Group Testing using left-and-right-regular sparse-graph codes

Avinash Vem, Nagaraj T. Janakiraman, Krishna R. Narayanan
Department of Electrical and Computer Engineering
Texas A&M University
{vemavinash,tjnagaraj,krn}@tamu.edu

***Abstract*—We consider the problem of non-adaptive group testing of N items out of which K or less items are known to be defective. We propose a testing scheme based on left-*and*-right-regular sparse-graph codes and a simple iterative decoder. We show that for any arbitrarily small $\epsilon > 0$ our scheme requires only $m = c_\epsilon K \log \frac{c_1 N}{K}$ tests to recover $(1-\epsilon)$ fraction of the defective items with high probability (w.h.p) i.e., with probability approaching 1 asymptotically in N and K, where the value of constants c_ϵ and ℓ are a function of the desired error floor ϵ and constant $c_1 = \frac{\ell}{c_\epsilon}$ (≈ 1 for various values of ϵ). More importantly the iterative decoding algorithm has a sub-linear computational complexity of $O(K \log \frac{N}{K})$ which is known to be optimal. Also for $m = c_2 K \log K \log \frac{N}{K}$ tests our scheme recovers the *whole* set of defective items w.h.p. These results are valid for both noiseless and noisy versions of the problem as long as the number of defective items scales sub-linearly with the total number of items, i.e., $K = o(N)$. The simulation results validate the theoretical results by showing a substantial improvement in the number of tests required when compared to the testing scheme based on left-regular sparse-graphs.**

I. Introduction

The problem of Group Testing (GT) refers to testing a large population of N items for K defective items (or sick people) where grouping multiple items together for a single test is possible. The output of the test is *negative* if all the grouped items are non-defective or else the output is *positive.* In the scenario when $K \ll N$, the objective of GT is to design the testing scheme such that the total number of tests to be performed m, is minimized.

This problem was first introduced to the field of statistics by Dorfman [**?**] during World War II for testing the soldiers for syphilis without having to test each soldier individually. Since then, in the literature on GT, three kinds of reconstruction guarantees have been considered: combinatorial, probabilistic and approximate. In the combinatorial designs for the GT problem, the probability of recovering the defective set should be equal to 1, whereas for the probabilistic version, one is interested in recovering *all* the defective items with high probability (w.h.p). For the approximate recovery version, one is interested in only recovering a $(1-\epsilon)$ fraction of the defective items (not the whole set) w.h.p.

For combinatorial GT, the best known lower bound on the number of tests required is $\Omega(K^2 \frac{\log N}{\log K})$ [**?**], [**?**] whereas the best known achievability bound is $\Omega(K^2 \log N)$ tests [**?**], [**?**]. Most of these results were based on algorithms relying on exhaustive searches and hence, have a high computational complexity of at least $O(K^2 N \log N)$. Only recently a scheme with efficient decoding was proposed by Indyk et al., [**?**] where all the defective items are guaranteed to be recovered using $m = O(K^2 \log N)$ tests in poly$(K) \cdot O(m \log^2 m) + O(m^2)$ time.

For the probabilistic version of the problem, it was shown in [**?**], [**?**] that the number of tests necessary is $\Omega(K \log \frac{N}{K})$ which is the best known lower bound in the literature. And regarding the best known achievability bound, Mazumdar [**?**] proposed a construction that requires $O(K \frac{\log^2 N}{\log K})$ tests to recover the defective set w.h.p. For the approximate version, it was shown [**?**] that the required number of tests scale as $O(K \log N)$ and to the best of our knowledge, this is the tightest bound known.

