbb{P}\left(i^n(\mathbf{Y}|\mathbf{X}_{s_{\text{eq}}}, \mathbf{X}_{s_{\text{dif}}}) \leq (1 - \delta)I_\ell^n\right) \leq \exp\left(-\frac{C_{\rho, \nu, \alpha}(I_\ell^n)^2 \delta^2}{n}\right), \quad (329)$$

where $C_{\rho, \nu, \alpha} > 0$ is a constant depending on (ρ, ν, α) , and I_ℓ^n satisfies (327) when $\alpha = 0$, or (328) when $\alpha \in (0, 1]$.

The proof is given in Appendix J.1. With this result in place, the rest of the analysis again follows that of the Bernoulli design to establish the first term in (11).

I High- ℓ Converse Analysis for Both Designs

Recall that the thresholds in (4) and (11) contain two terms; this appendix concerns only the first, which corresponds to the high- ℓ regime.

I.1 Bernoulli Design

For the Bernoulli design, the first term in (4) was already given in [39, Thm. 7 & App. C], and mainly comes down to studying the lower tail probability of the information density $i^n(\mathbf{X}_s, \mathbf{Y})$

(i.e., (18) with $s_{\text{eq}} = \emptyset$, corresponding to $\ell = k$). Here establishing concentration is relatively straightforward due to the i.i.d. nature of the Bernoulli design. We omit any further details, since (i) they are already known from [39], and (ii) for the near-constant weight design (analyzed below) we will follow a similar structure and we will give more detail for it.

I.2 Near-Constant Weight Design

Unlike the Bernoulli design, the first term in (11) does not readily follow from [39], since their analysis was specific to the Bernoulli design. However, it turns out that we can get the desired result by combining certain tools from [39] with a fairly simple analysis of the information density.

We start with the following non-asymptotic bound from [39, Thm. 5], which only relies on the design being symmetric with respect to re-ordering items:

$$\mathbb{P}(\text{err}) \geq \mathbb{P}\left(i^n(\mathbf{X}_s, \mathbf{Y}) \leq \log\left(\delta_1 \binom{p}{k}\right)\right) - \delta_1 \quad (330)$$

for any $\delta_1 > 0$, where $s = [k]$ (without loss of generality), and $i^n(\mathbf{X}_s, \mathbf{Y}) = \log \frac{\mathbb{P}(\mathbf{Y}|\mathbf{X}_s)}{\mathbb{P}(\mathbf{Y})}$ is a shorthand for the information density (320) with $s_{\text{eq}} = \emptyset$ (and $s_{\text{dif}} = [k]$). Hence, its mean is I_k^n (i.e., I_ℓ^n from (19) with $\ell = k$), which behaves as $n(H_2(e^{-\nu} \star \rho) - H_2(\rho))(1 + o(1))$ by Lemma 10.

The main step of the proof is to show the following one-sided concentration bound for i^n with arbitrarily small $\delta_2 > 0$:

$$\mathbb{P}(i^n(\mathbf{X}_s, \mathbf{Y}) \leq (1 + \delta_2)I_k^n) \rightarrow 1. \quad (331)$$

To do so, write $i^n(\mathbf{X}_s, \mathbf{Y}) = \log \mathbb{P}(\mathbf{Y}|\mathbf{X}_s) - \log \mathbb{P}(\mathbf{Y})$ and handle the two terms separately. For the first term, we note that given \mathbf{X}_s , the sequence \mathbf{Y} has independent entries depending on the Bernoulli(ρ) noise, from which a simple concentration argument (e.g., Hoeffding's inequality) gives $\log \mathbb{P}(\mathbf{Y}|\mathbf{X}_s) = -nH_2(\rho)(1 + o(1))$ with probability approaching one. For the second term, we decompose $\mathbb{P}(\mathbf{Y}) = \mathbb{P}(V)\mathbb{P}(\mathbf{Y}|V)$, where V is the number of 1s in \mathbf{Y} . We will utilize two useful lemmas regarding V , the first of which is the following concentration bound.

Lemma 12. (Concentration of V) *Under the preceding setup and notation, we have with probability approaching one (as $p \rightarrow \infty$) that $V = (e^{-\nu} \star \rho + o(1))n$.*

Proof. See (343) in the proof of Lemma 11 below (Appendix J). Note that (343) considers general $(s_{\text{eq}}, s_{\text{dif}})$ and defines V to be the number of 1s in tests that contain no items from s_{eq} , but this reduces to the above definition when $s_{\text{eq}} = \emptyset$ (and thus $\alpha = 1$ in (343)). \square

Now observe that for any fixed v satisfying the high-probability condition $v = (e^{-\nu} \star \rho + o(1))n$, and any corresponding \mathbf{y} whose number of 1s is v , we have by symmetry that $\mathbb{P}(\mathbf{y}|v) = \frac{1}{\binom{n}{v}}$, which in turn behaves as $e^{n(H_2(e^{-\nu} \star \rho) + o(1))}$.

Combining the preceding two findings with the fact that $I_k^n = n(H_2(e^{-\nu} \star \rho) - H_2(\rho))(1 + o(1))$ (Lemma 10), we see that in order to establish (331), it only remains to show that the impact of the remaining term $\log P(V)$ is asymptotically negligible; namely, it should scale as $o(n)$ to be

insignificant compared to the leading $\Theta(n)$ terms. This is stated in the following lemma, which is proved in Appendix J.

Lemma 13. (Behavior of $P(V)$) *Under the preceding setup and notation, we have with probability approaching one (as $p \rightarrow \infty$) that $P(V) = e^{-o(n)}$.*

Intuitively, this result holds because V is produced via binomial random variables (due to the independent noise variables); it can readily be deduced from anti-concentration bounds (Lemma 1) that for $U \sim \text{Bin}(N, q)$ with $N \rightarrow \infty$ and constant $q \in (0, 1)$, we have that $P(U = u) \geq e^{-o(N)}$ whenever $u = Nq(1 + o(1))$ (which holds with high probability). Since the full details are slightly more technical without being much more insightful than this intuition, we defer the formal proof of Lemma 13 to Appendix J.3.

Combining the above findings, we have established that (331) holds for arbitrarily small δ_2 . Moreover, in (330), we can take δ_1 to decay with p sufficiently slowly such that $\log(\delta_1(\frac{p}{k})) = (k \log \frac{p}{k})(1 + o(1))$ (i.e., the impact of δ_1 is asymptotically negligible), and comparing this $(k \log \frac{p}{k})(1 + o(1))$ term with the term $I_k^n = n(H_2(e^{-\nu} \star \rho) - H_2(\rho))(1 + o(1))$ in (331), we deduce the desired threshold on n given by the first term in (11).

J Proofs of Technical Results in the High- ℓ Analysis for the Near-Constant Weight Design

Recall that each item is placed in $\Delta := \frac{\nu n}{k}$ tests for some $\nu = \Theta(1)$, and that the information density $i^n(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}})$ is defined in (18). Throughout much of the analysis in this appendix, it will be useful to condition on $\mathbf{X}_{s_{\text{eq}}}$ (or its specific realization $\mathbf{x}_{s_{\text{eq}}}$ depending on the context), and we accordingly adopt the shorthand $P(\cdot) = \mathbb{P}(\cdot | \mathbf{X}_{s_{\text{eq}}})$ (or $P(\cdot) = \mathbb{P}(\cdot | \mathbf{x}_{s_{\text{eq}}})$).

Given $\mathbf{X}_{s_{\text{eq}}}$, we divide \mathbf{Y} into \mathbf{Y}_1 and \mathbf{Y}_2 , where \mathbf{Y}_1 represents the results of tests corresponding to $X_{s_{\text{eq}}}^{(i)} \neq 0$, and \mathbf{Y}_2 represents the remaining results corresponding to $X_{s_{\text{eq}}}^{(i)} = 0$. Due to the “OR” operation in the group testing model, \mathbf{Y}_1 is completely determined by the noise in its tests, which implies that \mathbf{Y}_1 and \mathbf{Y}_2 are conditionally independent given $\mathbf{X}_{s_{\text{eq}}}$. Therefore, we have

$$i^n(\mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}, \mathbf{X}_{s_{\text{dif}}}) = \log \frac{\mathbb{P}(\mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}, \mathbf{X}_{s_{\text{dif}}})}{\mathbb{P}(\mathbf{Y} | \mathbf{X}_{s_{\text{eq}}})} \quad (332)$$

$$= \log \frac{P(\mathbf{Y}_1 | \mathbf{X}_{s_{\text{dif}}}) P(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{dif}}})}{P(\mathbf{Y}_1) P(\mathbf{Y}_2)} \quad (333)$$

$$= \log \frac{P(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{dif}}})}{P(\mathbf{Y}_2)}, \quad (334)$$

where (334) follows from $P(\mathbf{Y}_1 | \mathbf{X}_{s_{\text{dif}}}) = P(\mathbf{Y}_1)$ due to the fact that \mathbf{Y}_1 only depends on the noise. We let n_i be the length of \mathbf{Y}_i ($i = 1, 2$), which is deterministic given $\mathbf{X}_{s_{\text{eq}}}$.

