

Constructions for Nonadaptive Tropical Group Testing

Nicholas Kwan and Lele Wang

University of British Columbia, Vancouver, BC V6T1Z4, Canada

`nickwan@student.ubc.ca`, `lelewang@ece.ubc.ca`

Abstract—PCR testing is an invaluable diagnostic tool that has most recently seen widespread use during the COVID-19 pandemic. A recent work by Wang, Gabrys and Vardy proposed tropical codes as a model for group PCR testing. For a known but arbitrary number of infected persons, a sufficient condition on the underlying block design of a zero-error tropical code, called double disjunction, is proposed. Despite this, the parameters for which the construction of doubly disjunct block designs is known to exist are very limited. In this paper, we define probabilistic tropical codes and consider random block designs that are doubly disjunct with high probability. We also provide a deterministic construction for a doubly disjunct block design given a disjunct block design. We show that for certain choices of parameters, our probabilistic construction has vanishing error. Our constructions, combined with existing methods, give us three different ways to construct tropical codes. We compare the number of tests required by each, and bounds on the error.

I. INTRODUCTION

Polymerase Chain Reaction (PCR) is a method by which genetic material may be rapidly amplified. PCR has a wide range of applications, perhaps most recently in reliable testing for the virus causing COVID-19. PCR testing is typically done through thermal cycling, where the viral load of a test containing some specimens is repeatedly doubled in concentration. Each doubling event is known as a cycle, and the load is doubled until it reaches a detectable threshold or fails to reach this threshold within a preset number of cycles. The cycle number at which this threshold is reached is called the *Ct value*. The Ct value of a test is thus

$$\text{Ct value} = \lfloor -\log_2(\text{viral load}) + \text{constant} \rfloor,$$

and is a measure of the degree of infection of a test, with smaller values indicating higher degrees of infection. If a test fails to reach the threshold after a preset number of cycles, then it is not infected and its Ct-value is ∞ .

Consider some finite set S of people with an infected subset $\mathcal{D} \subset S$. Group testing deals with the problem of identifying $\mathcal{D}$ by performing tests on subsets of S whose size may be greater than one. By nonadaptive group testing, we consider a setting where all group tests are given at once. Group testing can be divided into combinatorial and probabilistic group testing. Combinatorial group testing identifies $\mathcal{D}$ with zero error, while probabilistic group testing identifies $\mathcal{D}$ with high probability. Traditionally, group testing takes place in the binary setting so that a test on some subset $\mathcal{U} \subset S$ is positive when $\mathcal{U} \cap \mathcal{D}$ is nonempty and negative otherwise. On the other hand, a PCR

test on a subset $\mathcal{U}$ returns some aggregate Ct-value, which tells us more than whether or not $\mathcal{U} \cap \mathcal{D}$ is nonempty. For instance, when two specimens with Ct values x and y are placed into a PCR test at the same time, the resulting Ct value is z satisfying

$$2^{-x} + 2^{-y} = 2^{-z}.$$

Due to exponential decay, we have $z \approx \min\{x, y\}$. In performing a group PCR test, we are also able to add a specimen to a test after a certain number of cycles have elapsed. The operation of adding a specimen to a test after some δ cycles is known as delaying, and we say that the specimen has been delayed by δ cycles. When a specimen with Ct value x is delayed by δ cycles, the number of cycles it takes to reach the same detectable threshold becomes $x + \delta$.

Based on the above properties of PCR testing, Wang, Gabrys, and Vardy proposed the tropical semiring as a model for group PCR testing [1]. The tropical semiring $(\mathbb{R} \cup \{\infty\}, \oplus, \odot)$ consists of the real numbers and infinity equipped with the following operations:

$$\begin{aligned} x \oplus y &:= \min(x, y), \\ x \odot y &:= x + y. \end{aligned}$$

The additive identity is ∞ and the multiplicative identity is 0. In this model, for two specimens x and y (we have identified the specimens with their individual Ct-values), the Ct-value $(x \odot \delta_x) \oplus (y \odot \delta_y) = \min(x + \delta_x, y + \delta_y)$ is given by a test containing both x and y , delayed each by δ_x and δ_y respectively. In this way, $\odot$ indicates delay and $\oplus$ indicates presence. This model works because viral loads vary widely, and the Ct-value of a test with two specimens with differing viral loads is dominated by the specimen with the higher viral load due to the exponential nature of the test. Since Ct values and delays only take nonnegative integer values, all the analysis done in this paper is over the subsemiring $(\mathbb{N} \cup \{0, \infty\}, \oplus, \odot)$.