In [**?**], Lee, Pedarsani and Ramchandran proposed a non-adaptive group testing scheme based on *left-regular sparse-graph* codes and an elegant *peeling* based iterative decoder, which are popular tools in channel coding [**?**]. They refer to the scheme as SAFFRON(**S**parse-gr**A**ph codes **F**ramewrok **F**or g**RO**up testi**N**g), a reference which we will use throughout this paper. The authors proved that using the SAFFRON scheme, $m = c_\epsilon K \log_2 N$ number of tests suffice to identify at least $(1-\epsilon)$ fraction of defective items (the approximate version of GT) w.h.p. The precise value of constant c_ϵ as a function of the required error floor ϵ is also given. More importantly the computational complexity of the proposed iterative decoder is only $O(K \log N)$. They also showed that with $m = c_\alpha K \log K \log N$ tests i.e., with an additional $\log K$ factor, the *whole* defective set (the probabilistic version of GT) can be recovered with an asymptotically high probability of $1 - O(K^{-\alpha})$.

Our Contributions

In this work, we propose a non-adaptive GT scheme that is similar to the SAFFRON, but we employ *left-and-right-regular sparse-graph* codes instead of the left-regular sparse-graph codes, and show that we only require $c_\epsilon K \log \frac{N}{K}$ number of tests for an error floor of ϵ in the approximate GT problem. To the best of our knowledge, this is the first scheme which meets the lower bound (We must be careful here since it is not exactly a lower bound but only order optimal lower bound, right?)) for approximate GT problem with optimal computational complexity. We also show that for $m = \Omega(K \log K \log \frac{N}{K})$ tests i.e., with an additional $\log K$ factor the *whole* defective set can be recovered w.h.p $1 - O(K^{-\alpha})$. Note that the testing complexity is only a $\log K$ factor away from the best known lower bound of $\Omega(K \log \frac{N}{K})$ [**?**] for the probabilistic GT problem.

II. Review: Prior Work

Formally, the group testing problem can be stated as below. Given a total number of N items out of which K are defective, the objective is to perform m tests and identify the location of the K defective items from the test outputs. Each item can participate in multiple tests and the result of a test is positive if and only if at least one defective item is present.

Let the support vector $\mathbf{x} \in \{0,1\}^N$ denote the list of items with the non-zero values corresponding to the defective items. A non-adaptive testing scheme consisting of m tests can be represented by a matrix $\mathbf{A} \in \{0,1\}^{m \times N}$ where each row $\mathbf{a}_i$ corresponds to a test. The output corresponding to vector $\mathbf{x}$ and the testing scheme $\mathbf{A}$ can be expressed as

$$\mathbf{y} = \mathbf{A} \odot \mathbf{x}$$

where $\odot$ is the usual matrix multiplication in which the arithmetic addition and multiplication operations are replaced by the boolean OR and AND operations.

Testing Scheme

In this section we will briefly review the SAFFRON testing scheme [**?**] and the decoder. The testing scheme consists of two stages: the first stage is based on a left-regular sparse graph code which pools the N items into non-disjoint M bins where each item belongs to exactly ℓ bins. The second stage comprises producing h testing outputs at each bin where the h different combinations of the pooled items at the respective bin are tested according to a universal signature matrix. For the first stage the authors consider a bipartite graph with N variable nodes (corresponding to the N items) and M bin nodes. Each variable node is connected to ℓ bin nodes chosen uniformly at random from the M available bin nodes. All the variable nodes (historically depicted on the left side of the graph in coding theory) have a degree ℓ (hence, the left-regular graph), whereas the degree of a bin node on the right is a random variable ranging from $[1:n]$.

Definition 1 (Left-regular sparse graph ensemble)**.** Let $\mathcal{G}_\ell(N, M)$ be the ensemble of left-regular bipartite graphs where for each variable node the ℓ right node connections are chosen uniformly at random from the M right nodes.

