medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

A Recovery Algorithm and Pooling Designs for One-Stage Noisy Group Testing under the Probabilistic Framework *

Yining Liu[0000−0002−8779−2906], Sachin Kadyan[0000−0002−6079−7627], and Itsik Pe'er[0000−0002−6128−7231]

Department of Computer Science, Columbia University, New York, USA
{yl4536,sk4835,ip2169}@columbia.edu

Abstract. Group testing saves time and resources by testing each pre-assigned group instead of each individual, and one-stage group testing emerged as essential for cost-effectively controlling the current COVID-19 pandemic. Yet, the practical challenge of adjusting pooling designs based on infection rate has not been systematically addressed. In particular, there are both theoretical interests and practical motivation to analyze one-stage group testing at finite, practical problem sizes, rather than asymptotic ones, under noisy, rather than perfect tests, and when the number of positives is randomly distributed, rather than fixed.
Here, we study noisy group testing under the probabilistic framework by modeling the infection vector as a random vector with Bernoulli entries. Our main contributions include a practical one-stage group testing protocol guided by maximizing pool entropy and a maximum-likelihood recovery algorithm under the probabilistic framework. Our findings highlight the implications of introducing randomness to the infection vectors – we find that the combinatorial structure of the pooling designs plays a less important role than the parameters such as pool size and redundancy.

Keywords: Non-adaptive Group Testing · COVID-19 · Experimental Designs

1 Introduction

Group testing is a procedure to find positives in a cohort by applying tests for the presence of any positives to cohort subsets (groups), instead of testing each individual separately. When the fraction of positives among the samples is low, implementing group testing saves time and resources. Group testing has applications in genetics [21], drug screening [13] communications [25] and epidemiology [23].

One-stage group testing [6] addresses the scenario of groups being set in advance, independently of test results - a common practical requirement, e.g. due to testing speed constraints. Existing algorithms on infection vector recovery for

* Columbia University unrestricted funds

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

noisy group testing under the combinatorial prior (the number of positives among samples is fixed) include LP relaxation [16], belief propagation [20], Markov Chain Monte Carlo (MCMC) [9], Noisy Combinatorial Orthogonal Matching Pursuit (NCOMP) [4], separate decoding [18] and Definite positives (DD) [19]. However, group testing under a fully probabilistic framework has not been extensively studied. In particular, to the best of our knowledge, there has been a lack of work on non-asymptotic results on noisy group testing in the realistic scenario where number of positives in not known in advance, as in [2, 3], but rather is a random variable.

As part of the global efforts to control the COVID-19 pandemic, there has been an emerging body of work on implementing one-stage group testing for COVID-19 [10, 11, 22–24]. However, testing scenarios vary significantly due to fluctuating infection rates across time and geography often caused by emerging variants, as well as different testing objectives, such as screening of health care workers versus large scale monitoring of the community. Group testing protocols should thus be adjusted according to the infection rate among the tested samples, measurement error rates, and the recovery error tolerance level [5].

As a result, systematically studying one-stage noisy group testing with a random number of positives is of both theoretical interests and practical importance. In this paper, we study noisy group testing under the probabilistic framework in order to address the practical challenges posted by group testing implementation for COVID-19.

The rest of the paper is organized as follows. We begin with an introduction of the noisy group testing under the probabilistic framework in Section 2, including a group testing protocol and a novel recovery algorithm under the probabilistic framework. The performance of the recovery algorithm and the pooling designs is in Section 3. Finally, we conclude with a discussion on future work in Section 4.

2 Methods

2.1 Noisy Group Testing under the Probabilistic Framework

We assume a tested cohort of n individuals, with some infection rate f among the population the cohort is sampled for. Note, that the actual number of true positives is not known in advance, as combinatorial priors unrealistically assume [2, 3]. Also, this is the ground truth rate of infected individuals, as opposed to the observed positivity rate. Specifically:

Definition 1. *An infection vector X is a random vector with n i.i.d. Bernoulli(f) entries.*

We aim to design a protocol to test n samples with $t < n$ tests. We arrange the pool assignments into a pooling matrix $M \in \{0, 1\}^{t \times n}$ such that $M_{ij} = 1$ if and only if individual j is included in pool i. Notice that each row i sum of M correspond to a pool size s_i, and each column j sum corresponds to the number of pools that a sample participates, which we define as redundancy r_j.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Following [1] and [8], we focus on the following four classes of pooling designs that are of practical interests (see Fig. 1):

1. (constant-pool design) each test combines a fixed number $s_i = s$ of samples, where s samples are chosen uniformly at random among the n samples;
2. (constant-redundancy design) each individual participates in a fixed number $r_j = r$ of tests, where r tests are chosen uniformly at random among the t tests;
3. (doubly-regular design) draw a design uniformly at random from the designs that are both constant-pool and constant-redundancy;
4. (Reed-Solomon design) the pool assignment for each individual is obtained as a concatenation of Reed-Solomon error correcting code; see [8] for more details on the explicit construction.