In general, a test involving specimens $x_1, \dots, x_k$ with delays $\delta_1, \dots, \delta_k$ will return the Ct-value $\bigoplus_{i=1}^k (x_i \odot \delta_i)$. In this way, we can define tests using tropical matrix multiplication. Consider a set of n individuals, and let $x = [x_1, x_2, \dots, x_n]^\top \in (\mathbb{N} \cup \{0, \infty\})^n$ be a list of their Ct values. A set of t tests on these individuals can be formally written as a matrix of delay values $S \in (\mathbb{N} \cup \{0, \infty\})^{t \times n}$, known as the schedule matrix. If $S_{ij} = \infty$, then the i^{th} test did not involve the j^{th} individual. The results of the tests are then given by

$S \odot x$, where here, $\odot$ denotes tropical matrix multiplication; that is,

$$S \odot x = \begin{bmatrix} \bigoplus_{j=1}^N (S_{1j} \odot x_j) \\ \vdots \\ \bigoplus_{j=1}^N (S_{tj} \odot x_j) \end{bmatrix}.$$

If, for any distinct $x, y \in (\mathbb{N} \cup \{0, \infty\})^n$ with at most d finite values, $S \odot x \neq S \odot y$, then our schedule matrix is a valid (t, n, d) -tropical code in the combinatorial setting. The code is said to have maximum delay l , where l is the largest finite element in S . In a real-world setting, l is typically held at 40 cycles. For general d , zero-error tropical codes are hard to construct.

Existing work in nonadaptive group testing considers the construction of binary matrices satisfying certain properties, such as disjunction and separability, that correspond to binary group testing schemes. These properties are necessary, but not sufficient, for binary matrices that correspond to the schedule matrix of a tropical code. Furthermore, the sufficient doubly disjunct property given in definition 1 by [1] is new, and to our knowledge, existing constructions do not exist for doubly disjunct matrices. In this paper, we propose the construction of *random* schedule matrices that are doubly disjunct with high probability, as well as the construction of doubly disjunct matrices from disjunct matrices.

Notation. Throughout we use calligraphic letters such as $\mathcal{K}$ to refer to sets and block designs. We use boldfaced letters for random matrices $\mathbf{S}$ and random vectors $\mathbf{s}$ when they are capitalized and in lowercase, respectively. $\mathbf{P}(A)$ refers to the probability of the event A , and we define $p_X(x) \triangleq \mathbf{P}(X = x)$. Plain capital letters are used for matrices and blocks, and $\vec{p}$ is used for vectors.

II. SUMMARY OF EXISTING RESULTS

This paper relies heavily on the results given by [1], which was the first paper to consider using the tropical semiring as a model for group PCR testing, and produced results on tropical codes in the combinatorial setting.

For some $t \in \mathbb{N}$, a block design on t vertices is a subset $\mathcal{F} \subset 2^{[t]}$ of the power set $2^{[t]}$ satisfying certain properties with respect to unions, intersections, and differences. Each element $B \in \mathcal{F}$ is called a block. The set operation properties satisfied by the blocks vary between block designs. An example of a block design on 7 vertices is the Fano plane $\mathcal{F} = \{\{1, 2, 4\}, \{1, 3, 7\}, \{1, 5, 6\}, \{2, 3, 5\}, \{2, 6, 7\}, \{3, 4, 6\}, \{4, 5, 7\}\}$. Note that the intersection of any two blocks has exactly one element, and that each block has size three. Furthermore, each element of $[7]$ is contained in exactly three blocks.

We may associate a block design $\mathcal{F}$ on T vertices with $|\mathcal{F}| = N$ with an incidence matrix M , a $T \times N$ binary matrix where each column corresponds to a block. Then

$$M_{ij} = \begin{cases} 0 & \text{if } i \notin B_j \text{ for the block } B_j \in \mathcal{F}, \\ 1 & \text{if } i \in B_j \text{ for the block } B_j \in \mathcal{F}. \end{cases}$$

Note that this representation is unique up to permutation of the columns M , whereas the permutation of the rows of M results in an equivalent block design. For instance, the incidence matrix of the previously described Fano plane is

$$\begin{bmatrix} 1 & 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \end{bmatrix}.$$

We may use the incidence matrices of suitable block designs to construct tropical codes by replacing each 0 with ∞ and each 1 with a suitable finite delay value. Conversely, every tropical code has an associated block design.