Let $\mathbf{T}_G \in \{0,1\}^{M \times N}$ be the adjacency matrix corresponding to a graph $G \in \mathcal{G}_\ell(N, M)$ i.e., each column in $\mathbf{T}_G$ corresponds to a variable node and has exactly ℓ ones. Let the rows in matrix $\mathbf{T}_G$ be given by $\mathbf{T}_G = [\mathbf{t}_1^T, \mathbf{t}_2^T, \ldots, \mathbf{t}_{M_1}^T]^T$. For the second stage, let the universal signature matrix defining the h tests at each bin be $\mathbf{U} \in \{0,1\}^{h \times N}$. Then the overall testing matrix $\mathbf{A} := [\mathbf{A}_1^T, \ldots, \mathbf{A}_M^T]^T$ where $\mathbf{A}_i = \mathbf{U}\,\mathrm{diag}(\mathbf{t}_i)$ of size $h \times N$ defines the h tests at i^{th} bin. Thus the total number of tests is $m = M \times h$.

The signature matrix $\mathbf{U}$ employed by the authors in a general setting with parameters r and p can be given by

$$\mathbf{U}_{r,p} = \begin{bmatrix} \mathbf{b}_1 & \mathbf{b}_2 & \cdots & \mathbf{b}_r \\ \overline{\mathbf{b}}_1 & \overline{\mathbf{b}}_2 & \cdots & \overline{\mathbf{b}}_r \\ \mathbf{b}_{\pi_1^1} & \mathbf{b}_{\pi_2^1} & \cdots & \mathbf{b}_{\pi_r^1} \\ \overline{\mathbf{b}}_{\pi_1^1} & \overline{\mathbf{b}}_{\pi_2^1} & \cdots & \overline{\mathbf{b}}_{\pi_r^1} \\ \cdots & & \vdots & \\ \mathbf{b}_{\pi_1^{p-1}} & \mathbf{b}_{\pi_2^{p-1}} & \cdots & \mathbf{b}_{\pi_r^{p-1}} \\ \overline{\mathbf{b}}_{\pi_1^{p-1}} & \overline{\mathbf{b}}_{\pi_2^{p-1}} & \cdots & \overline{\mathbf{b}}_{\pi_r^{p-1}} \end{bmatrix} \quad (1)$$

where $\mathbf{b}_i \in \{0,1\}^{\lceil \log_2 r \rceil}$ is the binary expansion vector for i and $\overline{\mathbf{b}}_i$ is the complement of $\mathbf{b}_i$. $\pi^k = [\pi_1^k, \pi_2^k, \ldots, \pi_r^k]$ denotes a permutation chosen at random from the symmetric group S_r. In the SAFFRON scheme, the authors employ signature matrix with parameters $r = N$ and $p = 3$ thus resulting in a $\mathbf{U}$ of dimension $h \times N$ where $h = 6 \log_2 N$.

Decoding

Before describing the decoding process let us review some terminology. A bin node is referred to as a *singleton* if there is exactly one non-zero variable node connected, as a *double-ton* if two non-zero variable nodes are connected and a *multi-ton* if more than two non-zero variable nodes. For a double-ton if we know the identity of one of the two non-zero variables nodes leaving the decoder to decode the identity of the other one, the bin is referred to as a *resolvable double-ton*.

First part of the decoder which we will refer to as bin decoder will be able to detect and decode the identity of the connected non-zero variable nodes exactly if the bin is a singleton or a resolvable double-ton. If the bin is a multi-ton the bin decoder will detect it neither as a singleton nor a resolvable double-ton w.h.p. The second part of the decoder which we will refer to as the peeling decoder [**?**], when given the identities of some of the non-zero variable nodes by the bin decoder, identifies the bins connected to the recovered variable nodes and looks for newly uncovered resolvable double-tons in these bins. This process of recovering new non-zero variable nodes from already discovered non-zero variable nodes proceeds in an iterative manner. For details of the decoder we refer the reader to [**?**]. Note that we are not literally peeling off the decoded nodes from the graph because of the *non-linear* OR operation on the non-zero variable nodes at each bin thus preventing us in subtracting the effect of the non-zero node from the measurements of the bin node unlike in the problems of compressed sensing or LDPC codes on binary erasure channel.