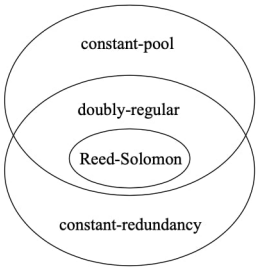

Fig. 1. Four classes of pooling designs.

Under the noiseless setting, a test result is negative if and only if all samples in the pool are negative. In practice, test results might suffer from measurement errors. Suppose the test we use for each pool has a false negative rate of f_0 and a false positive rate of f_1, i.e. consider the asymmetric noisy channel (where f_0 and f_1 can be distinct) that combines the additive model and the dilution model in [2]. See Fig. 2 for an example of a doubly-regular pooling matrix under the probabilistic framework.

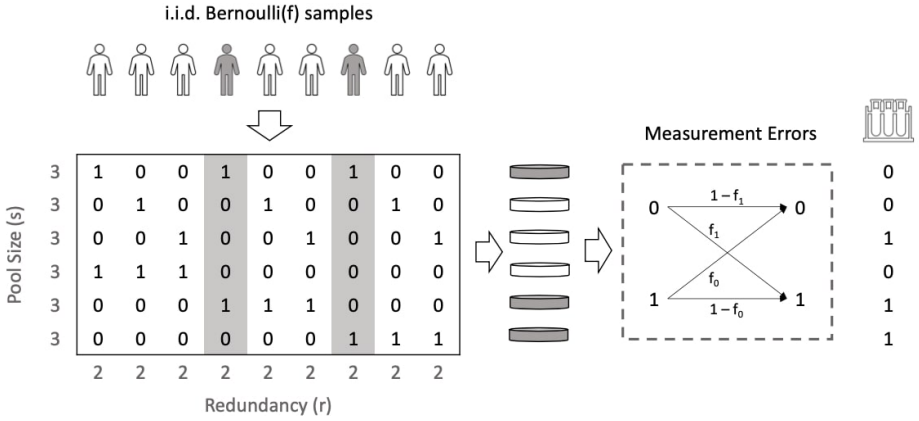

Fig. 2. Noisy group testing under the probabilistic framework. The pooling matrix is a doubly-regular design with $s = 3$ and $r = 2$. A pool is negative if and only if there is no positive sample participates in the pool. Each pool is then passed to an asymmetric noisy channel to model measurement errors.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

2.2 A Group Testing Protocol

Let Y_i be the indicator that the test result for the pool i is negative; then

$$Y_i \sim \text{Bernoulli}(f_0(1-(1-f)^{s_i}) + (1-f_1)(1-f)^{s_i}).$$

Extending the idea in [22] to the noisy setting, a pool design should maximize the entropy of Y_i; hence the optimal pool size is

$$\arg\max_{s_i} H(y_i; s_i) = \frac{\ln(\frac{0.5-f_0}{1-f_1-f_0})}{\ln(1-f)}. \tag{1}$$

Infection rates therefore dramatically affect the theoretical optimal pool size, while measurement errors slightly offsets such optima (see Fig. 3).

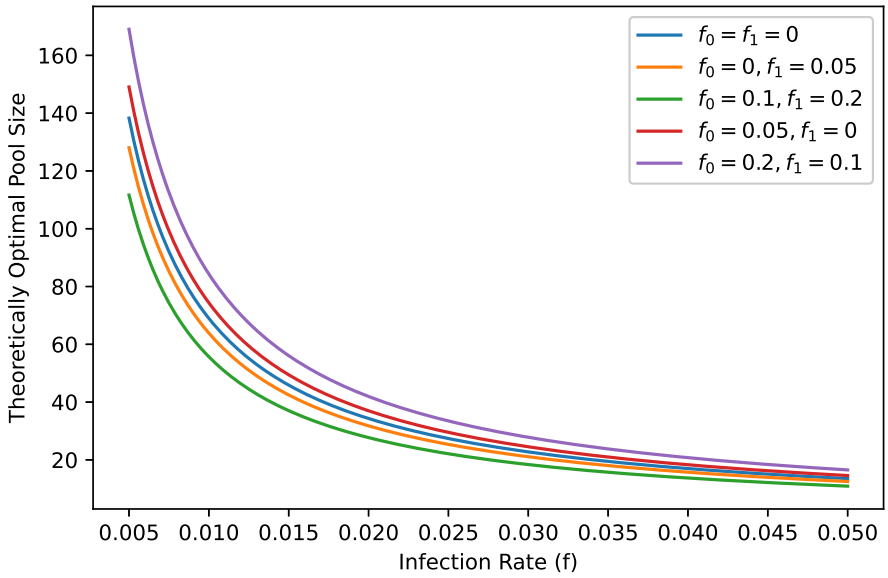

Fig. 3. Theoretically optimal pool size as a function of infection rate; false positives (resp. negatives) slightly reduce (increase) optimal pool size.

