PCR, Tropical Arithmetic, and Group Testing

Hsin-Po Wang
University of California San Diego
La Jolla, California, United States
Email: hsw001@eng.ucsd.edu

Ryan Gabrys
University of California San Diego
La Jolla, California, United States
Email: rgabrys@eng.ucsd.edu

Alexander Vardy
University of California San Diego
La Jolla, California, United States
Email: avardy@eng.ucsd.edu

***Abstract*—Polymerase chain reaction (PCR) testing is the gold standard for diagnosing COVID-19. Unfortunately, the outputs of these tests are imprecise and therefore quantitative group testing methods, which rely on precise measurements, are not applicable. Motivated by the ever-increasing demand to identify individuals infected with SARS-CoV-19, we propose a new model that leverages tropical arithmetic to characterize the PCR testing process. In many cases, some of which are highlighted in this work, tropical group testing is provably more powerful than traditional binary group testing in that it requires fewer tests than classical approaches, while additionally providing a mechanism to identify the viral load of each infected individual.**

A full version of this paper is accessible at: https://arxiv.org/abs/2201.05440

TABLE I
FOUR WAYS TO QUANTIFY AND COMBINE TEST OUTPUTS. BINARY TESTS OUTPUT "NEGATIVE" OR "POSITIVE"; COMBINING SAMPLES MEANS LOGICAL OR. QUANTITATIVE TESTS OUTPUT NUMBERS; COMBINING SAMPLES MEANS ADDITION. THE OTHER TWO REGIMES LIE IN BETWEEN.

Regime	Reading	Remixing
Binary	Negative, Positive	Neg $\vee$ Pos = Pos
Tropical	$2^{-\infty}, 2^{-40}, \dots, 2^{-12}$	$\min(30, 15) = 15$
Semiquantitative	$[0,3), [3,6), [6,9), \dots$	$[0,3) + [3,6) = [3,9)$
Quantitative	$0, 1, 2, 3, 4, 5, \dots$	$8 + 9 = 17$

I. INTRODUCTION

The COVID-19 pandemic has highlighted the critical role that widely-accessible testing can play in controlling the spread of infectious diseases. Efficient testing schemes have the potential to simultaneously reduce the time to diagnosis while improving both the reliability and accuracy of the testing procedure. This subject has attracted significant attention in the open literature [2]; however, existing works do not accurately model the semiquantitative information available at the output of the *polymerase chain reaction* (PCR) testing methods used to detect the presence of SARS-CoV-19.

PCR tests output *cycle threshold* (Ct) values, which, as a result of the testing mechanism itself, are typically represented as semiquantitative measurements in the $\log$ domain. In this instance, the term *quantitative* refers to the fact that the tests' readings are non-binary and *semi* means that the readings are noisy or inaccurate. Previous semiquantitative approaches are ill-suited for modeling the output of a PCR test as previous works mostly rely on the assumption that test measurements are reported on a linear (rather than a $\log$) scale [5]–[10].

As an illustration of the potential problem of modeling the PCR test outputs on a linear scale, consider the Ct value of a test as the dB value of a sound wave or the pH value of a liquid. When adding a 50 dB white noise with a 30 dB one, we get a 50.04 dB white noise that is indistinguishable from 50 dB. One might as well pretend that 50 dB plus 30 dB equals 50 dB. Due to the wide range in viral load between infected individuals, the same phenomenon for Ct values has been observed and is often referred to as *masking* [4].

In order to address the masking issue and also to take advantage of the semiquantitative outputs available from PCR, we propose introducing *delays* during the DNA amplification process. The basic idea will be to generate tests where each of the samples within a test can be inserted at different times. As a simple example of how this would work, suppose we design a test that consists of a single sample from an infected individual that has a Ct value of X. Then, if we delay inserting the sample by Δ cycles, the output of the resulting test would be $\Delta + X$.

We use tropical multiplication $x \odot y = x + y$ and tropical addition $x \oplus y = \min(x, y)$ to model the behavior of the Ct values that are provided as output of each of the PCR tests. See Table I for a comparison of this versus other models. According to our model, one can *match* the pattern of the Ct values against the pattern of the delays. See Figure 1 for an illustration of this process.