III. Proposed Scheme

The main difference between the SAFFRON scheme and our proposed scheme is that we use a left-and-right-regular sparse-graph in the first stage for the binning operation.

Definition 2 (Left-and-right-regular sparse graph ensemble)**.** Let $\mathcal{G}_{\ell,r}(N, M)$ be the ensemble of left and right regular bipartite graphs where the $N\ell$ edge connections from the left and $Mr(= N\ell)$ edge connections from the right are paired up according to a permutation $\pi_{N\ell}$ chosen at random from $S_{N\ell}$.

Let $\mathbf{T}_G \in \{0,1\}^{M\times N}$ be the adjacency matrix corresponding to a graph $G \in \mathcal{G}_{\ell,r}(N, M)$ i.e., each row and column in $\mathbf{T}_G$ has r and ℓ ones respectively. And let the universal signature matrix be $\mathbf{U} \in \{0,1\}^{h\times r}$ chosen from the $\mathbf{U}_{r,p}$ ensemble in Eq. (1). Then the overall testing matrix $\mathbf{A} := [\mathbf{A}_1^T, \ldots, \mathbf{A}_M^T]^T$ where $\mathbf{A}_i \in \{0,1\}^{h\times N}$ defining the h tests at i^{th} bin is given by

$$\begin{aligned}\mathbf{A}_i &= [\mathbf{0}, \ldots, \mathbf{0}, \mathbf{u}_1, \mathbf{0}, \ldots, \mathbf{u}_2, \mathbf{0}, \ldots, \mathbf{u}_r], \quad \text{where} \\ \mathbf{t}_i &= [0, \ldots, 0,\ 1,\ 0, \ldots,\ 1,\ 0, \ldots,\ 1].\end{aligned} \tag{2}$$

Note that $\mathbf{A}_i$ is defined by placing the r columns of $\mathbf{U}$ at the r non-zero indices of $\mathbf{t}_i$ and the remaining are padded with zero columns. We can observe that the total number of tests for this testing scheme is $m = M \times h$ where $h = 2p \log_2 r$.

Definition 3 (Regular SAFFRON)**.** Let the ensemble of testing matrices be $\mathcal{G}_{\ell,r}(N, M) \times \mathbf{U}_{r,p}$ where a graph G from $\mathcal{G}_{\ell,r}(N, M)$ and a signature matrix $\mathbf{U}$ from $\mathbf{U}_{r,p}$ are chosen at random and the testing matrix $\mathbf{A}$ is defined according to Eq. (2). Note that the total number of tests is $2pM \log_2 r$ where $r = \frac{N\ell}{M}$.

Analysis

For the regular SAFFRON testing ensemble defined in Def. 3, we employ the iterative decoder described in Sec. II. We will analyze the peeling decoder and the bin decoder separately and union bound the total error probability of the overall decoder. As carried out in [**?**] analysis of the peeling part of the decoder can be simplified by considering a *oracle-based peeling decoder* which decodes a variable node if it is connected to a bin node with degree one or degree two with one of them already decoded, in an iterative fashion. It is easy to verify that the original decoder with accurate bin decoding is equivalent to the oracle based peeling decoder on a pruned graph containing only the non-zero variable nodes.

Definition 4 (Pruned graph ensemble)**.** Let the pruned graph ensemble $\tilde{\mathcal{G}}_{\ell,r}(N, K, M)$ be the set of all graphs obtained by removing a random $N - K$ subset of variable nodes from a graph from the ensemble $\mathcal{G}_{\ell,r}(N, M)$.

Note that graphs from the pruned ensemble have K variable nodes with a degree ℓ whereas the right degree is not regular anymore. Before we analyze the irregular right degree let us define the right degree distribution (d.d) of an ensemble from node (edge) perspective as $R(x) = \sum_i R_i x^i$ ($\rho(x) = \sum_i \rho_i x^{i-1}$) where R_i (ρ_i) is the probability that a random right-node (edge) in a graph from the ensemble has degree i (is connected to a right-node of degree i).