We will show in Section 3 that given a fixed set of parameters r, s, t, the four classes of designs (and that of each member of a given class) perform comparably under this probabilistic framework. Hence, practitioners can safely choose a design class with practical advantages and then optimize the parameters r, s, t.

We thus propose the following protocol for implementing one-stage group testing under a given infection rate:

1. From the potential pool sizes (subject to practical constraints), choose s to be closest to the theoretical optimal pool size, per Equation 1.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

2. Simulate the performance curve using the recovery algorithm in Section 2.3.
3. Choose the number of pools t based on the error tolerance level.
4. Perform one-stage group testing with the same recovery algorithm in step 2.

2.3 Recovery Algorithm

We now introduce an algorithm to recover the infection vector for noisy group testing under the probabilistic framework.

Assume we are given a pooling design M with t pools; the pool results is

$$Y := \text{sign}(MX) - E^{(0)} + E^{(1)}$$

where $E^{(0)}, E^{(1)}$ are false-negatives vector and false-positives vector respectively, i.e. for $i = 1, \ldots, t$,

$$\begin{cases} E_i^{(0)} \sim \text{Bernoulli}(f_0) & \text{if } (MX)_i > 0, \\ E_i^{(0)} \sim \text{Bernoulli}(0) & \text{if } (MX)_i = 0, \\ E_i^{(1)} \sim \text{Bernoulli}(f_1) & \text{if } (MX)_i = 0, \\ E_i^{(1)} \sim \text{Bernoulli}(0) & \text{if } (MX)_i > 0. \end{cases}$$

A recovery algorithm is given a pool result $y \in \{0,1\}^t$ and outputs $(\hat{x}, \hat{e}^{(0)}, \hat{e}^{(1)})$, an estimate for $(X, E^{(0)}, E^{(1)})$. Our recovery algorithm (**Algorithm 1**) is given by solving the following integer linear program:

$$\begin{aligned} \text{minimize} \quad & -||\hat{x}|| \ln(\tfrac{f}{1-f}) - ||\hat{e}^{(0)}|| \ln(\tfrac{f_0}{1-f_0}) - ||\hat{e}^{(1)}|| \ln(\tfrac{f_1}{1-f_1}) \\ \text{subject to} \quad & 0.5\hat{e}_i^{(0)} \le (M\hat{x})_i \le t\hat{e}_i^{(0)} + 0.5 && i \in \{1, \ldots, t\} \text{ s.t. } y_i = 0 \\ & -0.5\hat{e}_i^{(1)} + 0.5 \le (M\hat{x})_i \le t(1 - \hat{e}_i^{(1)}) + 0.5 && i \in \{1, \ldots, t\} \text{ s.t. } y_i = 1 \\ & \hat{x} \in \{0,1\}^n \\ & \hat{e}^{(0)}, \hat{e}^{(1)} \in \{0,1\}^t \end{aligned}$$

Here, for a given binary vector $x \in \{0,1\}^n$, we use $||x||$ to denote the Hamming weight of x, i.e. $||x|| = \sum_{i=1}^n x_i$.

We now analyze the correctness and optimality of Algorithm 1.

Theorem 1. *Algorithm 1 returns a realizable output.*

Proof. In order for an output $(\hat{x}, \hat{e}^{(0)}, \hat{e}^{(1)})$ to be realizable, each row of the equation $y = \text{sign}(M\hat{x}) - \hat{e}^{(0)} + \hat{e}^{(1)}$ must fall into one of the following cases:

$$\begin{cases} y_i = 0, (M\hat{x})_i = 0, \hat{e}_i^{(0)} = 0; \\ y_i = 0, (M\hat{x})_i > 0, \hat{e}_i^{(0)} = 1; \\ y_i = 1, (M\hat{x})_i > 0, \hat{e}_i^{(1)} = 0; \\ y_i = 1, (M\hat{x})_i = 0, \hat{e}_i^{(1)} = 1. \end{cases}$$

Notice that for $y_i = 0$, we have

$$\begin{aligned} \hat{e}_i^{(0)} = 1 &\iff 0.5 \le (M\hat{x})_i \le t + 0.5 \iff 1 \le (M\hat{x})_i \le t, \\ \hat{e}_i^{(0)} = 0 &\iff 0 \le (M\hat{x})_i \le 0.5 \iff (M\hat{x})_i = 0. \end{aligned}$$

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Similarly, for $y_i = 1$, we have

$$\hat{e}_i^{(1)} = 0 \iff 0.5 \le (M\hat{x})_i \le t + 0.5 \iff 1 \le (M\hat{x})_i \le t,$$
$$\hat{e}_i^{(1)} = 1 \iff 0 \le (M\hat{x})_i \le 0.5 \iff (M\hat{x})_i = 0.$$

Therefore, the constraints in the integer linear program guarantees the output is realizable.