J.1 Proof of Lemma 11 (Concentration Bound for $\frac{\ell}{k} \rightarrow \alpha \in [0, 1]$)

We consider the fixed defective set $S = s = [k]$ without loss of generality, since the design is symmetric with respect to re-ordering items.

The case $\alpha \in (0, 1]$

Concentration behavior of various quantities: We first consider the case $\alpha \in (0, 1]$ and establish (329). Recall that we let n_i be the length of \mathbf{Y}_i for $i = 1, 2$. Let M denote the number of tests corresponding to “ $X_{s_{\text{eq}}}^{(i)} = 0, X_{s_{\text{dif}}}^{(i)} \neq 0$ ”,¹⁰ meaning that $n_2 - M$ is the number of tests corresponding to “ $X_{s_{\text{eq}}}^{(i)} = 0, X_{s_{\text{dif}}}^{(i)} = 0$ ”, or equivalently $X_s^{(i)} = 0$. This definition of M coincides with $M_{s_{\text{dif}}}$ in the notation introduced above Lemma 3, but we omit the subscript to lighten notation, since we only consider one particular choice of s_{dif} here.

Note that $\frac{|s_{\text{eq}}|}{k} = \frac{k-\ell}{k} = 1 - \frac{\ell}{k}$, and \mathbf{Y}_1 corresponds to the tests where “ $X_{s_{\text{eq}}}^{(i)} \neq 0$ ” so that we can write $n_1 = W^{(s_{\text{eq}})}$ in the notation of Lemma 7. Thus, for some $\delta > 0$, (299) and (301) in Lemma 7 give that with probability at least $1 - 2\exp(-2\nu^{-1}\delta^2n)$, we have

$$n_1 = \left(1 - e^{-\nu(1-\frac{\ell}{k})(1+o(1))} + \delta' \sqrt{1 - \frac{\ell}{k}}\right)n, \text{ for some } |\delta'| < \delta, \quad (335)$$

which implies

$$n_2 = n - n_1 = \left(e^{-\nu(1-\frac{\ell}{k})(1+o(1))} - \delta' \sqrt{1 - \frac{\ell}{k}}\right)n, \text{ for some } |\delta'| < \delta. \quad (336)$$

We note that (336) holds for all $\ell = 1, \dots, k$. For the case of $\frac{\ell}{k} \rightarrow \alpha \in (0, 1]$, we obtain from (336) that the event

$$n_2 = (e^{-\nu(1-\alpha)} + o(1) - \delta')n, \text{ for some } |\delta'| < \delta. \quad (337)$$

holds with probability at least $1 - 2\exp(-2\nu^{-1}\delta^2n)$.

Next, we seek to obtain a similar concentration result for M . Note that $n - (n_2 - M)$ represents the number of tests where $X_s^{(i)} \neq 0$, so we can write $n - (n_2 - M) = W^{(s)}$. Hence, (299) and (301) in Lemma 7 yield that with probability at least $1 - 2\exp(-2\nu^{-1}\delta^2n)$, we have

$$n - (n_2 - M) = (1 - e^{-\nu} + o(1) + \delta'')n, \text{ for some } |\delta''| < \delta. \quad (338)$$

Then, (337) and (338) imply

$$M = (e^{-\nu(1-\alpha)} - e^{-\nu} + o(1) + \delta^{(3)})n, \text{ for some } |\delta^{(3)}| < 2\delta. \quad (339)$$

In the subsequent analysis, we will frequently suppose that (337) and (339) hold, which is justified via a simple union bound for the multiple high-probability events.

¹⁰Here, M represents the random variable, while we will use lowercase letter m for its specific value or actual realization; we will use similar convention for other variables below, such as V and v .

Let V be the number of positive tests in \mathbf{Y}_2 . Conditioned on some $\mathbf{X} = \mathbf{x}$ (and hence on some $M = m$), the randomness of the noise gives $(V|M = m) \sim \text{Bin}(m, 1 - \rho) + \text{Bin}(n_2 - m, \rho)$, which in turn can be interpreted as a sum of n_2 independent (but non-identical) Bernoulli variables. This allows us to utilize Hoeffding's inequality [12, Thm. 2.8] with respect to the randomness of the noise to obtain that, for any $t \geq 0$,

$$\mathbb{P}\left(|V - [M(1 - \rho) + (n_2 - M)\rho]| \geq t\right) \leq 2 \exp\left(-\frac{2t^2}{n_2}\right). \quad (340)$$

We take $t = n_2\delta$ to obtain that $|V - [M(1 - \rho) + (n_2 - M)\rho]| < n_2\delta$, or equivalently, $V = (1 - 2\rho)M + (\rho + \delta^{(4)})n_2$ for some $|\delta^{(4)}| < \delta$, with probability at least $1 - 2 \exp(-2n_2\delta^2)$. Substituting the high-probability values of M and n_2 in (337) and (339), when $\delta < \frac{1}{4}e^{-\nu(1-\alpha)}$, with probability at least $1 - 2 \exp(-e^{-\nu(1-\alpha)}\delta^2 n)$ we have

$$V = (1 - 2\rho)(e^{-\nu(1-\alpha)} - e^{-\nu} + o(1) + \delta^{(3)})n + (\rho + \delta^{(4)})(e^{-\nu(1-\alpha)} + o(1) - \delta')n \quad (341)$$

$$= ne^{-\nu(1-\alpha)}\left((1 - 2\rho)(1 - e^{-\nu\alpha}) + \rho + o(1) + c_{1,\rho,\nu,\alpha}\delta\right) \quad (342)$$

$$= ne^{-\nu(1-\alpha)}(\rho \star e^{-\nu\alpha} + o(1) + c_{1,\rho,\nu,\alpha}\delta), \quad (343)$$

where in (342) we introduce $c_{1,\rho,\nu,\alpha}$ that is bounded by an absolute constant depending on (ρ, ν, α) ,¹¹ in (343) we make the observation $(1 - 2\rho)(1 - e^{-\nu\alpha}) + \rho = \rho \star e^{-\nu\alpha}$ (recalling that $a \star b = ab + (1 - a)(1 - b)$).

Recall that (334) gives $v^n(\mathbf{Y}|\mathbf{X}_{\text{seq}}, \mathbf{X}_{\text{dif}}) = \log \mathbb{P}(\mathbf{Y}_2|\mathbf{X}_{\text{dif}}) - \log \mathbb{P}(\mathbf{Y}_2)$, and in both terms we implicitly condition on \mathbf{X}_{seq} . To establish the concentration bound for $v^n(\mathbf{Y}|\mathbf{X}_{\text{seq}}, \mathbf{X}_{\text{dif}})$, the main work is to bound these two terms separately.

Bounding $\log \mathbb{P}(\mathbf{Y}_2)$: To handle the term $\mathbb{P}(\mathbf{Y}_2)$, consider a fixed realization \mathbf{y}_2 of \mathbf{Y}_2 . This implies that the realization v of V is also fixed (since it is a function of \mathbf{y}_2), and we suppose that it satisfies the high-probability condition (343). By the symmetry of the test design with respect to re-ordering the test indices, all $\binom{n_2}{v}$ possibilities of \mathbf{Y}_2 having weight v occur with equal probability, and hence, we have $\mathbb{P}(\mathbf{y}_2) = \frac{\mathbb{P}(v)}{\binom{n_2}{v}}$. Then, applying¹² $\mathbb{P}(v) \leq 1$ and substituting (337) and (343), we obtain

$$\mathbb{P}(\mathbf{y}_2) \leq \frac{1}{\binom{ne^{-\nu(1-\alpha)}(1+o(1)+c_{2,\rho,\nu,\alpha}\delta)}{ne^{-\nu(1-\alpha)}(e^{-\alpha\nu}\star\rho+o(1)+c_{1,\rho,\nu,\alpha}\delta)}}. \quad (344)$$

From Lemma 2, we have $\binom{N}{\beta N} \geq \frac{\exp(NH_2(\beta))}{\sqrt{2N}} = \exp(N[H_2(\beta) + o(1)])$ for constant $\beta \in (0, 1)$, which implies that if δ is smaller than some constant depending on (α, ν, ρ) such that $\frac{e^{-\alpha\nu}\star\rho+o(1)+c_{1,\rho,\nu,\alpha}\delta}{1+o(1)+c_{2,\rho,\nu,\alpha}\delta}$ is sufficiently close to $e^{-\alpha\nu} \star \rho$, we can upper bound the right-hand side of (344) by $\exp\left(-[1 + o(1) + c_{3,\rho,\nu,\alpha}\delta]ne^{-\nu(1-\alpha)}H_2(e^{-\alpha\nu}\star\rho)\right)$ for some constant $c_{3,\rho,\nu,\alpha}$. Since this holds for any realization

¹¹Similarly, we will use $c_{2,\rho,\nu,\alpha}$, $c_{3,\rho,\nu,\alpha}$, etc. to denote constants depending on (ρ, ν, α) .