Lemma 5 (Edge d.d of pruned graph)**.** For the pruned graph ensemble $\tilde{\mathcal{G}}_{\ell,r}(N, K, M)$ it can be shown in the limit $K, N \to \infty$ that edge d.d coefficients are $\rho_1 = e^{-\lambda}$ and $\rho_2 = \lambda e^{-\lambda}$ where $\lambda = \ell/c$ for the choice of $M = cK$, c being some constant.

Proof. See . □

Lemma 6. For the pruned graph ensemble $\tilde{\mathcal{G}}_{\ell,r}(N, K, M)$ the oracle-based peeling decoder fails to peel off atleast $(1 - \epsilon)$ fraction of the variable nodes with exponentially decaying probability for $M = c_\epsilon K$ where c_ϵ for various ϵ is given in Table. I.

Proof. From Lemma. 5 the edge d.d. coefficients ρ_1 and ρ_2 are identical to that of the SAFFRON scheme. Thus we can employ the proof same as [**?**, Thm. 4.1] where the authors use the density evolution equations based on ρ_1 and ρ_2 to show the required result. For details we refer the reader to. □

ϵ	10^{-3}	10^{-4}	10^{-5}	10^{-6}	10^{-7}	10^{-8}	10^{-9}
c_ϵ	6.13	7.88	9.63	11.36	13.10	14.84	16.57
ℓ	7	9	10	12	14	15	17

TABLE I
Constants for various error floor values

Lemma 7 (Bin decoder analysis)**.** For a signature matrix $\mathbf{U}_{r,p}$ as described in (1), the bin decoder successfully detects and resolves if the bin is either a singleton or a resolvable double-ton. In the case of the bin being a multi-ton, the bin decoder declares a wrong hypothesis of either a singleton or a resolvable double-ton with a probability no greater than $\frac{1}{r^{p-1}}$.

Proof. This result was proved in [**?**] for the choice of parameters $r = N$ and $p = 3$. The extension of the result to general r, p parameters is straight forward. □

Theorem 8. Let $p \in \mathbb{Z}$ such that $K = o(N^{1-1/p})$. A random testing matrix from the proposed regular SAFFRON ensemble $\mathcal{G}_{\ell, \frac{N\ell}{c_\epsilon K}}(N, c_\epsilon K) \times \mathbf{U}_{\frac{N\ell}{c_\epsilon K}, p}$ with $m = c \cdot K \log_2 \frac{c_2 N}{K}$ tests recovers atleast $(1 - \epsilon)$ fraction of the defective items w.h.p. The computational complexity of the decoding scheme is $O(K \log \frac{N}{K})$. The constants are $c = 2pc_\epsilon, c_2 = \frac{\ell}{c_\epsilon}$ where ℓ and c_ϵ for various values of ϵ are given in Table. I.

Proof. It remains to be shown that the total probability of error decays asymptotically in K and N. Let E_1 be the event of oracle-based peeling decoder terminating without recovering atleast $(1 - \epsilon)K$ variable nodes. Let E_2 be the event of the bin decoder making an error during the entirety of the peeling process and E_{bin} be the event of one instance of bin decoder making an error. The total probability of error P_e can be upper bounded by

$$\begin{aligned} P_e &\le \Pr(E_1) + \Pr(E_2) \\ &\le \Pr(E_1) + K\ell \Pr(E_{\text{bin}}) \\ &\in O\left(\frac{K^p}{N^{p-1}}\right) \end{aligned}$$

where the second inequality is due to the union bound over a maximum of $K\ell$ (number of edges in the pruned graph) instances of bin decoding. The third line is due to the fact that $\Pr(E_1)$ is exponentially decaying in K (see Lemma. 6) and $\Pr(E_{\text{bin}}) = (\frac{c_\epsilon K}{N\ell})^{p-1}$ (see Lemma. 7 and Def. 3) $\square$