Theorem 2. *Algorithm 1 returns a maximum likelihood estimate (MLE), i.e. an output* $(\hat{x}, \hat{e}^{(0)}, \hat{e}^{(1)})$ *maximizes* $\Pr(X = x, E^{(0)} = e^{(0)}, E^{(1)} = e^{(1)})$.

Proof. Let n_0 be the number of pools with no positive sample and n_1 be the number of pools containing at least one positive sample. The log-likelihood is

$$\begin{aligned}
&\ln \Pr(X = x, E^{(0)} = e^{(0)}, E^{(1)} = e^{(1)}) \\
=&\ln \Pr(X = x) + \ln \Pr(E^{(0)} = e^{(0)}, E^{(1)} = e^{(1)} | X = x) \\
=&||x|| \ln f + (n - ||x||) \ln(1 - f) + ||e^{(0)}|| \ln f_0 + (n_1 - ||e^{(0)}||) \ln(1 - f_0) \\
&+ ||e^{(1)}|| \ln f_1 + (n_0 - ||e^{(1)}||) \ln(1 - f_1) \\
=&||x|| \ln(\frac{f}{1-f}) + ||e^{(0)}|| \ln(\frac{f_0}{1-f_0}) + ||e^{(1)}|| \ln(\frac{f_1}{1-f_1}) + \text{const.}
\end{aligned}$$

Therefore, the objective function in the integer linear program maximizes the likelihood.

Open source code implementing all our methods is available at [15].

3 Results

For biological experiments, sample sizes $n = 96, 384, 1536$ are of particular interests because the experiments tend to be conducted on 96-well plates, 384-well plates, or 1536-well plates [8]. The savings in resources offered by pooled design compared to individual testing stops being worth the increased complexity of the experimental paradigm when infection rates are high. For example, under an infection rate $f = 0.1$, using a constant-pool design with optimal pool size $s = 7$, even a 2-fold saving, with $T = n/2 = 192$ pools to recover $n = 384$ samples, the average accuracy over 1000 trials is unacceptably low at 96.5%. As a result, we believe our group testing protocol should not be applied when the infection rate is over 0.1, and limit our analysis to infection rate up to 0.025 in this section.

3.1 Performance of the Recovery Algorithm

In order to test the performance of the recovery algorithm, we used the pooling design $M \in \{0, 1\}^{48 \times 384}$ (testing 384 individuals in 48 pools) in [23].

We simulated 1000 infection vectors $x \in \{0, 1\}^{384}$ under the population infection rate $f = 0.02$ and evaluated recovery under different rates of false positives and false negatives. We observe error rate to only slightly detract from the accuracy of noiseless measurements (see Fig. 4).

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

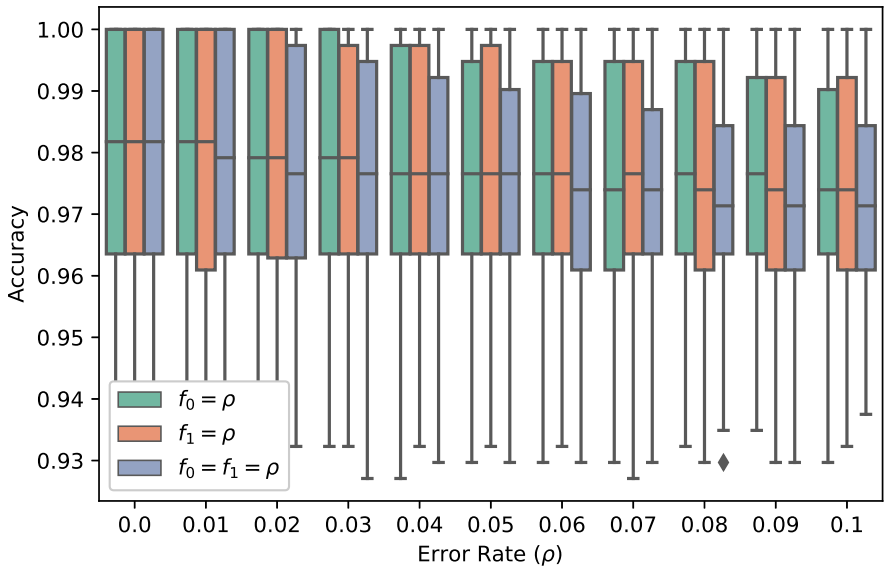

Fig. 4. Robustness against measurements errors.