This paper is organized as follows: Section II reviews PCR, tropical arithmetic, and group testing; and it states this paper's goal. Section III and IV discuss the case of one and two infected individuals, respectively. Section V discusses adaptive strategies. Section VI displays simulation results.

II. BACKGROUND

A. PCR testing

A PCR test, abstractly speaking, consists of two components: a DNA amplifier that duplicate the virus DNA every minute and a sensor that lights up whenever the concentration of DNA reaches 1 arbitrary unit. Suppose a specimen contains $1 \cdot 10^{-6}$ unit of DNA and is undergoing a PCR test. After one minute, it contains $2 \cdot 10^{-6}$ unit of DNA. After ten minutes, it contains $1.024 \cdot 10^{-3}$ unit of DNA. After twenty minutes, finally, the amount of DNA reaches $1.048 \cdot 10^{0} > 1$ units and the sensor lights up. We report that this specimen required 20 minutes to trigger the sensor; "20" is its Ct value.

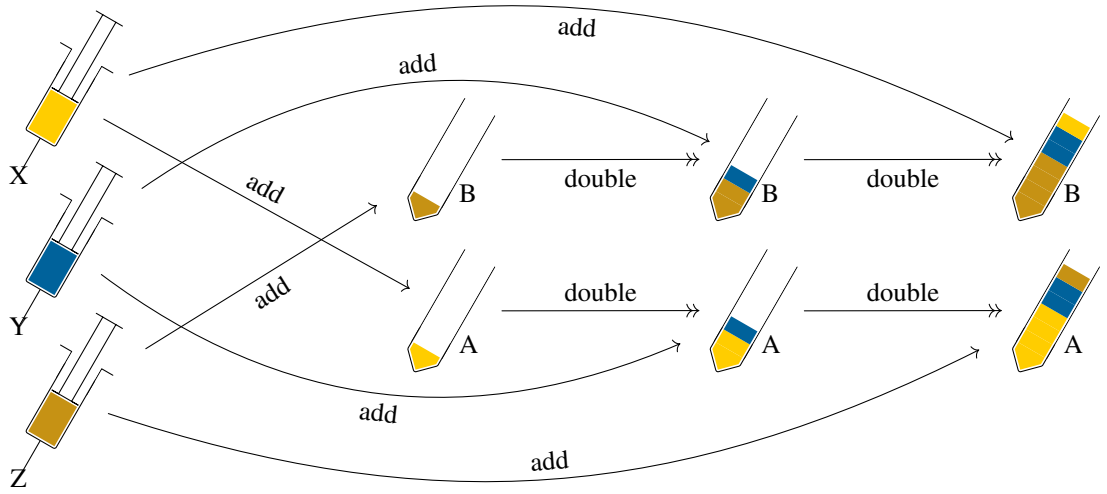

Fig. 1. Syringes: Specimens are extracted from three people. Tubes: Specimens will be remixed in two tubes at the same time the PCR machine is amplifying DNA. We double the colored area to represent the fact that the virus DNA, if any, is replicated. If X is the only patient, A will have four times as much virus as B. If Y is the only patient, A and B will have equal amount of virus. If Z is the only patient, A will have a quarter as much virus as B.

Owing to some technicality [1, Figure 2], the sensor always lights up after some whole minutes. A Ct value of 20 implies that the tested sample has a DNA concentration anywhere from $2^{-20.5}$ unit to $2^{-19.5}$ unit. This is the source of the multiplicative fuzziness, which we will model with the tropical semiring.