IV. Total Recovery: Singleton-Only Variant

In this section we will look at the proposed regular-SAFFRON scheme but with a decoder that operates only on the singleton bins. To elaborate, the only difference is in the decoder which is not iterative in this framework and recovers only the variable nodes connected to the singleton bins and terminates. The trade-off is that now we can recover the *whole* defective set instead of just a large fraction of the defective items with an additional $\log K$ factor tests. Since we do not need to be able to recover resolvable double-tons we only need $2\log_2 r$ number of tests at each bin i.e. we choose $p = 1$ in Eqn. (1).

Theorem 9. For $M = c_\alpha K \log K$ and $(\ell, r) = (c_\alpha \log K, \frac{N}{K})$ a random testing matrix from the regular SAFFRON ensemble $\mathcal{G}_{\ell,r}(N, M) \times \mathbf{U}_{r,1}$ with $m = 2c_\alpha K \log K \log_2 \frac{N}{K}$ tests with the singleton-only decoder fails to recover all the non-zero variable nodes with a vanishing probability of $O(\frac{1}{K^\alpha})$ where $c_\alpha = e(1+\alpha)$.

Proof. For proof see . $\square$

V. Robust Group Testing

In this section we extend our scheme to the group testing problem with noisy test results,

$$\mathbf{y} = \mathbf{A} \odot \mathbf{x} + \mathbf{w},$$

where the addition is over binary field and $\mathbf{w} \in \{0,1\}^N$ is an i.i.d noise vector distributed according to Bernoulli distribution with parameter $0 < q < \frac{1}{2}$.

Testing Scheme

Let $\mathcal{C}_n$ be a binary error-correcting code with the encoder and decoder functions $f : \{0,1\}^n \to \{0,1\}^{\frac{n}{R}}$ and $g : \{0,1\}^{\frac{n}{R}} \to \{0,1\}^n$ respectively where R is the rate of the code. For the choice of $\mathcal{C}_n$ consider any capacity achieving codes, polar or spatially-coupled codes, whose properties can be summarized as follows:

- There exists a sequence of codes $\{\mathcal{C}_n\}$ with the rate of each code being R satisfying $R < C_{\text{Sh}}(q) - \delta$ for any arbitrary small constant δ such that the probability of error $\Pr(g(\mathbf{x}+\mathbf{w}) \neq \mathbf{x}) < 2^{-\kappa n}$ for some $\kappa > 0$. Here $C_{\text{Sh}}(q) = 1 - h_2(q)$ is the Shannon capacity of the BSC channel model.

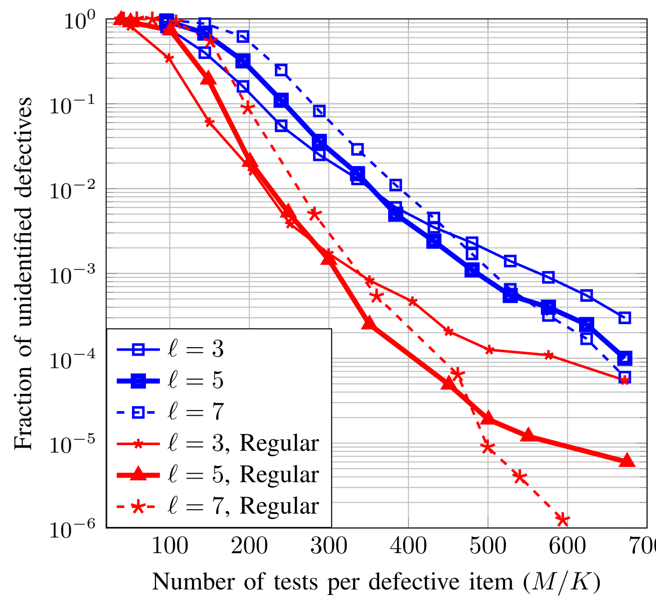

Fig. 1. MonteCarlo simulations for $K = 100, N = 2^{16}$. We compare the SAFFRON scheme with our regular SAFFRON scheme for various left degrees $\ell \in \{3,5,7\}$. For a given ℓ the bin detection size is fixed and we vary the number of bins. The plots in blue indicate the SAFFRON scheme and the plots in red indicate our regular SAFFRON scheme based on left-and-right-regular bipartite graphs.