¹²This step may appear loose, but the idea is that if we were to consider all v and take the dominant one, it would have a “large” value of $\mathbb{P}(v)$ anyway, i.e., not decaying fast enough to impact the final result.

of \mathbf{y}_2 whose corresponding v value satisfies (343), we conclude that

$$\mathbf{P}(\mathbf{Y}_2) \leq \exp \left(- [1 + o(1) + c_{3,\rho,\nu,\alpha}\delta] n e^{-\nu(1-\alpha)} H_2(e^{-\alpha\nu} \star \rho) \right) \quad (345)$$

with probability at least $1 - 2 \exp(-e^{-\nu(1-\alpha)} \delta^2 n)$ (over the randomness of \mathbf{Y}_2).

Bounding $\log \mathbf{P}(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{dif}}})$: Next, we consider $\mathbf{P}(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{dif}}})$. Here the entries of \mathbf{Y}_2 are conditionally independent, since given $(\mathbf{X}_{s_{\text{eq}}}, \mathbf{X}_{s_{\text{dif}}})$ they only depend on the noise variables (which are independent across tests). This allows us to decompose

$$\log \mathbf{P}(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{dif}}}) = \log \prod_{i=1}^{n_2} \mathbf{P}(Y_2^{(i)} | X_{s_{\text{dif}}}^{(i)}) = \sum_{i=1}^{n_2} \log \mathbf{P}(Y_2^{(i)} | X_{s_{\text{dif}}}^{(i)}), \quad (346)$$

where without loss of generality we index the tests in \mathbf{Y}_2 from 1 to n_2 . Note that (346) indicates that $\log \mathbf{P}(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{dif}}})$ is independent sum of n_2 random variables, which have a common expected value $-H_2(\rho)$ and are bounded by $|\log \rho|$ (since for any $\rho \in (0, \frac{1}{2})$ we have $|\rho \log \rho + (1-\rho) \log(1-\rho)| \leq \rho |\log \rho| + (1-\rho) |\log \rho| = |\log \rho|$), so using Hoeffding's inequality [12, Thm. 2.8] with respect to the noise randomness gives the following for any $t > 0$:

$$\mathbf{P} \left(\left| \log \mathbf{P}(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{dif}}}) + n_2 H_2(\rho) \right| \geq t \right) \leq 2 \exp \left(- \frac{2t^2}{n_2 \log^2 \rho} \right). \quad (347)$$

Moreover, we use (347) and let $t = \delta n_2 H_2(\rho)$ to obtain that $|\log \mathbf{P}(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{dif}}}) + n_2 H_2(\rho)| \leq \delta n_2 H_2(\rho)$ with probability at least $1 - 2 \exp(-2\delta^2 (\log \rho)^{-2} (H_2(\rho))^2 n_2)$. Further substituting (337), when $\delta < \frac{1}{4} e^{-\nu(1-\alpha)}$, we obtain that

$$\log \mathbf{P}(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{dif}}}) = -n e^{-\nu(1-\alpha)} H_2(\rho) (1 + c_{4,\rho,\nu,\alpha} \delta) \quad (348)$$

for some $c_{4,\rho,\nu,\alpha} = \Theta(1)$, with probability at least $1 - 2 \exp(-\delta^2 (\log \rho)^{-2} (H_2(\rho))^2 e^{-\nu(1-\alpha)} n)$.

Putting the pieces together: We have shown that (345) and (348) hold with probability at least $1 - \exp(-C_{\rho,\nu,\alpha} \delta^2 n)$ for some $C_{\rho,\nu,\alpha} > 0$ depending on (ρ, ν, α) only (here we suppose that δ is given and n is large enough, so that any leading constant before $\exp(-C_{\rho,\nu,\alpha} \delta^2 n)$ can be absorbed into the exponent). Now we start with (334) and proceed as

$$I^n(\mathbf{Y} | \mathbf{X}_{s_{\text{eq}}}, \mathbf{X}_{s_{\text{dif}}}) = \log \mathbf{P}(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{dif}}}) - \log \mathbf{P}(\mathbf{Y}_2) \quad (349)$$

$$\geq -n e^{-\nu(1-\alpha)} H_2(\rho) (1 + c_{4,\rho,\nu,\alpha} \delta) + [1 + o(1) + c_{3,\rho,\nu,\alpha} \delta] n e^{-\nu(1-\alpha)} H_2(e^{-\alpha\nu} \star \rho) \quad (350)$$

$$\geq (1 - o(1)) I_\ell^n + (c_{3,\rho,\nu,\alpha} H_2(e^{-\alpha\nu} \star \rho) - c_{4,\rho,\nu,\alpha} H_2(\rho)) \delta n e^{-\nu(1-\alpha)} \quad (351)$$

$$\geq (1 - o(1)) I_\ell^n - C_{1,\rho,\nu,\alpha} n \delta \quad (352)$$

$$= \left(1 - o(1) - \frac{C_{1,\rho,\nu,\alpha} n \delta}{I_\ell^n} \right) I_\ell^n, \quad (353)$$

where in (350) we substitute (345) and (348), in (351) we recall I_ℓ^n given in (328) in Lemma 10, and in (352) $C_{1,\rho,\nu,\alpha}$ is a constant depending on (ρ, ν, α) . Thus, given any $\delta_1 > 0$, we can set the above

δ as $\delta = \frac{I_\ell^n \delta_1}{2nC_{1,\rho,\nu,\alpha}}$, and we obtain $i^n(\mathbf{Y}|\mathbf{X}_{\text{seq}}, \mathbf{X}_{\text{dif}}) \geq (1 - o(1) - \frac{\delta_1}{2})I_\ell^n$ with probability at least $1 - 4 \exp\left(-\frac{C'_{\rho,\nu,\alpha}(I_\ell^n)^2 \delta_1^2}{n}\right)$ for some $C'_{\rho,\nu,\alpha} > 0$ only depending on (ρ, ν, α) , as desired.

The case $\alpha = 0$

It remains to deal with the case of $\frac{\ell}{k} \rightarrow 0$, for which the proof is generally similar to the case $\alpha \in (0, 1]$ above but needs some minor modifications. We focus on the differences and avoid repeating most near-identical steps.

To get started, we note that (336) remains valid when $\alpha = 0$, which yields that with probability at least $1 - 2 \exp(-2\nu^{-1}\delta^2 n)$, we have

$$n_2 = \left(e^{-\nu(1-\frac{\ell}{k})} + o(1) - \delta'\right)n, \text{ for some } |\delta'| < \delta. \quad (354)$$

Next, we note that (338) remains valid, which together with (354) yields the following with probability at least $1 - 2 \exp(-2\nu^{-1}\delta^2 n)$:

$$M = \left(e^{-\nu(1-\frac{\ell}{k})} - e^{-\nu} + o(1) + \delta^{(3)}\right)n, \text{ for some } |\delta^{(3)}| < 2\delta. \quad (355)$$

Moreover, via the randomness of noise, Hoeffding's inequality still gives (340), in which we set $t = n_2\delta$, substitute (354) and (355), and then perform some simple algebra (as in (341)–(343)) to obtain that

$$V = ne^{-\nu(1-\frac{\ell}{k})}(\rho \star e^{-\frac{\nu\ell}{k}} + o(1) + c_{1,\rho,\nu}\delta) \quad (356)$$

holds with probability at least $1 - 2 \exp(-e^{-\nu}\delta^2 n)$, where $c_{1,\rho,\nu}$ is a quantity bounded by a constant only depending on (ρ, ν) . By a union bound, the events (354), (355) and (356) collectively hold with probability at least $1 - \exp(-C_{\rho,\nu}\delta^2 n)$ for some $C_{\rho,\nu}$ depending on (ρ, ν) .