3.2 Pooling Matrices

Optimal Pool Size In order to test the theoretically optimal pool size given in Equation 1, we simulated 1000 infection vectors under various infection rates and evaluated the performance of constant-pool matrix with $t = 48$ and $s \in \{8, 16, 24, \ldots, 80\}$. The performance of the pooling matrices is strongly correlated with the entropy of each pool, which supports the guidance of maximizing pool entropy in one-stage group testing designs (see Fig. 5).

In order to test the robustness of the optimal pool size s with respect to varying the number of pools t, we tested the performance of constant-pool designs with $s \in \{8, 16, 24, \ldots, 80\}$ based on 1000 simulated infection vectors under $f = 0.02$. The optimal pool size ($s = 32$) achieves nearly maximal recovery accuracy for all t (see Fig. 6).

Explicit Constructions Combinatorial structure of pooling matrices has been studied under the combinatorial framework [7]. Recent development of explicit pooling matrices construction for noisy group testing is based on error correcting codes [1]. Pooling design based on Reed-Solomon code was introduced in [14] and shown to be optimal under a more restricted probabilistic setting (assuming a random set of positives of a fixed size) [12].

Authors in [23] use a design based on Reed-Solomon error correcting code; the matrix assigns each individual in 6 pools such that each pool has 48 individuals. We compare the performance of the Reed-Solomon design, a randomly drawn

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

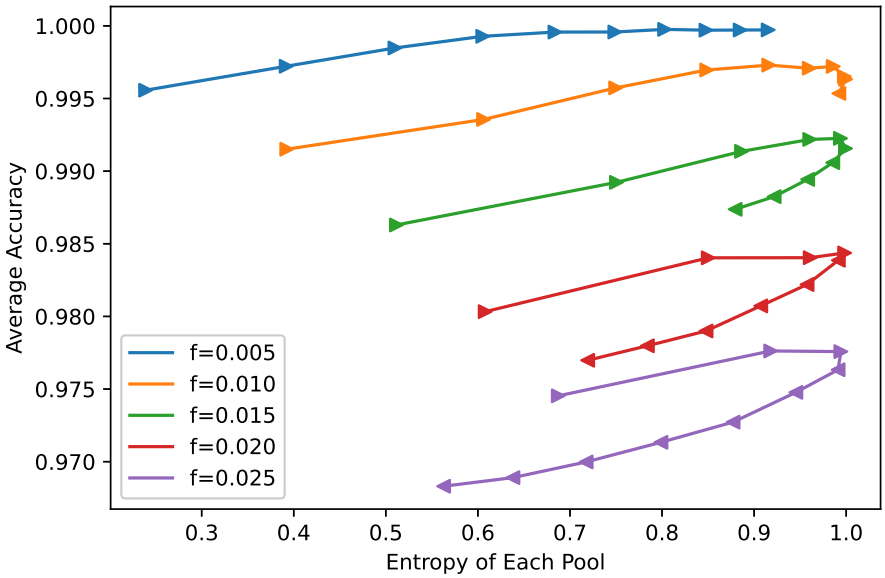

Fig. 5. Entropy of each pool against the performance of constant-pool designs with pool size $s \in \{8, 16, \ldots, 80\}$ and $t = 48$ pools; infection vectors are simulated with 384 Bernoulli(f) entries. The direction of the triangular markers indicates increasing s.