The modified signature matrix $\mathbf{U}'_{r,p}$ can be described via $\mathbf{U}_{r,p}$ given in Eq. (1) and encoding function f for $\mathcal{C}_n$ where $n = \lceil \log_2 r \rceil$ as follows:

$$\mathbf{U}'_{r,p} := \begin{bmatrix} f(\mathbf{b}_1) & f(\mathbf{b}_2) & \cdots & f(\mathbf{b}_r) \\ \overline{f(\mathbf{b}_1)} & \overline{f(\mathbf{b}_2)} & \cdots & \overline{f(\mathbf{b}_r)} \\ f(\mathbf{b}_{\pi_1^1}) & f(\mathbf{b}_{\pi_2^1}) & \cdots & f(\mathbf{b}_{\pi_r^1}) \\ \overline{f(\mathbf{b}_{\pi_1^1})} & \overline{f(\mathbf{b}_{\pi_2^1})} & \cdots & \overline{f(\mathbf{b}_{\pi_r^1})} \\ \cdots & & \vdots & \\ f(\mathbf{b}_{\pi_1^{p-1}}) & f(\mathbf{b}_{\pi_2^{p-1}}) & \cdots & f(\mathbf{b}_{\pi_r^{p-1}}) \\ \overline{f(\mathbf{b}_{\pi_1^{p-1}})} & \overline{f(\mathbf{b}_{\pi_2^{p-1}})} & \cdots & \overline{f(\mathbf{b}_{\pi_r^{p-1}})} \end{bmatrix} \quad (3)$$

Then the overall testing matrix $\mathbf{A}$ is defined similar to the noiseless case in Sec. II except that $\mathbf{U}$ will be replaced by $\mathbf{U}'$ in Eqn. (2).

Lemma 10 (Robust Bin Decoder Analysis)**.** For $\mathbf{U'}_{r,p}$ in (3), the robust bin decoder misses a singleton with probability no greater than $\frac{p}{r^\kappa}$. The robust bin decoder wrongly declares a singleton with probability at most $\frac{1}{r^{p(1+\kappa)}}$.

The fraction of missed singletons can be compensated by using $M(1+\frac{p}{r^\kappa})$ instead of M such that the total number of singletons decoded will be $M(1+\frac{p}{r^\kappa})(1-\frac{p}{r^\kappa}) \approx M$.

Theorem 11. Let $p \in \mathbb{Z}$ such that $K = o\left(N^{(p-1)/p}\right)$. The proposed robust regular SAFFRON scheme using $m =$

$c \cdot K \log_2 \frac{N\ell}{c_\epsilon K}$ tests recovers atleast $(1-\epsilon)$ fraction of the defective items w.h.p. where $c = 2p\beta(q)c_\epsilon$ and $\beta(q) = 1/R$. The computational complexity of the decoding scheme is $O(K \log \frac{N}{K})$.

Proof. Similar to the noiseless case the total probability of error P_e is dominated by the performance of bin decoder.

$$\begin{aligned} P_e &\leq \Pr(E_1) + K\ell \Pr(E_{\text{bin}}) \\ &\leq \Pr(E_1) + \frac{K^{1+p(1+\kappa)}}{N^{p(1+\kappa)}} \\ &= O(N^{-\kappa-1/p}) \end{aligned}$$

where the last line is due to the fact that $\Pr(E_1)$ is exponentially decaying in K and $K \leq N^{(p-1)/p}$ for large enough K, N. □

VI. Simulation Results

In this section we will evaluate the performance of our proposed regular SAFFRON scheme via Monte Carlo simulations and compare it with the results for SAFFRON scheme provided in [**?**].