Then, it is not hard to check the arguments for bounding $\log \mathbf{P}(\mathbf{Y}_2)$ and $\log \mathbf{P}(\mathbf{Y}_2|\mathbf{X}_{\text{dif}})$ directly carry over here, which give the following (corresponding to the earlier (345) and (348)):

$$\log \mathbf{P}(\mathbf{Y}_2) \leq -[1 + o(1) + c_{3,\rho,\nu}\delta]ne^{-\nu(1-\frac{\ell}{k})}H_2(e^{-\frac{\nu\ell}{k}} \star \rho), \quad (357)$$

$$\log \mathbf{P}(\mathbf{Y}_2|\mathbf{X}_{\text{dif}}) = -ne^{-\nu(1-\frac{\ell}{k})}H_2(\rho)(1 + c_{4,\rho,\nu}\delta), \quad (358)$$

which hold with probability at least $1 - \exp(-C'_{\rho,\nu}\delta^2 n)$ for some $C'_{\rho,\nu} > 0$ depending on (ρ, ν) . Starting with (334), we can proceed as in (349)–(353) and obtain

$$i^n(\mathbf{Y}|\mathbf{X}_{\text{seq}}, \mathbf{X}_{\text{dif}}) = \log \mathbf{P}(\mathbf{Y}_2|\mathbf{X}_{\text{dif}}) - \log \mathbf{P}(\mathbf{Y}_2) \quad (359)$$

$$\geq ne^{-\nu(1-\frac{\ell}{k})}[H_2(e^{-\frac{\nu\ell}{k}} \star \rho) - H_2(\rho)] - nC_{5,\rho,\nu}\delta(1 + o(1)) \quad (360)$$

$$\geq (1 - o(1))I_\ell^n - 2nC_{5,\rho,\nu}\delta \quad (361)$$

$$= \left(1 - o(1) - \frac{2nC_{5,\rho,\nu}\delta}{I_\ell^n}\right)I_\ell^n, \quad (362)$$

where in (360) we use (357) and (358), and in (361) we apply Lemma 14 below with $C_{5,\rho,\nu}$ being some constant depending on (ρ, ν) . Thus, given any $\delta_1 > 0$, we can take the above δ as $\delta = \frac{I_\ell^n \delta_1}{4nC_{5,\rho,\nu}}$ to ensure $\iota^n(\mathbf{Y}|\mathbf{X}_{\text{seq}}, \mathbf{X}_{\text{dif}}) \geq (1 - o(1) - \frac{\delta_1}{2})I_\ell^n$, which holds with probability at least $1 - 4 \exp\left(-\frac{C'_{\rho,\nu}(I_\ell^n)^2 \delta_1^2}{n}\right)$. The proof of Lemma 11 is now complete.

In the final steps of the above analysis, we used the following.

Lemma 14. *If $\frac{\ell}{k} \rightarrow 0$, then we have*

$$ne^{-\nu(1-\frac{\ell}{k})}[H_2(e^{-\frac{\nu\ell}{k}} \star \rho) - H_2(\rho)] \geq (1 - o(1))I_\ell^n, \quad (363)$$

where I_ℓ^n satisfies (327).

Proof. Since $\frac{\ell}{k} \rightarrow 0$, we have $ne^{-\nu(1-\frac{\ell}{k})} = ne^{-\nu}(1 + o(1))$. In view of the desired inequality (363) and the asymptotics of I_ℓ^n in (327), some simple algebra reveals that it suffices to prove

$$H_2(e^{-\frac{\nu\ell}{k}} \star \rho) - H_2(\rho) \geq (1 - o(1))\frac{\nu\ell(1 - 2\rho)}{k} \log \frac{1 - \rho}{\rho}. \quad (364)$$

To this end, by using $e^{-\frac{\nu\ell}{k}} = 1 - (1 + o(1))\frac{\nu\ell}{k}$, we first write

$$e^{-\frac{\nu\ell}{k}} \star \rho = \rho e^{-\frac{\nu\ell}{k}} + (1 - \rho)(1 - e^{-\frac{\nu\ell}{k}}) \quad (365)$$

$$= \rho \left(1 - (1 + o(1))\frac{\nu\ell}{k}\right) + (1 - \rho)(1 + o(1))\frac{\nu\ell}{k} \quad (366)$$

$$= \rho + (1 - 2\rho + o(1))\frac{\nu\ell}{k}. \quad (367)$$

Then, we only need to apply a Taylor expansion to $H_2(e^{-\frac{\nu\ell}{k}} \star \rho) - H_2(\rho)$ with $H_2'(\rho) = \log \frac{1-\rho}{\rho}$ to obtain the desired bound (364). \square

J.2 Proof of Lemma 10 (Mutual Information for the Near-Constant Weight Design)

We are interested in $I_\ell^n = \mathbb{E}(\iota^n(\mathbf{X}_{\text{dif}}; \mathbf{Y}|\mathbf{X}_{\text{seq}}))$ when \mathbf{X} is generated via the near-constant column design with parameter ν (i.e., $\Delta = \frac{\nu n}{k}$ tests per item). We define

$$I(\mathbf{X}_{\text{dif}}; \mathbf{Y}|\mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}) := \mathbb{E}_{\mathbf{X}_{\text{dif}}, \mathbf{Y}}(\iota^n(\mathbf{X}_{\text{dif}}; \mathbf{Y}|\mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}})) \quad (368)$$

where the expectation is taken with respect to the randomness of $\mathbf{X}_{\text{dif}}, \mathbf{Y}$ with a given $\mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}$. Further averaging over \mathbf{X}_{seq} will provide the quantity I_ℓ^n of interest, i.e.,

$$I_\ell^n = \mathbb{E}_{\mathbf{X}_{\text{seq}}}[I(\mathbf{X}_{\text{dif}}; \mathbf{Y}|\mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}})]. \quad (369)$$

Recall that $|s_{\text{dif}}| = \ell$ and $|s_{\text{eq}}| = k - \ell$. We first consider $I(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}})$ for fixed $\mathbf{x}_{s_{\text{eq}}}$, which we expand as (see (18))

$$I(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) = H(\mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) - H(\mathbf{Y} | \mathbf{X}_{s_{\text{dif}}}, \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) \quad (370)$$

$$= H(\mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) - nH_2(\rho), \quad (371)$$

where (371) holds because with the full conditioning on \mathbf{X}_s , the only remaining uncertainty is the i.i.d. Bernoulli(ρ) noise. It remains to bound $H(\mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}})$. Given $\mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}$, let \mathbf{Y}_1 be the collection of outcomes of tests containing at least one item from s_{eq} , let \mathbf{Y}_2 be the remaining test outcomes, and let the corresponding lengths of \mathbf{Y}_1 and \mathbf{Y}_2 be n_1 and n_2 , respectively. Note also that given $\mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}$, \mathbf{Y}_1 only depends on the corresponding n_1 noise terms, so the vectors \mathbf{Y}_1 and \mathbf{Y}_2 are conditionally independent. Thus, we get

$$H(\mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) = H(\mathbf{Y}_1 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) + H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) \quad (372)$$

$$= n_1 H_2(\rho) + H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}), \quad (373)$$

where (373) is again because under given $\mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}$, \mathbf{Y}_1 only depends on the randomness of the n_1 noise variables. Substituting (373) into (371) yields

$$I(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) = H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) - n_2 H_2(\rho). \quad (374)$$

In addition, note that we can upper bound

$$H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) \leq n_2 H(Y' | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) \quad (375)$$

where Y' is an arbitrary single test in \mathbf{Y}_2 (e.g., the first one). This follows by sub-additivity of entropy [19, Thm. 2.5.1 & 2.6.5] and the fact that each test in \mathbf{Y}_2 has the same probability of being positive by symmetry.