Definition 1 (Doubly Disjunct Block Design). A block design $\mathcal{F}$ with n blocks on t vertices is said to be d -doubly disjunct if, for any distinct blocks $Z, B_1, \dots, B_d \in \mathcal{F}$, $|Z \setminus (\bigcup_{i=1}^d B_i)| \geq 2$.

If instead $|Z \setminus (\bigcup_{i=1}^d B_i)| \geq 1$, then we call $\mathcal{F}$ d -disjunct.

Remark. Clearly if $\mathcal{F}$ is a d -doubly disjunct block design, it is also d -disjunct.

The Fano plane is a 1-doubly disjunct block design on 7 vertices. A construction for 1-doubly disjunct block design on some arbitrary number of vertices t is given by Graham and Sloane [2]. Let $\mathcal{S}_w^t$ be the set of all binary vectors of length t and Hamming weight w , consider the map

$$f : \mathcal{S}_w^t \longrightarrow \mathbb{Z}/t\mathbb{Z}, \\ \vec{v} \longmapsto \left(\sum_{i=1}^n i v_i \right) \mod t,$$

taking a binary vector v to the sum of the indices of its ones, modulo t . Then the preimage of any residue class will be a 1-doubly disjunct block design. At least one of these preimages will have cardinality larger than $\frac{1}{t} \binom{t}{w}$.

Theorem 1 (Theorem 31 in [1]). Let $\mathcal{F}$ be a $(d-1)$ -doubly-disjunct block design with n blocks on t vertices. Let M be the incidence matrix of $\mathcal{F}$. Then there exists a (t, n, d) -tropical code whose associated block design is $\mathcal{F}$. That is, the finite entries of the schedule matrix S coincide exactly with the nonzero entries of M .

In the case where $d = 2$ and each block $B \in \mathcal{F}$ has size at least three, there exists a decoding algorithm, given by Theorem 17 in [1], whose maximum delay is a prime bounded above by n . Otherwise, for $d > 2$, the decoding algorithm given in Theorem 31 in [1] uses a maximum delay of $2^{n(t+1)}$, with the schedule matrix

$$S_{ij} = \begin{cases} 2^{i+jt}, & \text{if } M_{ij} = 1, \\ \infty, & \text{if } M_{ij} = 0. \end{cases}$$

although the authors conjecture that there is a polynomial delay code using the same underlying block design. The given decoding algorithm considers sets of potentially infected persons of size d . We may construct a bipartite graph using a plausible set of d infected persons and the actual set of d infected persons as our two sets of vertices. The double disjointness of $\mathcal{F}$ guarantees that this graph contains an even cycle. The alternating sum along this cycle will correspond to the sum of delays for the actual set of infected persons.

To the best of our knowledge, there is no construction for a $(d-1)$ -doubly disjoint block design for $d > 2$ with arbitrary parameters. Existing constructions for d -disjoint matrices, which are suitable for binary group testing, exist. Combinatorial constructions exist with $t = O(d^2 \log n)$ [3], while Monte-Carlo constructions, like our own, exist for $O(d^2 \min\{\log n, (\log_t n)^2\})$ [4]. Probabilistic binary group testing schemes exist with $t = \Theta(d \log n)$ [5].

III. RESULT I: A DETERMINISTIC CONSTRUCTION FROM DISJUNCT BLOCK DESIGNS

Recall from Definition 1 what it means for a block design $\mathcal{F}$ to be disjoint and doubly disjoint. Our first result provides a simple way to construct d -doubly disjoint block designs from d -disjoint block designs by doubling the number of vertices. These can then be used to produce (t, n, d) -tropical codes by Theorem 1. Methods for constructing d -disjoint block designs are given in [3]–[5].

Theorem 2. *Let $\mathcal{F} \subset 2^{[t]}$ be a d -disjoint block design on t vertices. Then the image $\mathcal{G}$ of the function*

$$f : \mathcal{F} \longrightarrow 2^{[2t]}$$

$$B \longmapsto \{2x : x \in B\} \cup \{2x-1 : x \in B\},$$

is a d -doubly disjoint block design on $2t$ vertices.