Noiseless Group Testing

The system parameters are summarized below:

- We fix $N = 2^{16}$ and $K = 100$
- For $\ell \in \{3, 5, 7\}$ we vary the number of bins $M = cK$.
- In Eqn. 1 the parameter $p = 2$ is chosen for matrix $\mathbf{U}$
- Thus the bin detection size is $h = 6 \log_2 \frac{N\ell}{cK}$
- Hence the total number of tests $m = 6cK \log_2 \left(\frac{N\ell}{cK}\right)$

The results are shown in Fig. 1. We observe that there is clear improvement in performance for our regular SAFFRON scheme when compared to the SAFFRON scheme for each $\ell \in \{3, 5, 7\}$.

Noisy Group Testing

Similar to the noiseless group testing problem we simulate the performance of our robust regular SAFFRON scheme and compare it with that of the SAFFRON scheme. For convenience of comparison we choose our system parameters identical to the choices in [**?**]. The system parameters are summarized below:

- $N = 2^{32}, K = 2^7$. We fix $\ell = 12, M = 11.36K$
- BSC noise parameter $q \in \{0.03, 0.04, 0.05\}$
- In Eqn. 1 the parameter $p = 1$ is chosen for matrix $\mathbf{U}$
- Thus the bin detection size is $h = 4 \log_2 \frac{N\ell}{M}$

Note that for the above set of parameters the right degree $r = \frac{N\ell}{M} \approx 26$. We choose to operate in field $GF(2^7)$ thus giving us a message length of 4 symbols. For the choice of code we use a $(4 + 2e, 4)$ Reed-Solomon code for $e \in [0 : 8]$ thus giving us a column length of $4 \times 7(4 + 2e)$ bits at each bin and the total number of tests $m = 28M(4 + 2e)$.

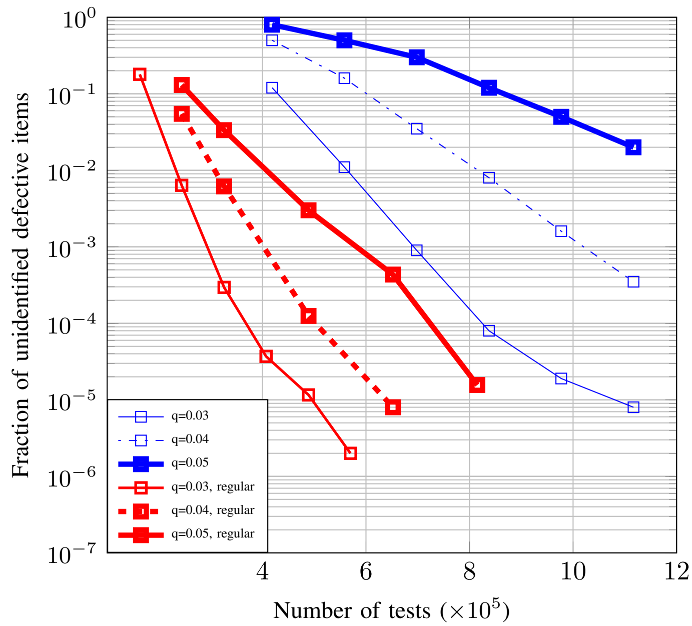

Fig. 2. MonteCarlo simulations for $K = 100, N = 2^{16}$. We compare the SAFFRON scheme [**?**] with the proposed regular SAFFRON scheme for various left degrees $\ell \in \{3, 5, 7\}$. The plots in blue indicate the SAFFRON scheme and the plots in red indicate our regular SAFFRON scheme based on left-and-right-regular bipartite graphs.