The case $\frac{\ell}{k} \rightarrow \alpha > 0$

We use (375) to upper bound the conditional entropy $H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}})$. Without noise, a given test in \mathbf{Y}_2 would be negative if and only if it does not contain any item from s_{dif} . Thus, the conditional probability of a single test in \mathbf{Y}_2 being negative *without noise* would be $(1 - \frac{1}{n})^{\ell\Delta} = (1 - \frac{1}{n})^{\frac{\ell\nu n}{k}} = e^{-\alpha\nu(1+o(1))}$. After noise, the probability of a single test being positive becomes $e^{-\alpha\nu(1+o(1))} \star \rho$ (recalling the notation $a \star b = ab + (1-a)(1-b)$). Thus, the conditional entropy of a single test is $H_2(e^{-\alpha\nu(1+o(1))} \star \rho)$, and overall (375) yields

$$H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) \leq n_2 H_2(e^{-\alpha\nu(1+o(1))} \star \rho). \quad (376)$$

Substituting this into (374), we find that

$$I(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) \leq n_2(H_2(e^{-\alpha\nu} \star \rho) - H_2(\rho))(1 + o(1)). \quad (377)$$

Since (299) in Lemma 7 gives $\mathbb{E}(n_2) = n - \mathbb{E}(n_1) = n - (1 - e^{-\nu(1-\alpha)} + o(1))n = (e^{-\nu(1-\alpha)} + o(1))n$, by further averaging (377) over $\mathbf{X}_{s_{\text{eq}}}$ as in (369), we obtain

$$I_\ell^n \leq ne^{-\nu(1-\alpha)}(H_2(e^{-\alpha\nu} \star \rho) - H_2(\rho))(1 + o(1)). \quad (378)$$

For the lower bound, we define V to be the number of positive tests in \mathbf{Y}_2 and first count (n_2, V) as in first part of the proof of Lemma 11 (Appendix J.1). In particular, for any given $\delta > 0$, (337) states that

$$n_2 = (e^{-\nu(1-\alpha)} + o(1) - \delta')n, \text{ for some } |\delta'| < \delta \quad (379)$$

with probability at least $1 - 2\exp(-2\nu^{-1}\delta^2n)$; on this event, (343) further gives with probability at least $1 - 2\exp(-e^{-\nu(1-\alpha)}\delta^2n)$ that

$$V = ne^{-\nu(1-\alpha)}(\rho \star e^{-\nu\alpha} + o(1) + c_{\rho,\nu,\alpha}\delta), \text{ for some } c_{\rho,\nu,\alpha} \leq c_1(\rho, \nu, \alpha), \quad (380)$$

where $c_1(\rho, \nu, \alpha)$ is a constant depending on (ρ, ν, α) . Thus, we can set $\delta = n^{-1/4} = o(1)$ to obtain that the two events

$$\mathcal{A}_1 = \{n_2 = ne^{-\nu(1-\alpha)}(1 + o(1))\}, \quad (381)$$

$$\mathcal{A}_2 = \{V = ne^{-\nu(1-\alpha)}(\rho \star e^{-\nu\alpha})(1 + o(1))\}, \quad (382)$$

hold with probability at least $\mathbb{P}(\mathcal{A}_1) \geq 1 - 2\exp(-2\nu^{-1}\sqrt{n}) = 1 - o(1)$ and $\mathbb{P}(\mathcal{A}_2) \geq 1 - 2\exp(-e^{-\nu(1-\alpha)}\sqrt{n}) = 1 - o(1)$. Since conditioning reduces entropy [19, Thm. 2.6.5], conditioning on V gives

$$H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) \geq H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}, V) \quad (383)$$

$$= \sum_{v=0}^{n_2} \mathbb{P}(V = v) H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}, V = v) \quad (384)$$

$$\geq \mathbb{P}(\mathcal{A}_2) H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}, V = ne^{-\nu(1-\alpha)}(\rho \star e^{-\nu\alpha})(1 + o(1))) \quad (385)$$

$$\geq (1 - o(1)) \log \binom{n_2}{ne^{-\nu(1-\alpha)}(\rho \star e^{-\nu\alpha})(1 + o(1))}, \quad (386)$$

where in (385) we only count the summands in which $V = ne^{-\nu(1-\alpha)}(\rho \star e^{-\nu\alpha})(1 + o(1))$ as specified in the event \mathcal{A}_2 (382), and (386) holds because given $V = v$ (and $\mathbf{x}_{s_{\text{eq}}}$), \mathbf{Y}_2 is uniformly distributed over the $\binom{n_2}{v}$ possibilities of n_2 -dimensional $\{0, 1\}$ -valued vector with v 1's. Substituting (386) into

(374) yields

$$I(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) \geq (1 - o(1)) \log \left(\frac{n_2}{ne^{-\nu(1-\alpha)}(\rho \star e^{-\nu\alpha})(1 + o(1))} \right) - n_2 H_2(\rho). \quad (387)$$

To evaluate I_ℓ^n , we further average over $\mathbf{X}_{s_{\text{eq}}}$ as in (369) to obtain that

$$I_\ell^n \geq \mathbb{E}_{\mathbf{X}_{s_{\text{eq}}}} (I(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}})) \quad (388)$$

$$\geq (1 - o(1)) \log \left(\frac{ne^{-\nu(1-\alpha)}(1 + o(1))}{ne^{-\nu(1-\alpha)}(\rho \star e^{-\nu\alpha})(1 + o(1))} \right) - (ne^{-\nu(1-\alpha)}(1 + o(1))) H_2(\rho) \quad (389)$$

$$= ne^{-\nu(1-\alpha)} (H_2(\rho \star e^{-\nu\alpha}) - H_2(\rho))(1 + o(1)), \quad (390)$$

where (389) follows by using the non-negativity of mutual information to sum only over $\mathbf{x}_{s_{\text{eq}}}$ satisfying the event \mathcal{A}_1 (381) and then using (387) along with $n_2 = ne^{-\nu(1-\alpha)}(1 + o(1))$ (under \mathcal{A}), and in (390) we apply $\log \binom{N}{\beta N} = NH_2(\beta)(1 + o(1))$ (see Lemma 2). Since this matches the upper bound we derived above, (328) follows.

The case $\frac{\ell}{k} \rightarrow 0$

For the upper bound, we again use (375) to upper bound $H(\mathbf{Y}_2 | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}})$, but now trivially use the fact that the probability of a particular test containing an item from s_{dif} is at most $\frac{\nu\ell}{k} \rightarrow 0$ (a union bound of $\frac{\nu n\ell}{k}$ events having probability $\frac{1}{n}$ each). Hence, we have

$$\mathbb{P}[Y' = 1] \leq \frac{\nu\ell}{k}(1 - \rho) + \left(1 - \frac{\nu\ell}{k}\right)\rho = \rho + \frac{\nu\ell}{k}(1 - 2\rho). \quad (391)$$

The corresponding entropy upper bound is

$$H(Y' | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) \leq H_2\left(\rho + \frac{\nu\ell}{k}(1 - 2\rho)\right) = H_2(\rho) + \frac{\nu\ell}{k}(1 - 2\rho) \log\left(\frac{1 - \rho}{\rho}\right)(1 + o(1)), \quad (392)$$

where we use a Taylor expansion with $H'_2(\rho) = \log \frac{1-\rho}{\rho}$ in the last step. Combining with (374) and (375) gives

$$I(\mathbf{X}_{s_{\text{dif}}}; \mathbf{Y} | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}) \leq n_2 \frac{\nu\ell}{k}(1 - 2\rho) \log\left(\frac{1 - \rho}{\rho}\right)(1 + o(1)). \quad (393)$$

Note that $\frac{|s_{\text{eq}}|}{k} = 1 - \frac{\ell}{k} \rightarrow 1$, so (299) in Lemma 7 gives

$$\mathbb{E}(n_2) = n - \mathbb{E}(n_1) = n - \mathbb{E}(W^{(s_{\text{eq}})}) = n - (1 - e^{-\nu} + o(1))n = ne^{-\nu}(1 + o(1)). \quad (394)$$

Thus, by averaging over $\mathbf{X}_{s_{\text{eq}}}$ in (393), we obtain $I_\ell^n \leq \frac{n\nu e^{-\nu}\ell}{k}(1 - 2\rho) \log\left(\frac{1 - \rho}{\rho}\right)(1 + o(1))$.