Proof: Let $B_0, \dots, B_d$ be any $d+1$ distinct blocks in $\mathcal{G}$, and for each $i \in \{0, \dots, d\}$ let A_i be the unique element of $\mathcal{F}$ mapping to B_i . Since $\mathcal{F}$ is disjoint, $|A_0 \setminus (\bigcup_{i=1}^d A_i)| \geq 1$. For each block, the maps $x \mapsto 2x$ and $x \mapsto 2x-1$ are injective, and the images of the two maps are disjoint. Let $a \in A_0 \setminus (\bigcup_{i=1}^d A_i)$. Then note that $2a \in B_0$ and $2a-1 \in B_0$. If $2a \in \bigcup_{i=1}^d B_i$, then $2a \in B_i$ for some $i \geq 1$, but this would imply that $a \in A_i$ for some $i \geq 1$, which cannot be the case. Similarly, if $2a-1 \in \bigcup_{i=1}^d B_i$, then $2a-1 \in B_i$ for some $i \geq 1$, but this would imply that $a \in A_i$ for some $i \geq 1$, which cannot be the case. Hence, $2a \in B_0 \setminus (\bigcup_{i=1}^d B_i)$ and $2a-1 \in B_0 \setminus (\bigcup_{i=1}^d B_i)$, so $|B_0 \setminus (\bigcup_{i=1}^d B_i)| \geq 2$ and $\mathcal{G}$ is d -doubly disjoint. ■

Remark. This construction extends to the construction of n -fold disjoint block designs, but it is not optimal. For instance, consider the Fano plane as described in section II. The Fano plane is an “optimal” 2-disjunctive block design on 7 vertices in the sense that no blocks can be added to it while preserving 2-disjunctiveness, but applying Theorem 2 to the Fano plane

and adding the block $\{2, 4, 6, 8, 10, 12, 14\}$ still yields a 2-doubly disjoint block design on 14 vertices.

IV. RESULT II: A PROBABILISTIC CONSTRUCTION

In this section we define tropical codes in a probabilistic setting, and construct random matrices that are doubly disjoint with high probability.

A. Definition

Consider a random $t \times n$ schedule matrix $\mathbf{S}$. We define the probability of error as

$$P_e^{n,d}(\mathbf{S}) \triangleq \mathbb{P}(\mathbf{S} \odot \vec{x} = \mathbf{S} \odot \vec{y} \text{ for some distinct } \vec{x}, \vec{y} \in \mathcal{R}_{n,d}),$$

where $\mathcal{R}_{n,d} \triangleq \{\vec{x} \in (\mathbb{N} \cup \{0, \infty\})^n : x \text{ contains at most } d \text{ finite values}\}$. In other words, $P_e^{n,d}(\mathbf{S})$ is the probability that a group testing scheme with schedule matrix $\mathbf{S}$ confuses x and y . To formalize the definition of a tropical code in a probabilistic setting, we introduce a parameter ϵ to bound the probability of error $P_e^{n,d}(\mathbf{S})$.

Definition 2. A random $t \times n$ schedule matrix $\mathbf{S}$ is a (t, n, d, ϵ) -tropical code if $P_e^{n,d} \leq \epsilon$.

A $(t, n, d, 0)$ -tropical code then corresponds to the definition of a (t, n, d) -tropical code given in [1]. This definition naturally motivates us to find relationships between the parameters t, n, d and ϵ .

B. Construction

In this section we find parameters (t, n, d) for which we are able to construct doubly disjoint matrices with high probability, then compute ϵ for our construction, based on its probability of generating a matrix that is not doubly disjoint.

We show the existence of a joint distribution $p_s(\vec{v})$ for the first row of delays $\mathbf{s}$ in $\mathbf{S}$. We then generate the remaining rows of $\mathbf{S}$ i.i.d. according to the same distribution $p_s(\vec{v})$.