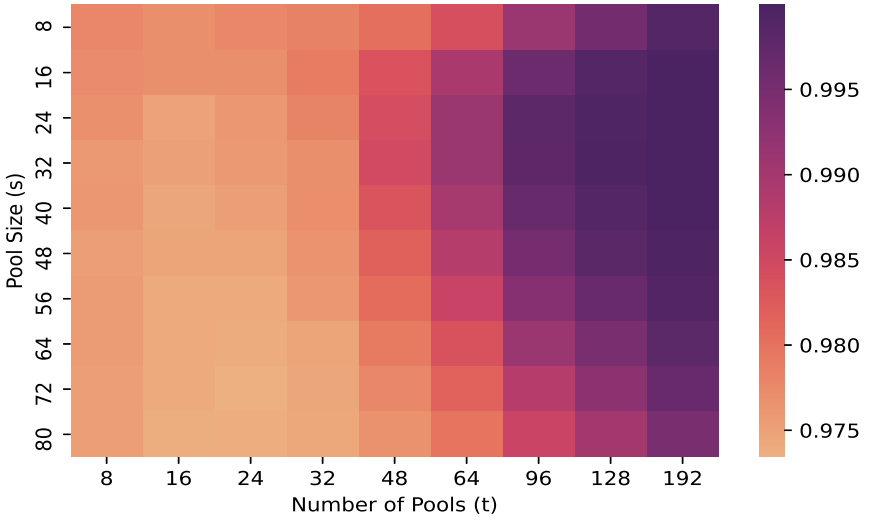

Fig. 6. Average recovery accuracy of designs with the pool size $s \in \{8, 16, 24, \ldots, 80\}$ and varying the number of pools under the infection rate $f = 0.02$; the pool size $s = 32$ achieves the maximum pool entropy.

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

doubly-regular design (with $r = 6$ and $s = 48$), a randomly drawn constant-redundancy design (with $r = 6$ and varying pool sizes), and a randomly drawn constant-pool design (with $s = 48$ and varying redundancies).

We found that under the probabilistic framework, the performance of the three random designs is comparable to that of the Reed-Solomon design in terms of both the average and the standard deviation of the recovery accuracy (see Fig. 7). For each of the three random designs, we computed the distribution of

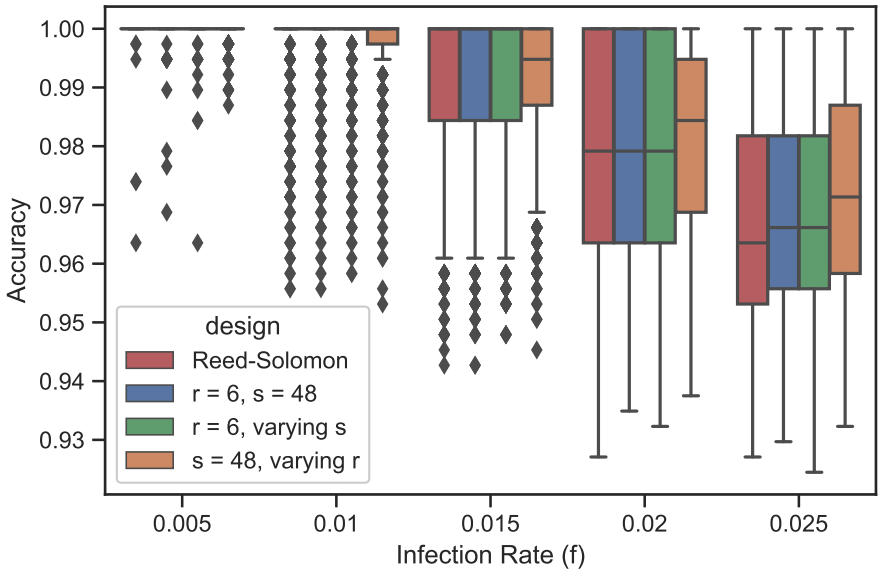

Fig. 7. Performance of the Reed-Solomon (red), the doubly-regular $r = 6, s = 48$ (blue), the constant redundancy $r = 6$ (green), and the constant pool $s = 48$ (orange) designs under simulated data with $n = 384$ individuals, $t = 48$ pools. The box shows the four quartiles, the whiskers show the rest of the distribution, and the points show the outliers.

the inner products of column vectors (which are derived from code words) of 1000 randomly-drawn matrices. The performance guarantees of Reed-Solomon design under the combinatorial setting follows from the minimum pairwise distance of the code words [1]. The authors in [17] studied explicit pooling matrices and showed the dependency of the error probability on both the minimum and the average distance under the assumption of a fixed number of positives. In the same spirit, we found that maximizing the number of pairs of column vectors with zero inner products could lead to performance guarantees under the probabilistic framework (see Fig. 8).

As a first step towards understanding the role of combinatorial structures of the pooling matrices under the probabilistic framework, we randomly drew