It remains to prove a matching lower bound. To this end, we write

$$H(\mathbf{Y}_2 | \mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}) = H(\mathbf{Y}_2, V | \mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}) \quad (395)$$

$$= H(\mathbf{Y}_2 | \mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}, V) + H(V | \mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}) \quad (396)$$

$$\geq \sum_{v=0}^{n_2} \mathbb{P}(V = v) H(\mathbf{Y}_2 | \mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}, V = v) + \left(\frac{1}{2} + o(1)\right) \log(n_2), \quad (397)$$

where (395) is because V is deterministic given \mathbf{Y}_2 , (396) holds by the chain rule for entropy [19, Thm. 2.5.1], and (397) holds due to the following argument showing that the entropy of V (when given $\mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}$) is at least $(\frac{1}{2} + o(1)) \log(n_2)$:

- Since conditioning reduces entropy [19, Thm. 2.6.5], we can lower bound the entropy of V by a conditional entropy of V given the *noiseless* test results (still conditioned on the fixed \mathbf{x}_{seq}).
- Given any fixed realization of noiseless test results, we have $V = B_1 + B_2$, where B_1, B_2 are independent binomial random variables with probability parameters $\rho, 1 - \rho$, and count parameters n', n'' satisfying $n' + n'' = n_2$. In particular, $\max\{n', n''\} \geq \frac{n_2}{2}$.
- Again using that conditioning reduces entropy, we deduce that the resulting entropy for V (with the conditioning described above) is lower bounded by the higher of the two entropies associated with B_1 and B_2 . Specifically, after conditioning on B_2 the uncertainty in V reduces to that in B_1 alone, and vice versa.
- The desired claim then follows from the fact that $\text{Binomial}(N, q)$ has entropy $\frac{1}{2} \log(2\pi e N q(1 - q)) + O(1/N)$ as $N \rightarrow \infty$ [1], which simplifies to $(\frac{1}{2} + o(1)) \log N$ when q is constant. (Here we use $q \in \{\rho, 1 - \rho\}$ and $N = \max\{n', n''\} \geq \frac{n_2}{2}$, which we established above.)

As before, we seek to identify high-probability events that simplify the further analysis of (397). Fortunately, what we need has already been shown in the proof of Lemma 11, as we now detail.

We first quantify n_2 . By setting $\delta = \Theta(\sqrt{\frac{\ell}{k}})$ in (354) in the proof of Lemma 11, the event

$$\mathcal{A}_3 = \{n_2 = ne^{-\nu}(1 + o(1))\} \quad (398)$$

holds with probability at least $1 - 2 \exp(-\frac{C_{*,1}\ell n}{k})$, where $C_{*,1}$ can be chosen arbitrarily large as a function of (ν, ρ) . Similarly, regarding V , we set $\delta = \Theta(\sqrt{\frac{\ell}{k}})$ in (356) in the proof of Lemma 11, that the event

$$\mathcal{A}_4 = \{V = \rho ne^{-\nu}(1 + o(1))\} \quad (399)$$

holds with probability at least $1 - 2 \exp(-\frac{C_{*,2}\ell n}{k})$, where we can set $C_{*,2}$ to be arbitrarily large as a function of (ν, ρ) .

We argue that it suffices to consider $\mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}$ such that $n_2 = ne^{-\nu}(1 + o(1))$, as given in the high-probability event \mathcal{A}_1 . Recall that our goal is to characterize $I(\mathbf{X}_{\text{dif}}; \mathbf{Y} | \mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}})$ in (368)

and then average it over \mathbf{X}_{seq} (as in (369)) to characterize I_ℓ^n . Since \mathbf{Y} is binary-valued and has length n , we always have $I(\mathbf{X}_{\text{dif}}; \mathbf{Y} | \mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}) \leq n \log 2 = O(n)$, which means that the total contribution to I_ℓ^n from the low-probability event \mathcal{A}_1^c is bounded by at most the following:

$$O(n) \exp\left(-\frac{C_{*,1}\nu^{-1}\ell n}{k}\right) = \exp\left(O(\log k) - C_{*,1}\nu^{-1}\ell \cdot \Omega(\log k)\right) = o\left(\frac{\ell n}{k}\right), \quad (400)$$

where the first equality holds since $n = \Theta(k \log k)$, and the second equality follows by taking $C_{*,1}$ to be sufficiently large. Thus, the total contribution under \mathcal{A}_1^c does not affect the desired bound (327).

For specific $\mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}$ (with $n_2 = ne^{-\nu}(1+o(1))$ as in the event \mathcal{A}_1), we define $\mathcal{V}_{\mathbf{x}_{\text{seq}}}$ as the corresponding set of V values belonging to the event \mathcal{A}_2 (all satisfying $V = \rho ne^{-\nu}(1+o(1))$). Then, we obtain from (397) that

$$H(\mathbf{Y}_2 | \mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}) \geq \sum_{v \in \mathcal{V}_{\mathbf{x}_{\text{seq}}}} \mathbb{P}(V = v) H(\mathbf{Y}_2 | \mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}, V = v) + \left(\frac{1}{2} + o(1)\right) \log(n_2) \quad (401)$$

$$\geq \sum_{v \in \mathcal{V}_{\mathbf{x}_{\text{seq}}}} \mathbb{P}(V = v) \log\left(\binom{n_2}{v}\right) + \left(\frac{1}{2} + o(1)\right) \log(n_2) \quad (402)$$

$$\geq \sum_{v \in \mathcal{V}_{\mathbf{x}_{\text{seq}}}} \mathbb{P}(V = v) \left[n_2 H_2\left(\frac{v}{n_2}\right) - \frac{1}{2} \log(n_2) - \log(2) \right] + \left(\frac{1}{2} + o(1)\right) \log(n_2) \quad (403)$$

$$= \sum_{v \in \mathcal{V}_{\mathbf{x}_{\text{seq}}}} \mathbb{P}(V = v) n_2 H_2\left(\frac{v}{n_2}\right) + o\left(\frac{\ell n}{k}\right), \quad (404)$$

where (402) holds because $\mathbf{Y}_2 | \{V = v\}$ follows a uniform distribution over $\binom{n_2}{v}$ possibilities, in (403) we use $\log\left(\binom{N}{\beta N}\right) \geq N H_2(\beta) - \frac{1}{2} \log N - \log(2)$ (see Lemma 2), and in (404) we perform some rearrangement to cancel out the terms of $\frac{1}{2} \log(n_2)$ and then write $o(\log n_2) = o(\log n) = o(\log k) = o\left(\frac{\ell n}{k}\right)$, which can be seen analogously to (400). Combining with (374), it follows that

$$I(\mathbf{X}_{\text{dif}}; \mathbf{Y} | \mathbf{X}_{\text{seq}} = \mathbf{x}_{\text{seq}}) \geq \sum_{v \in \mathcal{V}_{\mathbf{x}_{\text{seq}}}} \mathbb{P}(V = v) \left[n_2 H_2\left(\frac{v}{n_2}\right) - n_2 H_2(\rho) \right] \quad (405)$$

$$- \left(\sum_{v \in [n_2] \setminus \mathcal{V}_{\mathbf{x}_{\text{seq}}}} \mathbb{P}(V = v) \right) n_2 H_2(\rho) + o\left(\frac{\ell n}{k}\right) := \mathcal{T}_1 - \mathcal{T}_2 + o\left(\frac{\ell n}{k}\right). \quad (406)$$

By using $n = \Theta(k \log k)$ and choosing $C_{*,2}$ large enough, analogously to (400), we obtain

$$|\mathcal{T}_2| \leq \mathbb{P}(\mathcal{A}_2^c) n_2 H_2(\rho) \leq 2 \exp\left(-\frac{C_{*,2}\ell n}{k}\right) n_2 H_2(\rho) \quad (407)$$

$$= O\left(\exp\left(O(\log k) - C_{*,2}\ell \cdot \Omega(\log k)\right)\right) = o\left(\frac{\ell n}{k}\right), \quad (408)$$

which only amounts to a lower-order term in the desired result (327).