We produce sufficient properties on $p_s(\vec{v})$ for $\mathbf{S}$ to be doubly disjoint with high probability. Since $p_s(\vec{v})$ is a probability distribution, we require

$$\sum_{\vec{v} \in \{0,1\}^n} p_s(\vec{v}) = 1, \quad (1)$$

and

$$p_s(\vec{v}) \geq 0 \text{ for all } \vec{v} \in \{0,1\}^n. \quad (2)$$

For $\mathbf{S}$ to be $d-1$ -doubly disjoint on average, we require, for every distinct $i_1, i_2, \dots, i_d \in [n]$

$$\mathbb{P}\left(\begin{matrix} s_{i_1} = 1 \text{ and} \\ s_{i_2} = \dots = s_{i_d} = 0 \end{matrix}\right) = \sum_{\substack{\vec{v} \in \{0,1\}^n \\ v_{i_1}=1, \\ v_{i_2}=\dots=v_{i_d}=0}} p_s(\vec{v}) = \frac{2+\Delta}{t}. \quad (3)$$

Additionally, to ensure that no test is empty, we set $P(\mathbf{s} = \vec{0}) = 0$. Conditions (1), (2), and (3) reduce our problem to proving the existence of non-negative solutions to a linear system. To simplify our analysis, we restrict to the case where rows whose Hamming weights are equal have equal

probability of being generated. For instance, if $n = 3$, then we let $p_s(100) = p_s(010) = p_s(001)$ and $p_s(110) = p_s(101) = p_s(011)$. This allows us to write our first conditions (1) and (3) as linear combinations of the probabilities that s has hamming weight $w \leq n$, rather than the probabilities of $p_s(\vec{v})$ for each $\vec{v} \in \{0, 1\}^n$. Let $\vec{p}$ be the vector whose w^{th} element corresponds to the probability that $\vec{s}$ has hamming weight w . Conditions (1) and (3) can then be written as the following linear system, for which $\vec{p}$ is a solution

$$\begin{bmatrix} \binom{n}{1} & \binom{n}{2} & \cdots & \binom{n}{n-d+1} & \cdots & \binom{n}{n} \\ \binom{n-d}{0} & \binom{n-d}{1} & \cdots & \binom{n-d}{n-d} & 0 & \cdots & 0 \end{bmatrix} \vec{p} = \begin{bmatrix} 1 \\ \frac{2+\Delta}{t} \end{bmatrix}. \quad (4)$$

The first row of the matrix in equation (4) corresponds to the condition given by (1); its w^{th} component is the number of length n binary vectors with weight w . The second row corresponds to the condition given by (3); its w^{th} component is the number of weight w length n binary vectors s with $s_{i_1} = 1$ and $s_{i_2} = \dots = s_{i_d} = 0$ for any distinct $i_1, \dots, i_d \in [n]$.

The following lemma provides conditions on the parameters such that an elementwise nonnegative solution to the linear system (4) exists.

Lemma 1. *For a general pool of n patients with a maximum of d infected, and any $\Delta > 0$, if*

$$\frac{\binom{n-d}{\lfloor \frac{n}{d} \rfloor - 1}}{\binom{n}{\lfloor \frac{n}{d} \rfloor}} > \frac{2 + \Delta}{t},$$

then equation (4) has an elementwise nonnegative solution.

We will use Farkas' Lemma to establish Lemma 1.

Lemma 2 (Farkas' Lemma, [6, p. 263]). *Let $A \in \mathbb{R}^{m \times n}$ and $\vec{b} \in \mathbb{R}^m$. Then exactly one of the following is true:*

- 1) *There exists $\vec{p} \in \mathbb{R}^n$ such that $A\vec{p} = \vec{b}$ and $\vec{p} \geq 0$ elementwise*
- 2) *There exists $\vec{y} \in \mathbb{R}^m$ such that $A^T \vec{y} \geq 0$ elementwise and $\vec{b}^T \vec{y} < 0$*

Proof of Lemma 1: If no elementwise nonnegative vector $\vec{p}$ exists satisfying (4), then by the second condition of Farkas' lemma, there exists some $\vec{y} = [y_1 \ y_2]^T$ such that $y_1 \geq 0$, $y_1 + \frac{2+\Delta}{t}y_2 < 0$, and $\binom{n-d}{w}y_1 + \binom{n-d}{w-1}y_2 \geq 0$ for all $w = 1, 2, \dots, n-d+1$. Together, these imply