B. Tropical semiring

The *tropical semiring* [11] is a tuple $(\mathbb{R}\cup\{\infty\}, \oplus, \odot)$ where

$$\begin{aligned} x \oplus y &\coloneqq \min(x, y) && \text{(in particular } x \oplus \infty = x\text{)}, \\ x \odot y &\coloneqq x + y && \text{(in particular } x \odot \infty = \infty\text{)}. \end{aligned}$$

It captures the behavior of ordinary addition and ordinary multiplication in the log domain. The tropical arithmetic generalizes to matrix multiplication. For matrices A and B, define $A \odot B$ to be the matrix whose (i,k)th element is

$$\bigoplus_j (A_{ij} \odot B_{jk}) = \min_j (A_{ij} + B_{jk}).$$

Tropical matrix multiplication is a convenient language for describing the Ct value of a pool of specimens. Suppose we have N specimens with Ct values $x_1, x_2, \ldots, x_N$ and that we place each of them into the same PCR test, each delayed by $\delta_1, \delta_2, \ldots, \delta_N$ minutes, respectively. Then the final Ct value will be roughly

$$-\log_2\Big(\sum_j 2^{-\delta_j - x_j}\Big) \approx \min_j (\delta_j + x_j).$$

We formalize this as an assumption in the following.

Assumption 1 (Tropical model): Suppose we design a PCR test which contains N specimens with Ct values $x_1, x_2, \ldots, x_N$, each delayed by $\delta_1, \delta_2, \ldots, \delta_N$ minutes, respectively. The final Ct value is assumed to be exactly

$$\boldsymbol{\delta} \odot \boldsymbol{x} = \min_j (\delta_j + x_j).$$

Here, $\boldsymbol{\delta}$ is the row vector of the delays and $\boldsymbol{x}$ is the column vector of the Ct values. When there are T tests in parallel, we denote the test results by

$$S \odot \boldsymbol{x} = \begin{bmatrix} \min_j (S_{1j} + x_j) \\ \vdots \\ \min_j (S_{Tj} + x_j) \end{bmatrix}$$

where S is a $T \times N$ matrix and each row of S corresponds to a test. We refer to the matrix S as a *schedule*.

C. Group testing (GT)

Invented by Dorfman in 1943, GT has led to a handful of generalizations and applications [3]. Let $\boldsymbol{x}$ and A denote a boolean vector and a boolean matrix, respectively. Under the classical GT setup, the goal is to solve for $\boldsymbol{x}$ given the logical matrix–vector product $A \wedge \boldsymbol{x}$ whose ith row is $\bigvee_j (A_{ij} \wedge x_j)$.

Quantitative GT is a generalization of the classical (binary) GT setup. Under the quantitative GT setup, $\boldsymbol{x}$ is a real vector and $A\boldsymbol{x}$ is the ordinary matrix multiplication. A further generalization, known as *semiquantitative* GT [12], considers the setup where each entry of $\boldsymbol{x}$ is one of the intervals $[0, \theta_1)$, $[\theta_1, \theta_2)$, $[\theta_2, \theta_3)$, $\ldots$ for some predefined set of thresholds $\{\theta_j\}_j$, and the arithmetic in $A\boldsymbol{x}$ is interpreted as interval arithmetic.

In our tropical setting, we choose the powers of 2 to be $\{\theta_j\}_j$. Tropical GT sits between binary GT and semiquantitative GT, as shown in Table I, in that it has more than two outcomes and its addition is idempotent.

D. Problem statement

We are interested in both non-adaptive and adaptive testing schemes. For shorthand, the *support size* of a vector $\boldsymbol{x} \in \{0, 1, 2, \ldots, \infty\}^{N\times 1}$ is the number of entries of $\boldsymbol{x}$ that are finite.

Definition 2 (Nonadaptive testing): A (T, N, D)-*tropical code* is a matrix $S \in \{0, 1, 2, \ldots, \infty\}^{T\times N}$ such that $S \odot \boldsymbol{x} \neq S \odot \boldsymbol{y}$ for every distinct $\boldsymbol{x}, \boldsymbol{y} \in \{0, 1, 2, \ldots, \infty\}^{N\times 1}$ with