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

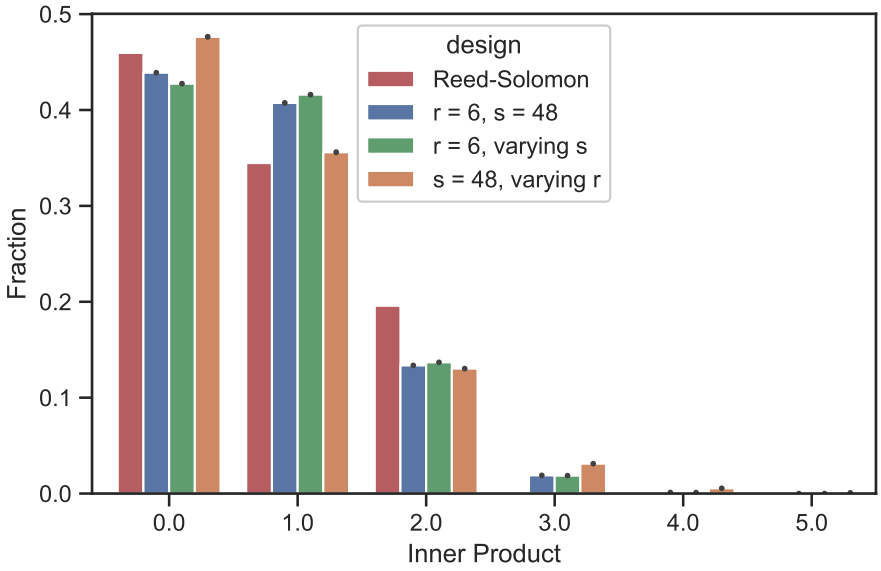

Fig. 8. Distributions of the inner products of column vector. Error bars are computed based on 1000 matrices.

1000 constant-pool designs with $s = 48$ to test $n = 384$ individuals with $t = 48$ pools, and computed the average accuracy on two sets of 1000 simulated infection vectors under the infection rate $f = 0.02$. We observe low correlation (≈ 0.0884; $p \approx 0.00516$) between the performance of designs on the two sets of vectors, which suggests that under the probabilistic framework, the combinatorial structure of the pooling matrices plays a less significant role than the parameters s, r, t.

4 Discussion

We presented a framework that addresses several practical issues, including randomness of the infection vector and noisy measurements. Our findings suggest that in practice, the combinatorial structure of pooling design matrices plays a lesser role compared to parameters such as redundancy, pool sizes, and the number of pools. Furthermore, we provide a protocol and an implementation for practitioners to choose the parameters in order to design pooling matrices.

From a theoretical perspective, there are several promising future directions for this work. The ILP problem presented in Section 2.3 can be viewed as a generalization of the smallest satisfying set (SSS) problem [1], which is NP-hard. One direction of future work is to design polynomial-time approximation recovery algorithm under the probabilistic framework.

In addition, non-asymptotic theoretical understanding of one-stage group testing under the probabilistic framework is still limited. An interesting future

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

research direction is to explore theoretical justifications on the effect of the number of pairs of column vectors with zero inner product on the performance, as observed in Section 3.

From a practical perspective, while no significant difference in performance among the four classes of designs is observed in the simulations performed in this study, future research should be conducted on further investigating the significance of the combinatorial structure of the pooling matrices under other scenarios such that parameter values with a higher range of recovery accuracy and sample sizes other than $n = 384$ that are germane to biological experiments.

As the COVID-19 pandemic continue to ravage large populations of resource limited nations, cost effective testing with simple designs is an absolute necessity for sustaining daily life in a safe manner.

References

1. Aldridge, M., Johnson, O., Scarlett, J.: Group testing: An information theory perspective. Foundations and Trends in Communications and Information Theory **15**(3-4), 196–392 (2019). https://doi.org/10.1561/0100000099
2. Atia, G., Saligrama, V.: Noisy group testing: An information theoretic perspective. In: 2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton). pp. 355–362. IEEE (2009). https://doi.org/10.1109/ALLERTON.2009.5394787
3. Atia, G.K., Saligrama, V.: Boolean compressed sensing and noisy group testing. IEEE Transactions on Information Theory **58**(3), 1880–1901 (2012). https://doi.org/10.1109/TIT.2011.2178156
4. Chan, C.L., Che, P.H., Jaggi, S., Saligrama, V.: Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms. In: 2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton). pp. 1832–1839. IEEE (2011). https://doi.org/10.1109/Allerton.2011.6120391
5. Deka, S., Kalita, D.: Effectiveness of Sample Pooling Strategies for SARS-CoV-2 Mass Screening by RT-PCR: A Scoping Review. Journal of Laboratory Physicians **12**(03), 212–218 (2020). https://doi.org/10.1055/s-0040-1721159
6. Du, D.: Pooling designs and nonadaptive group testing : important tools for DNA sequencing. World Scientific Publishing Company, New Jersey (2006)
7. Du, D., Hwang, F.: Combinatorial Group Testing and Its Applications. World Scientific Publishing Company, Singapore (1993)
8. Erlich, Y., Gilbert, A., Ngo, H., Rudra, A., Thierry-Mieg, N., Wootters, M., et al.: Biological screens from linear codes: theory and tools (2015), https://www.biorxiv.org/content/10.1101/035352v1.article-info, accessed 14 January 2021
9. Furon, T., Guyader, A., Cérou, F.: Decoding fingerprints using the Markov chain Monte Carlo method. In: 2012 IEEE International Workshop on Information Forensics and Security (WIFS). pp. 187–192 (2012). https://doi.org/10.1109/WIFS.2012.6412647
10. Ghosh, S., Agarwal, R., Rehan, M.A., Pathak, S., Agrawal, P., Gupta, Y., et al.: A Compressed Sensing Approach to Group-testing for COVID-19 Detection (2020), https://arxiv.org/abs/2005.07895, accessed 14 January 2021