$$\begin{aligned} 0 &> y_1 + \frac{2+\Delta}{t}y_2 \\ &\geq \frac{-\binom{n-d}{w-1}y_2}{\binom{n}{w}} + \frac{2+\Delta}{t}y_2 \\ &= y_2 \left(\frac{2+\Delta}{t} - \frac{\binom{n-d}{w-1}}{\binom{n}{w}} \right), \end{aligned}$$

for all $w = 1, 2, \dots, n-d+1$. Note that, since $y_1 \geq 0$ and $y_1 + \frac{2+\Delta}{t}y_2 < 0$, we must have $y_2 < 0$. This means

$$\frac{\binom{n-d}{w-1}}{\binom{n}{w}} < \frac{2+\Delta}{t}.$$

We compute the maximum of $f(w) \triangleq \frac{\binom{n-d}{w-1}}{\binom{n}{w}}$ over $w \in [d]$. Note that

$$f(w) = \frac{w \prod_{i=0}^{w-2} (n-d-i)}{\prod_{i=0}^{w-1} (n-i)},$$

so that

$$f(w+1) = \frac{(w+1)(n-d-w+1) \prod_{i=0}^{w-2} (n-d-i)}{(n-w) \prod_{i=0}^{w-1} (n-i)}.$$

We examine the terms outside of the products in the numerator and denominator, which differ between $f(w)$ and $f(w+1)$. For $f(w+1)$, we have

$$\frac{(w+1)(n-d-w+1)}{n-w} = w \left(1 + \frac{1}{w} \right) \left(1 - \frac{d-1}{n-w} \right).$$

If $w \geq \lfloor n/d \rfloor$ is an integer, then $wd \geq n$ so that

$$\begin{aligned} w(d-1) &\geq n-w \\ \frac{d-1}{n-w} &\geq \frac{1}{w}, \end{aligned}$$

and $f(w+1) \leq f(w)$. If $w < \lfloor n/d \rfloor$ is an integer, then $w \leq \frac{n}{d} - 1$ so that

$$\begin{aligned} n-w &\geq d(w+1) - w \\ &> (d-1)(w+1) \\ \frac{1}{w} &> \left(\frac{d-1}{n-w} \right) \left(1 + \frac{1}{w} \right), \end{aligned}$$

and $f(w+1) > f(w)$. As such, f is maximized when $w = \lfloor n/d \rfloor$. Hence, if

$$\frac{\binom{n-d}{\lfloor \frac{n}{d} \rfloor - 1}}{\binom{n}{\lfloor \frac{n}{d} \rfloor}} > \frac{2 + \Delta}{t},$$

then there exists some elementwise non-negative vector $\vec{p}$ satisfying (4). Thus, there exists a valid probability distribution on s , using which we can generate a (t, n, d) -tropical code with high probability. ■

C. Performance of the probabilistic construction

In this section, we establish conditions for which the proposed construction provides us with a valid tropical code.

Theorem 3. *The construction given in Section IV-B is a (t, n, d, ϵ) -tropical code if*

$$t > 2e \left(d^2 \log \left(\frac{ne}{d} \right) + d \log \left(\frac{d}{\epsilon} \right) + 2d \right).$$

We will use Chernoff bound to establish Theorem 3.

Lemma 3 (Chernoff bound [7]). *Let $X = \sum_{k=1}^t X_k$ where all $X_k \sim \text{Bern}(p)$ are independent. Let $\mu = \mathbb{E}(X) = tp$. Then*

$$\mathbb{P}(X \leq (1-\delta)\mu) \leq e^{-\mu\delta^2/2},$$

for all $0 < \delta < 1$.

Proof of Theorem 3: For a given $\epsilon \in (0, 1)$, choose $\Delta = (2 + 2 \log(\frac{d}{\epsilon} \binom{n}{d}))$. For some d fixed distinct blocks indexed $i_1, \dots, i_d \in [n]$ let $X_k = \mathbf{1}_{\{s_{k,i_1}=1, s_{k,i_2}=0, \dots, s_{k,i_d}=0\}} \sim$