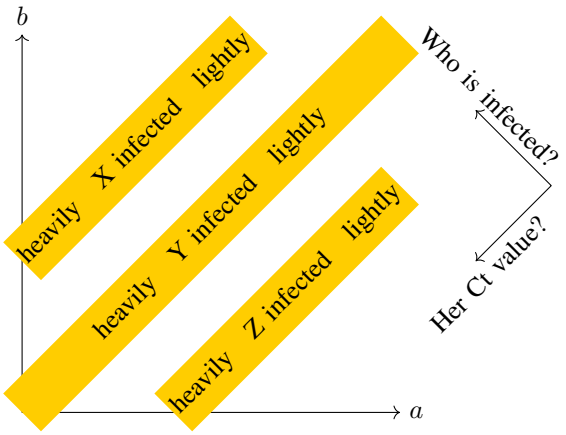

Fig. 2. The configuration space of test results of schedule (1).

support size at most D. A tropical code is said to be *within maximum delay*[1] ℓ if $S \in \{0, 1, \dots, \ell, \infty\}^{T\times N}$.

Definition 3 (Adaptive testing): An R-(T, N, D)-*tropical protocol* is a series of R functions, $\mathcal{S}^{(1)}, \mathcal{S}^{(2)}, \dots, \mathcal{S}^{(R)}$, that take past results as inputs and output schedules[2]

$$\begin{aligned} S^{(1)} &= \mathcal{S}^{(1)}(), \\ S^{(2)} &= \mathcal{S}^{(2)}(S^{(1)} \odot \boldsymbol{x}), \\ &\vdots \\ S^{(R)} &= \mathcal{S}^{(R)}\left(\begin{bmatrix} S^{(1)} \\ \vdots \\ S^{(R-1)} \end{bmatrix} \odot \boldsymbol{x}\right) \end{aligned}$$

such that (i) the numbers of rows may depend on past results but the total is at most T, (ii) the numbers of columns are N, and (iii) the final result

$$\begin{bmatrix} S^{(1)} \\ \vdots \\ S^{(R)} \end{bmatrix} \odot \boldsymbol{x}$$

is unique among all $\boldsymbol{x} \in \{0, 1, 2, \dots, \infty\}^N$ with support size at most D. A tropical protocol is said to be *within maximum delay* ℓ if all schedules assume alphabet $\{0, 1, \dots, \ell, \infty\}$.

III. One Infected Individual

Let there be three individuals with Ct values x, y and z. Suppose at most one of them is infected. Figure 1 applies the schedule:

$$\begin{bmatrix} a \\ b \end{bmatrix} := \begin{bmatrix} 0 & 1 & 2 \\ 2 & 1 & 0 \end{bmatrix} \odot \begin{bmatrix} x \\ y \\ z \end{bmatrix} \tag{1}$$

If all three individuals are healthy, the test results a and b will be infinity. Otherwise, a and b will be finite. If x is infected (finite), $a - b = -2$. Similarly, if y is infected, $a - b = 0$. Finally, if z is infected, $a-b = 2$. To summarize, the difference in the Ct values encodes the identity of the infected person, as also shown in Figure 2.

Using short and long delays we can assemble a schedule that can screen more people. Suppose ℓ is the maximum possible delay, then this is a $(2, 2\ell+3, 1)$-tropical code within maximum delay ℓ:

$$\begin{bmatrix} 0 & 0 & \cdots & 0 & 0 & 1 & \cdots & \ell & \infty \\ \infty & \ell & \cdots & 1 & 0 & 0 & \cdots & 0 & 0 \end{bmatrix} \tag{2}$$

More generally, we have the following theorem.

Theorem 4 ([1, Section III]): If $S \in \{0, 1, \dots, \ell, \infty\}^{2\times N}$ does not contain $\left[\begin{smallmatrix}\infty \\ \infty\end{smallmatrix}\right]$ and satisfies

$$\left|\{S_{1j} - S_{2j} \mid j = 1, \dots, N\}\right| = N,$$

then S is a $(2, N, 1)$-tropical code within maximum delay ℓ. Such code exists iff $N \le 2\ell + 3$.

If more tests can be afforded, we have the following.