Thus, by further averaging over \mathbf{X}_{seq} as in (369), we only need to prove $\mathbb{E}(\mathcal{T}_1)$ is lower bounded by the right-hand side of (327). Regarding \mathcal{T}_1 , since we are considering $n_2 = ne^{-\nu}(1 + o(1))$ and $v = \rho ne^{-\nu}(1 + o(1))$ (see the high-probability events in (381)–(382)), a Taylor expansion with derivative $H_2'(\rho) = \log \frac{1-\rho}{\rho}$ gives

$$H_2\left(\frac{v}{n_2}\right) - H_2(\rho) = \log\left(\frac{1-\rho}{\rho}\right)\left(\frac{v}{n_2} - \rho\right)(1 + o(1)). \quad (409)$$

We now substitute this into \mathcal{T}_1 , then by decomposing $\sum_{v \in \mathcal{V}_{\mathbf{X}_{\text{seq}}}} = \sum_{v=0}^{n_2} - \sum_{v \in [n_2] \setminus \mathcal{V}_{\mathbf{X}_{\text{seq}}}}$ we can write \mathcal{T}_1 as

$$\begin{aligned} \mathcal{T}_1 &= \sum_{v=0}^{n_2} \mathbb{P}(V = v) \left[\log\left(\frac{1-\rho}{\rho}\right)(v - n_2\rho)(1 + o(1)) \right] \\ &\quad - \sum_{v \in [n_2] \setminus \mathcal{V}_{\mathbf{X}_{\text{seq}}}} \mathbb{P}(V = v) \left[\log\left(\frac{1-\rho}{\rho}\right)(v - n_2\rho)(1 + o(1)) \right] := \mathcal{T}_{11} - \mathcal{T}_{12}. \end{aligned} \quad (410)$$

Note that \mathcal{T}_{12} is again insignificant, since analogously to (400) and (408) we have

$$|\mathcal{T}_{12}| \leq O(n) \mathbb{P}(\mathcal{A}_2^c) \leq O(n) \exp\left(-\frac{C_{*,2}\ell n}{k}\right) = o\left(\frac{\ell n}{k}\right). \quad (411)$$

Observe that \mathcal{T}_{11} can be written as an expectation:

$$\mathcal{T}_{11} = \mathbb{E}_{V|\mathbf{X}_{\text{seq}}=\mathbf{x}_{\text{seq}}} \left[\log\left(\frac{1-\rho}{\rho}\right)(V - n_2\rho)(1 + o(1)) \right] \quad (412)$$

$$= \log\left(\frac{1-\rho}{\rho}\right) (\mathbb{E}_{V|\mathbf{X}_{\text{seq}}=\mathbf{x}_{\text{seq}}}(V) - n_2\rho)(1 + o(1)). \quad (413)$$

Thus, it suffices to show $\mathbb{E}[\mathcal{T}_{11}]$ is lower bounded by the right-hand side of (327), and now we calculate $\mathbb{E}_{V|\mathbf{X}_{\text{seq}}=\mathbf{x}_{\text{seq}}}[V]$. While V represents the number of 1s in the noisy outcomes \mathbf{Y}_2 , it is useful to also define U as the number of tests in \mathbf{Y}_2 that would be positive in the absence of noise (i.e., the tests corresponding to $\mathbf{X}_{\text{seq}} = 0, \mathbf{X}_{\text{dif}} \neq 0$). Then, conditioned on $U = u$, the effect of noise gives

$$\mathbb{E}_{V|\mathbf{X}_{\text{seq}}=\mathbf{x}_{\text{seq}}, U=u}[V] = \mathbb{E}[\text{Bin}(u, 1-\rho) + \text{Bin}(n_2 - u, \rho)] \quad (414)$$

$$= u(1-\rho) + (n_2 - u)\rho \quad (415)$$

$$= (1-2\rho)u + \rho n_2. \quad (416)$$

By averaging over U , we obtain

$$\mathbb{E}_{V|\mathbf{X}_{\text{seq}}=\mathbf{x}_{\text{seq}}}[V] = (1-2\rho)\mathbb{E}_{U|\mathbf{X}_{\text{seq}}=\mathbf{x}_{\text{seq}}}[U] + \rho n_2. \quad (417)$$

To characterize $\mathbb{E}_{U|\mathbf{X}_{\text{seq}}=\mathbf{x}_{\text{seq}}}[U]$, we note that U represents the number of tests (among those corresponding to \mathbf{Y}_2) that get at least one placement from overall $\ell\Delta = \frac{\ell\nu n}{k}$ placements of the

ℓ items in s_{dif} . To understand the resulting expectation, we envision performing the placements sequentially, and note that for a given placement to increment U by one, it suffices that (i) the placement is into one of the n_2 tests that form \mathbf{Y}_2 , *and* (ii) the placement is not into the same test as any of the previous placements. Thus, we have

$$\mathbb{E}[U | \mathbf{X}_{s_{\text{eq}}} = \mathbf{x}_{s_{\text{eq}}}] \geq \mathbb{E}\left[\sum_{i=1}^{\ell\Delta} \mathbb{1}(\text{the } i\text{-th placement satisfies the preceding two conditions})\right] \quad (418)$$

$$\geq \sum_{i=1}^{\ell\Delta} \frac{n_2 - i + 1}{n} \quad (419)$$

$$= \frac{n_2 \ell \Delta}{n} - \frac{(\ell \Delta - 1) \ell \Delta}{2n} \quad (420)$$

$$\geq \frac{\ell \nu n_2}{k} \left(1 - \frac{\ell \nu n}{2k n_2}\right), \quad (421)$$

where (419) holds since at most $i-1$ out of the n_2 tests are ruled out by the previous $i-1$ placements. Substituting (417) and (421) into (413) gives

$$\mathcal{T}_{11} \geq (1 + o(1)) \log\left(\frac{1-\rho}{\rho}\right) (1-2\rho) \frac{\ell \nu n_2}{k} \left(1 - \frac{\ell \nu n}{2k n_2}\right) \quad (422)$$

$$= \left(\frac{n \nu e^{-\nu} \ell}{k} (1-2\rho) \log \frac{1-\rho}{\rho}\right) (1 + o(1)), \quad (423)$$

where (423) holds because we are considering $n_2 = n e^{-\nu} (1 + o(1))$ (as explained earlier in (400)).

We have now established that the right-hand side of (423) is a lower bound for I_ℓ^n (with other terms such as \mathcal{T}_{12} being factored into the $o(1)$ part). This matches the upper bound on I_ℓ^n derived above, and the proof is complete.

J.3 Proof of Lemma 13 (Characterization of $\mathbb{P}(V)$)

We specialize some of the developments from the proof of Lemma 11 to the case that $s_{\text{eq}} = \emptyset$, $s_{\text{dif}} = s$, and $\ell = k$. Hence, in the notation used therein, we have $\mathbf{Y}_1 = \emptyset$ and $\mathbf{Y}_2 = \mathbf{Y}$, M denotes the number of tests that would be positive if there were no noise, and V represents for the number of positive tests in the noisy results (i.e., the number of 1s in \mathbf{Y}). Our goal is to show that $\mathbb{P}(v) = e^{-o(n)}$ for all v within a high-probability set (i.e., V lies in that set with $1 - o(1)$ probability). Since we trivially have $\mathbb{P}(v) \leq 1$, it suffices to only lower bound $\mathbb{P}(v)$.

By setting $\delta = n^{-1/4} = o(1)$ in (339), we have with $1 - o(1)$ probability that

$$M = (1 - e^{-\nu} + o(1))n. \quad (424)$$

Moreover, by setting $\delta = n^{-1/4} = o(1)$ in (343), we have with $1 - o(1)$ probability that

$$V = (\rho \star e^{-\nu} + o(1))n, \quad (425)$$

We will consider values $M = m$ and $V = v$ that satisfy both of these high-probability events.

Give $M = m$, we have that V follows a sum of two independent binomial variables: $(V|M = m) = B_1 + B_2$, where $(B_1|M = m) \sim \text{Bin}(m, 1 - \rho)$ and $(B_2|M = m) \sim \text{Bin}(n - m, \rho)$. Given fixed m and v values satisfying (424)–(425), we can find b_1, b_2 with $b_1 + b_2 = v$ such that

$$b_1 = n(1 - \rho)(1 - e^{-\nu})(1 + o(1)), \quad b_2 = n\rho e^{-\nu}(1 + o(1)), \quad (426)$$

i.e., such that the two are asymptotically close to the mean values of $\text{Bin}(m, 1 - \rho)$ and $\text{Bin}(n - m, \rho)$, and also sum to v . Letting $\mathcal{M}_{\text{typical}}$ be a high-probability set of m values that all satisfy (424), we can then lower bound $\mathbb{P}(V = v)$ as follows:

$$\mathbb{P}(V = v) \geq \sum_{m \in \mathcal{M}_{\text{typical}}} \mathbb{P}(M = m) \mathbb{P}(B_1 = b_1 | M = m) \mathbb{P}(B_2 = b_2 | M = m). \quad (427)$$

We now observe that $\mathbb{P}(B_1 = b_1 | M = m)$ and $\mathbb{P}(B_2 = b_2 | M = m)$ both scale as $e^{-o(n)}$; this directly follows from the anti-concentration of binomial random variables (Lemma 1) and b_1, b_2 being arbitrarily close to the respective means (see (426)). Since $\mathcal{M}_{\text{typical}}$ is a high-probability set of m values, we deduce from (427) that $\mathbb{P}(V = v) \geq e^{-o(n)}$, which completes the proof.