$\text{Bern}(\frac{4+2\log(\frac{d}{\epsilon} \binom{n}{d})}{t})$ be the random variable indicating whether or not the i_1^{th} entry of the k^{th} row is 1 and the $i_2, \dots, i_d$ entries of the k^{th} row are 0. Then the probability of error for the chosen subset of blocks is $\mathbb{P}(\sum_{k \in [t]} X_k < 2)$, since the sum $\sum_{k \in [t]} X_k$ gives the size of the set $|B_{i_1} \setminus (\bigcup_{j=2}^d B_{i_j})|$. Then note that

$$\begin{aligned} & \mathbb{P}\left(\sum_{k \in [t]} X_k < 2\right) \\ &= \mathbb{P}\left(\sum_{k \in [t]} X_k < \left(4 + 2\log\left(\frac{d}{\epsilon} \binom{n}{d}\right)\right) \left(1 - \frac{2+2\log(\frac{d}{\epsilon} \binom{n}{d})}{4+2\log(\frac{d}{\epsilon} \binom{n}{d})}\right)\right). \end{aligned}$$

We may then use the Chernoff bound (lemma 3) with $\mu = 4 + 2\log(\frac{d}{\epsilon} \binom{n}{d})$ and $\delta = \frac{2+2\log(\frac{d}{\epsilon} \binom{n}{d})}{4+2\log(\frac{d}{\epsilon} \binom{n}{d})}$ to get:

$$\begin{aligned} & \mathbb{P}\left(\sum_{k \in [t]} X_k < 2\right) \\ &\leq \exp\left(-\left(4 + 2\log\left(\frac{d}{\epsilon} \binom{n}{d}\right)\right) \frac{\left(2+2\log\left(\frac{d}{\epsilon} \binom{n}{d}\right)\right)^2}{2\left(4+2\log\left(\frac{d}{\epsilon} \binom{n}{d}\right)\right)^2}\right) \\ &= \exp\left(-\frac{\left(2 + 2\log\left(\frac{d}{\epsilon} \binom{n}{d}\right)\right)^2}{2\left(4 + 2\log\left(\frac{d}{\epsilon} \binom{n}{d}\right)\right)}\right). \end{aligned}$$

Now we take the union bound over every subset of $[n]$ with cardinality d with one specific index chosen. This gives an upper bound for $P_e^{n,d}$.

$$\begin{aligned} P_e^{n,d} &\leq \sum_{\substack{I \subset [n], |I|=d \\ i_1 \in I}} \mathbb{P}\left(\sum_{k \in [t]} X_k < 2\right) \\ &\leq d \binom{n}{d} \exp\left(-\frac{\left(2 + 2\log\left(\frac{d}{\epsilon} \binom{n}{d}\right)\right)^2}{2\left(4 + 2\log\left(\frac{d}{\epsilon} \binom{n}{d}\right)\right)}\right) \\ &\leq d \binom{n}{d} \exp(-1 - \log(\frac{d}{\epsilon} \binom{n}{d})) \\ &\leq d \binom{n}{d} \exp\left(\log\left(\frac{\epsilon}{d \binom{n}{d}}\right)\right) \\ &= \epsilon \end{aligned}$$

To study the asymptotic behaviour of the construction in Lemma 1, we reduce our final expression using the following asymptotic approximation of the binomial coefficient

$$\binom{n}{\alpha n} \approx \frac{1}{\sqrt{2\pi\alpha(1-\alpha)n}} 2^{nh(\alpha)},$$

where $h(\cdot)$ is the binary entropy function. This gives

$$\frac{\binom{n}{\lfloor \frac{n}{d} \rfloor}}{\binom{n-d}{\lfloor \frac{n}{d} \rfloor - 1}} \approx \sqrt{\frac{n-d}{n}} 2^{dh(1/d)} < \sqrt{\frac{n-d}{n}} ed,$$

so that our construction is a (t, n, d, ϵ) -tropical code if

$$t > \sqrt{\frac{n-d}{n}} ed \left(4 + 2\log\left(\frac{d}{\epsilon} \binom{n}{d}\right)\right). \quad (5)$$

Using the bounds $\sqrt{\frac{n-d}{n}} < 1$ and $\binom{n}{d} \leq (\frac{en}{d})^d$ we can relax the condition in (5) so that our construction is a (t, n, d, ϵ) -tropical code if

$$t > 2e \left(d^2 \log\left(\frac{ne}{d}\right) + d \log\left(\frac{d}{\epsilon}\right) + 2d\right). \quad (6)$$

■

V. COMPARISON

Since both probabilistic and deterministic constructions exist for disjunct matrices, and our result in Theorem 1 provides a construction for a $(d-1)$ -doubly disjunct block design from a $(d-1)$ -disjunct block design, we have three ways to produce (t, n, d, ϵ) -tropical codes for some fixed d :

- 1) First construct a deterministic $(d-1)$ -disjunct matrix, then use Theorem 2 to construct a $(d-1)$ -doubly disjunct matrix;
- 2) Construct a $(d-1)$ -disjunct matrix probabilistically, then use Theorem 2 to construct a $(d-1)$ -doubly disjunct matrix;
- 3) Construct a $(d-1)$ -doubly disjunct random matrix directly using Lemma 1.