Theorem 5 ([1, Appendix A]): If $S \in \{0, 1, \dots, \ell, \infty\}^{T\times N}$ does not contain infinite column $\begin{bmatrix}\infty & \cdots & \infty\end{bmatrix}^\top$ and satisfies

$$\left|\left\{\begin{bmatrix} S_{1j} - \min_t S_{tj} \\ \vdots \\ S_{Tj} - \min_t S_{tj} \end{bmatrix} \middle| \; j = 1, \dots, N\right\}\right| = N,$$

then S is a $(T, N, 1)$-tropical code within maximum delay ℓ. Such a code exists iff $N \le (\ell + 2)^T - (\ell + 1)^T$.

IV. Two Infected Individuals

The case of two infected individuals is considerably more interesting. We will consider nonadaptive schemes for the following scenarios: (i) the number of tubes each person participates in is minimized, (ii) the total number of tests is minimized, and (iii) the maximum delay is minimized.

A. Every person participates in (only) two tubes

Since pipetting is laborious and prone to mistakes, one might prefer GT schemes wherein every person appear in two tubes, the minimal possible number of tubes before GT degenerates to individual testing[3]. Under this constraint, a tropical code can be expressed by a weighted digraph in the manner specified below.

Let G be a weighted digraph. Let G have T vertices, each corresponding to a test, and N weighted arrows, each corresponding to a person. Suppose that, for each $j = 1, \dots, N$, the jth person corresponds to an arrow $u_j \to v_j$ with weight $\delta_j \in \mathbb{Z}$. Assign the schedule

$$S_{tj} := \begin{cases} \max(0, -\delta_j) & \text{if } t = u_j, \\ \max(0, \delta_j) & \text{if } t = v_j, \\ \infty & \text{otherwise.} \end{cases}$$

We have a necessary condition.[4]

[1]Note that allowing negative delay adds nothing new to tropical GT. This is because S and $S + \mathbf{1}_{T\times N}$ have the same functionality.

[2]The first function, $\mathcal{S}^{(1)}$, takes empty input as there is no "past result".

[3]Note that if a sick person attends only one tube while a sicker person also attends that tube, the former will be masked. So every person that attends only one tube cannot share that tube with others.

[4]For every column of a schedule S, only the difference of the non-zero entries matters. It suffices to consider schedules induced by weighted digraphs.

Lemma 6 ([1, Section IV]): In order for S to be a $(T, N, 2)$-tropical code, the undirected version of G cannot have repeated edges or 3-cycles.

We also have a sufficient condition.

Lemma 7 ([1, Section IV]): Regard an arrow $u \to v$ with weight δ the same as the opposite arrow $v \to u$ with the opposite weight $-\delta$. Then G gives rise to a $(T, N, 2)$-tropical code if its girth is at least 4 and for all directed 4-cycles with weights δ_w, δ_x, δ_y, and δ_z, the sum $\delta_w + \delta_x + \delta_y + \delta_z$ is nonzero.

Knowing necessary and sufficient conditions, we can design a tropical code in two steps. Step one, to maximize the number of individuals being tested, we need a graph with as many edges as possible but contains no 3-cycles. A complete bipartite graph satisfies this condition (Turán's theorem). Step two: we need to ensure that $\delta_w + \delta_x + \delta_y + \delta_z \neq 0$ along any 4-cycle. The multiplication table

$$S_{tj} \coloneqq \begin{cases} \max(0, -u_j v_j \bmod p) & \text{if } t = u_j, \\ \max(0, u_j v_j \bmod p) & \text{if } t = v_j, \\ \infty & \text{otherwise.} \end{cases}$$

satisfies this condition, Here, all arrows goes from the left part to the right part of the graph; the jth person corresponds to the arrow from u_j to v_j; the number $p > T/2$ is a prime; and $\bmod p$ is the modulo operator that returns integers between $\pm(p-1)/2$. We conclude the theorem.

Theorem 8 ([1, Section IV]): Let $T \geq 2$. Let $p \geq T/2$ be an odd prime. There exists a $(T, \lfloor T/2 \rfloor \lceil T/2 \rceil, 2)$-tropical code within maximum delay $(p-1)/2$, wherein everyone participates in two tubes.