Acknowledgment

This work was supported by the Singapore National Research Foundation (NRF) under grant number A-0008064-00-00.

References

- [1] J. A. Adell, A. Lekuona, and Y. Yu, “Sharp bounds on the entropy of the Poisson law and related quantities,” *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2299–2306, 2010.
- [2] M. Aldridge, “Individual testing is optimal for nonadaptive group testing in the linear regime,” *IEEE Trans. Inf. Theory*, vol. 65, no. 4, pp. 2058–2061, April 2019.
- [3] M. Aldridge, L. Baldassini, and O. Johnson, “Group testing algorithms: Bounds and simulations,” *IEEE Trans. Inf. Theory*, vol. 60, no. 6, pp. 3671–3687, June 2014.
- [4] M. Aldridge, “The capacity of Bernoulli nonadaptive group testing,” *IEEE Trans. Inf. Theory*, vol. 63, no. 11, pp. 7142–7148, 2017.
- [5] M. Aldridge, L. Baldassini, and K. Gunderson, “Almost separable matrices,” *J. Comb. Opt.*, pp. 1–22, 2015.
- [6] M. Aldridge, O. Johnson, and J. Scarlett, “Group testing: An information theory perspective,” *Found. Trend. Comms. Inf. Theory*, vol. 15, no. 3–4, pp. 196–392, 2019.

- [7] N. Arenbaev, “Asymptotic behavior of the multinomial distribution,” *Theory of Probability & Its Applications*, vol. 21, no. 4, pp. 805–810, 1977.
- [8] R. Arratia and L. Gordon, “Tutorial on large deviations for the binomial distribution,” *Bulletin of Mathematical Biology*, vol. 51, no. 1, pp. 125–131, 1989.
- [9] R. Ash, *Information Theory*, ser. Dover books on advanced mathematics. Dover Publications, 1990.
- [10] L. Baldassini, O. Johnson, and M. Aldridge, “The capacity of adaptive group testing,” in *IEEE Int. Symp. Inf. Theory*, July 2013, pp. 2676–2680.
- [11] W. H. Bay, J. Scarlett, and E. Price, “Optimal non-adaptive probabilistic group testing in general sparsity regimes,” *Information and Inference: A Journal of the IMA*, vol. 11, no. 3, pp. 1037–1053, 2022.
- [12] S. Boucheron, G. Lugosi, and P. Massart, *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press, 2013.
- [13] 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 *Allerton Conf. Comm., Ctrl., Comp.*, Sep. 2011, pp. 1832–1839.
- [14] M. Cheraghchi, “Noise-resilient group testing: Limitations and constructions,” in *Int. Symp. Found. Comp. Theory*, 2009, pp. 62–73.
- [15] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, and P. Loick, “Information-theoretic and algorithmic thresholds for group testing,” in *Int. Colloq. Aut., Lang. and Prog. (ICALP)*, 2019.
- [16] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, and P. Loick, “Information-theoretic and algorithmic thresholds for group testing,” *IEEE Trans. Inf. Theory*, vol. 66, no. 12, pp. 7911–7928, 2020.
- [17] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, and P. Loick, “Optimal group testing,” in *Conf. Learn. Theory (COLT)*, 2020.
- [18] A. Coja-Oghlan, O. Gebhard, M. Hahn-Klimroth, A. S. Wein, and I. Zadik, “Statistical and computational phase transitions in group testing,” in *Conf. Learn. Theory (COLT)*, 2022.
- [19] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley & Sons, Inc., 2006.
- [20] D. Du and F. K. Hwang, *Combinatorial group testing and its applications*. World Scientific, 2000, vol. 12.
- [21] A. G. D’yachkov and V. V. Rykov, “A survey of superimposed code theory,” *Prob. Contr. Inf.*, vol. 12, no. 4, pp. 1–13, 1983.

- [22] A. G. D'yachkov and V. V. Rykov, "Bounds on the length of disjunctive codes," *Problemy Peredachi Informatsii*, vol. 18, no. 3, pp. 7–13, 1982.
- [23] A. Feinstein, "A new basic theorem of information theory," *IRE Prof. Group. on Inf. Theory*, vol. 4, no. 4, pp. 2–22, Sept. 1954.
- [24] O. Gebhard, O. Johnson, P. Loick, and M. Rolvien, "Improved bounds for noisy group testing with constant tests per item," *IEEE Trans. Inf. Theory*, vol. 68, no. 4, pp. 2604–2621, 2021.
- [25] T. S. Han, *Information-Spectrum Methods in Information Theory*. Springer, 2003.
- [26] W. Hoeffding, "Probability inequalities for sums of bounded random variables," *J. Amer. Stat. Assoc.*, vol. 58, no. 301, pp. 13–30, 1963.
- [27] F. Hwang, "A method for detecting all defective members in a population by group testing," *J. Amer. Stats. Assoc.*, vol. 67, no. 339, pp. 605–608, 1972.
- [28] F. Iliopoulos and I. Zadik, "Group testing and local search: Is there a computational-statistical gap?" in *Conf. Learn. Theory (COLT)*, 2021.
- [29] O. Johnson, M. Aldridge, and J. Scarlett, "Performance of group testing algorithms with near-constant tests-per-item," *IEEE Trans. Inf. Theory*, vol. 65, no. 2, pp. 707–723, Feb. 2019.
- [30] W. Kautz and R. Singleton, "Nonrandom binary superimposed codes," *IEEE Trans. Inf. Theory*, vol. 10, no. 4, pp. 363–377, 1964.
- [31] E. H. Lieb, "Concavity properties and a generating function for stirling numbers," *Journal of Combinatorial Theory*, vol. 5, no. 2, pp. 203–206, 1968.
- [32] M. B. Malyutov and P. S. Mateev, "Screening designs for non-symmetric response function," *Mat. Zametki*, vol. 29, pp. 109–127, 1980.
- [33] M. Malyutov, "The separating property of random matrices," *Math. Notes Acad. Sci. USSR*, vol. 23, no. 1, pp. 84–91, 1978.
- [34] C. McDiarmid *et al.*, "On the method of bounded differences," *Surveys in Combinatorics*, vol. 141, no. 1, pp. 148–188, 1989.
- [35] J. Niles-Weed and I. Zadik, "It was "all" for "nothing": Sharp phase transitions for noiseless discrete channels," in *Conf. Learn. Theory (COLT)*, 2021.
- [36] B. C. Rennie and A. J. Dobson, "On Stirling numbers of the second kind," *J. Comb. Theory*, vol. 7, no. 2, pp. 116–121, 1969.
- [37] J. Scarlett, "Noisy adaptive group testing: Bounds and algorithms," *IEEE Trans. Inf. Theory*, vol. 65, no. 6, pp. 3646–3661, June 2019.

- [38] J. Scarlett and V. Cevher, “Converse bounds for noisy group testing with arbitrary measurement matrices,” in *IEEE Int. Symp. Inf. Theory*, Barcelona, 2016.
- [39] J. Scarlett and V. Cevher, “Phase transitions in group testing,” in *ACM-SIAM Symp. Disc. Alg. (SODA)*, 2016.
- [40] J. Scarlett and V. Cevher, “Near-optimal noisy group testing via separate decoding of items,” *IEEE J. Sel. Topics Sig. Proc.*, vol. 2, no. 4, pp. 625–638, 2018.
- [41] J. Scarlett and O. Johnson, “Noisy non-adaptive group testing: A (near-) definite defectives approach,” *IEEE Trans. Inf. Theory*, vol. 66, no. 6, pp. 3775–3797, 2020.
- [42] N. Tan, W. Tan, and J. Scarlett, “Performance bounds for group testing with doubly-regular designs,” *IEEE Trans. Inf. Theory*, vol. 69, no. 2, pp. 1224–1243, 2022.
- [43] B. Teo and J. Scarlett, “Noisy adaptive group testing via noisy binary search,” *IEEE Trans. Inf. Theory*, vol. 68, no. 5, pp. 3340–3353, 2022.
- [44] L. V. Truong, M. Aldridge, and J. Scarlett, “On the all-or-nothing behavior of Bernoulli group testing,” *IEEE J. Sel. Areas in Inf. Theory*, vol. 1, no. 3, pp. 669–680, 2020.
- [45] R. C. Yavas, V. Kostina, and M. Effros, “Random access channel coding in the finite blocklength regime,” *IEEE Trans. Inf. Theory*, vol. 67, no. 4, pp. 2115–2140, 2020.