medRxiv preprint doi: https://doi.org/10.1101/2021.03.09.21253193; this version posted March 12, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

11. Ghosh, S., Rajwade, A., Krishna, S., Gopalkrishnan, N., Schaus, T., Chakravarthy, A., et al.: Tapestry: A Single-Round Smart Pooling Technique for COVID-19 Testing (2020), https://www.medrxiv.org/content/10.1101/2020.04.23.20077727v2, accessed 14 January 2021
12. Inan, H.A., Kairouz, P., Wootters, M., Ozgur, A.: On the optimality of the kautz-singleton construction in probabilistic group testing. In: 2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton). pp. 188–195. IEEE (2018). https://doi.org/10.1109/ALLERTON.2018.8635972
13. Kainkaryam, R.M., Woolf, P.J.: Pooling in high-throughput drug screening. Current Opinion in Drug Discovery and Development **12**(3), 339–350 (2009)
14. Kautz, W.H., Singleton, R.C.: Nonrandom Binary Superimposed Codes. IEEE Transactions on Information Theory **10**(4), 363–377 (1964). https://doi.org/10.1109/TIT.1964.1053689
15. Liu, Y., Kadyan, S., Pe'er, I.: Group testing under the probabilistic framework code, https://github.com/imyiningliu/group-testing-probabilistic-framework, accesed 28 February 2021
16. Malioutov, D., Malyutov, M.: Boolean compressed sensing: LP relaxation for group testing. In: 2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). pp. 3305–3308. IEEE (2012). https://doi.org/10.1109/ICASSP.2012.6288622
17. Mazumdar, A.: Nonadaptive group testing with random set of defectives. IEEE Transactions on Information Theory **62**(12), 7522–7531 (2016). https://doi.org/10.1109/TIT.2016.2613870
18. Scarlett, J., Cevher, V.: Near-Optimal Noisy Group Testing via Separate Decoding of Items. In: 2018 IEEE International Symposium on Information Theory (ISIT). pp. 2311–2315. IEEE (2018). https://doi.org/10.1109/ISIT.2018.8437667
19. Scarlett, J., Johnson, O.: Noisy non-Adaptive group testing: A (near-)definite defectives approach. IEEE Transactions on Information Theory **66**(6), 3775–3797 (2020). https://doi.org/10.1109/TIT.2020.2970184
20. Sejdinovic, D., Johnson, O.: Note on noisy group testing: Asymptotic bounds and belief propagation reconstruction. In: 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton). pp. 998–1003. IEEE (2010). https://doi.org/10.1109/ALLERTON.2010.5707018
21. Sham, P., Bader, J.S., Craig, I., O'Donovan, M., Owen, M.: DNA pooling: A tool for large-scale association studies. Nature Reviews Genetics **3**(11), 862–871 (2002). https://doi.org/10.1038/nrg930
22. Shani-Narkiss, H., Gilday, O.D., Yayon, N., Landau, I.D.: Efficient and Practical Sample Pooling for High-Throughput PCR Diagnosis of COVID-19 (2020), https://www.medrxiv.org/content/10.1101/2020.04.06.20052159v2, accessed 14 January 2021
23. Shental, N., Levy, S., Wuvshet, V., Skorniakov, S., Shemer-Avni, Y., Porgador, A., et al.: Efficient high throughput SARS-CoV-2 testing to detect asymptomatic carriers. Science Advances **6**(37) (2020). https://doi.org/10.1126/sciadv.abc5961
24. Täufer, M.: Rapid, large-scale, and effective detection of COVID-19 via non-adaptive testing. Journal of Theoretical Biology **506**(110450) (2020). https://doi.org/10.1016/j.jtbi.2020.110450, http://www.sciencedirect.com/science/article/pii/S0022519320303052
25. Wu, S., Wei, S., Wang, Y., Vaidyanathan, R., Yuan, J.: Achievable partition information rate over noisy multi-access Boolean channel. In: IEEE International Symposium on Information Theory - Proceedings. pp. 1206–1210. IEEE (2014). https://doi.org/10.1109/ISIT.2014.6875024