B. Minimizing the number of tests

When people are to appear in more than two tests, each person is not analogous to an edge but to a hyperedge (in the hypergraph terminology) or to a block (in the block design terminology). We stick to the latter.

Let $[T]$ be $\{1, 2, \ldots, T\}$. A *block design* $\mathcal{F} \subseteq 2^{[T]}$ is a family of subsets of $[T]$. A *block* $B \in \mathcal{F}$ is regarded as a person as well as the subset of tests she is in. There are certain configurations of blocks that make decoding pathological or even impossible. For instance, if two blocks $B, Z \in \mathcal{F}$ are such that $Z \subseteq B$, then B being heavily infected will mask Z. That is, we will not be able to tell if Z is healthy or slightly infected because B "pollutes" all tubes Z is in. Similarly, if three blocks $B, Y, Z \in \mathcal{F}$ are such that $Z \setminus B = Y \setminus B$. Then the only chance we can tell Y and Z apart is when $|Z \setminus B| \geq 2$ because learning who is infected and how infected she is costs two degrees of freedom.

The preceding discussion leads to this sufficient condition.

Theorem 9 ([1, Section V]): Suppose a block design $\mathcal{F} \subseteq 2^{[T]}$ is such that $|Z \setminus B| \geq 2$ for any distinct blocks $B, Z \in \mathcal{F}$. Then there exists a $(T, |\mathcal{F}|, 2)$-tropical code.

A nearly-identical statement holds for more infected people.

Theorem 10 ([1, Section VI]): Suppose a block design $\mathcal{F} \subseteq 2^{[T]}$ is such that $|Z \setminus (B_1 \cup \cdots \cup B_{D-1})| \geq 2$ for any distinct blocks $Z, B_1, \ldots, B_{D-1} \in \mathcal{F}$. Then there exists a $(T, |\mathcal{F}|, D)$-tropical code.

How to design a $\mathcal{F}$ with optimal number of blocks, unlike complete bipartite graphs having optimal number of edges in the previous subsection, remains unclear and is an active subject to date.

C. Minimizing, but not completely forbidding, the delays

In the previous subsection, the condition $|Z \setminus B| = 2$ was sufficient provided we could make the difference of delays $S_{uj} - S_{vj}$ distinct for $\{u, v\} = Z \setminus B$. Unlike in the first subsection wherein the multiplication table (and the associated bipartite graph) allowed us to design nearly optimal schedule matrices, in this case neither the structure of $\mathcal{F}$ nor the assignment of delays is straightforward. To begin, we consider a block design $\mathcal{F}$ that has less-than-optimal number of blocks. We will subsequently improve this initial attempt before presenting our main result.

In the following discussion, we assume that the two infected individuals have different Ct values. See the full version [1] for how to deal with the equal case.

Define the *Li index*[5] of a (T, N, D)-tropical code to be $D \log_2(N)/T$. Suppose S is a $(t, n, 2)$-tropical code whose Li index $2\log_2(n)/t$ is fairly large (e.g., ≥ 1). Consider the Kronecker products

$$\begin{bmatrix} S \otimes \mathbf{1}_{1\times n} \\ \mathbf{1}_{1\times n} \otimes S \end{bmatrix}$$

where $\mathbf{1}_{1\times n}$ is the all-one row vector of length n. The first t tests group collections of n consecutive individuals into a single pool and apply S to each of the n pools. The last t tests pool together 1 out of every n individuals so that last set of t tests are "orthogonal" to the first set of t tests.

Suppose $0 \leq Y, Z \leq n-1$ are the indices of the infected individuals where $Y = y_{10}n + y_1$ and $Z = z_{10}n + z_1$. As a result of the test design, the first t tests can tell us the tens digits y_{10} and z_{10}. The last t tests can tell us the ones digits y_1 and z_1. At this point, we know that one of the following holds:

$$\{Y, Z\} = \{y_{10}n + y_1, z_{10}n + z_1\}$$

or

$$\{Y, Z\} = \{y_{10}n + z_1, z_{10}n + y_1\}.$$

We now invoke the assumption that Y and Z have different Ct values. Under this assumption, y_1 and y_{10} will be associated with one Ct value while z_1 and z_{10} will be associated with another Ct value. Hence the decoder can simply match the Ct values to recover Y and Z.