Theorem 2 doubles the number of tests used, but does not change the asymptotic behaviour of the construction used for a $(d-1)$ -disjunct matrix. The deterministic constructions given by [3], [8] have $t = O(d^2 \log n)$ with $\epsilon = 0$ for specific parameter choices. As such, the first method also gives $t = O(d^2 \log n)$ with $\epsilon = 0$. For the second method, we can use our construction detailed in Section IV-B to construct $(d-1)$ -disjunct matrices by replacing the right hand side of Lemma 1 and all subsequent calculations with $\frac{1+\Delta}{t}$. This gives $(d-1)$ -disjunct matrices with $t > e(2d^2 \log(\frac{ne}{d}) + 2d \log(\frac{d}{\epsilon}) + 3d)$ for error bound ϵ . Then, using Theorem 2, we can get a $(d-1)$ -doubly disjunct matrix with $t > 2e(2d^2 \log(\frac{ne}{d}) + 2d \log(\frac{d}{\epsilon}) + 3d)$, which is greater than the bound in Theorem 3 for the construction of a $(d-1)$ -doubly disjunct matrix directly, as in the third method. From equation (5) in Theorem 3, the construction given in Section IV-B requires $t = O(d \log(\frac{n}{d}) + d \log(\frac{d}{\epsilon}))$. When $d = o(n)$ is sublinear, then we have $t = O(d^2 \log n + d \log(d/\epsilon))$. If $d = \Theta(n)$, then $t = o(d^2 \log n + d \log(d/\epsilon))$.

In a practical setting, if the situation requires zero error, then we can use the first method to construct a doubly disjunct matrix with the set of given parameters. Otherwise, if some error ϵ is tolerable, then our construction in Section IV-B may be preferable.

VI. CONCLUSION

In this paper we consider a probabilistic construction for a $(d-1)$ -doubly disjunct block design, as well as a deterministic method based on the existence of a $(d-1)$ -disjunct block design. We show that constructing a $(d-1)$ -doubly-disjunct block design probabilistically gives a similar performance to a fully deterministic construction based on a $(d-1)$ -disjunct block design, with the added benefit of being able to use arbitrary parameters.

REFERENCES

- [1] H.-P. Wang, R. Gabrys, and A. Vardy, "Tropical group testing," *IEEE Transactions on Information Theory*, vol. 69, no. 9, pp. 6098–6120, 2023.
- [2] R. Graham and N. Sloane, "Lower bounds for constant weight codes," *IEEE Transactions on Information Theory*, vol. 26, no. 1, pp. 37–43, 1980.
- [3] W. Kautz and R. Singleton, "Nonrandom binary superimposed codes," *IEEE Transactions on Information Theory*, vol. 10, no. 4, pp. 363–377, 1964.
- [4] M. Cheraghchi and V. Nakos, "Combinatorial group testing and sparse recovery schemes with near-optimal decoding time," *CoRR*, vol. abs/2006.08420, 2020. [Online]. Available: <https://arxiv.org/abs/2006.08420>
- [5] H. A. Inan, P. Kairouz, M. Wootters, and A. Özgür, "On the optimality of the kautz-singleton construction in probabilistic group testing," *IEEE Trans. Inf. Theor.*, vol. 65, no. 9, p. 5592–5603, sep 2019. [Online]. Available: <https://doi.org/10.1109/TIT.2019.2902397>
- [6] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.
- [7] M. Mitzenmacher and E. Upfal, *Probability and computing: Randomization and probabilistic techniques in algorithms and data analysis*. Cambridge university press, 2017.
- [8] E. Porat and A. Rothschild, "Explicit non-adaptive combinatorial group testing schemes," *CoRR*, vol. abs/0712.3876, 2007. [Online]. Available: <http://arxiv.org/abs/0712.3876>