It is not hard to observe that

$$\begin{bmatrix} S \otimes \mathbf{1}_{1\times n} \otimes \mathbf{1}_{1\times n} \\ \mathbf{1}_{1\times n} \otimes S \otimes \mathbf{1}_{1\times n} \\ \mathbf{1}_{1\times n} \otimes \mathbf{1}_{1\times n} \otimes S \end{bmatrix}$$

and its generalizations work the same way. The final theorem along this line is given below. We need $2k-3$ extra tests that function like the checksums of a Reed–Solomon code to deal with the case where Y and Z have the same Ct value.

[5]Note that this definition aligns with the classical result that $T = \Theta(D \log_2 N)$ when binary search is employed.

Theorem 11 ([1, Appendix B]): Let S be a $(t, n, 2)$-tropical code within maximum delay n, then there is an $(n^k, kt+2k-3, 2)$-tropical code within $q-1$ cycles, where $q \geq \max(3k-3, n)$ is a prime power.

Asymptotically, the Li index $2\log_2(n^k)/(kt+2k-3)$ converges to $2\log_2(n)/(t+2)$.

V. Adaptive Strategies

Adaptive testing significantly reduces the number of tests necessary to identify infected individuals. The intuition behind this is that if some individual has a relatively small Ct value, then it must be rather easy to locate her and isolate her specimen from future tests. For finding two infected individuals, in particular, we can reduce the number of tests to only four.

Theorem 12: For any N. There is a 2-$(4, N, 2)$-tropical protocol. (See [1, Section VII] for a proof.)

For general D, each infected individual requires roughly only three tests.

Theorem 13 ([1, Section VIII]): For any N, there exists a $(3N+1)$-$(3N+1, N, N)$-tropical protocol that is, simultaneously, a $(3N+1)$-$(3D+1, N, D)$-tropical protocol for all $D \leq N$.

Next, we present an overview of the adaptive strategy. Given any population, apply the schedule (2) to identify a potentially infected person. Then apply a confirmation test

$$\begin{bmatrix}\infty & \cdots & \infty & 0 & \infty & \cdots & \infty\end{bmatrix}$$

to verify if the person is actually infected. If the person is infected, remove her from the population and start over. If not, there are other techniques that can reuse the information provided by (2).

For the 2-$(4, N, 2)$-tropical protocol, there is another technique whereby we can superimpose two schedules so that one test can replace two of the existing tests. This reduces the number of tests required from $3 \cdot 2 + 1$ to 4. An open problem is whether it is possible for general $D > 1$ to design schemes that require two tests per individual, provided that our work confirms that three tests per individual is sufficient.

As for sub-linear delay, we have the following. The main idea is to apply the previous theorem with ℓ-ary search.

Theorem 14 ([1, Appendix C]): Fix an N and an ℓ. Let L be $\lceil \log_\ell N \rceil$. There is a $(4NL+1)$-$(4NL+1, N, N)$-tropical protocol within maximum delay ℓ that is, simultaneously, a $(4NL+1)$-$(4DL+1, N, D)$-tropical protocol for all $D \leq N$.

VI. Simulations

To generate pseudo random data to benchmark our decoder, we no longer impose Assumption 1. For this setup, we assume the number of virus particles is additive when two specimens are combined; however, our decoder still operates under Assumption 1. Our goal is to illustrate that even under realistic testing conditions, the tropical model is still functional.

Let x_j, for $j \in [N]$, be the Ct value of the jth person. With 90% probability she is healthy and $x_j := 99$. With 10% probability she is infected and x_j is drawn uniformly from the interval $[16, 32]$. Let $\boldsymbol{x}$ be the $N \times 1$ column vector of the x's, and let $2^{-\boldsymbol{x}}$ be the entry-wise exponentiation.

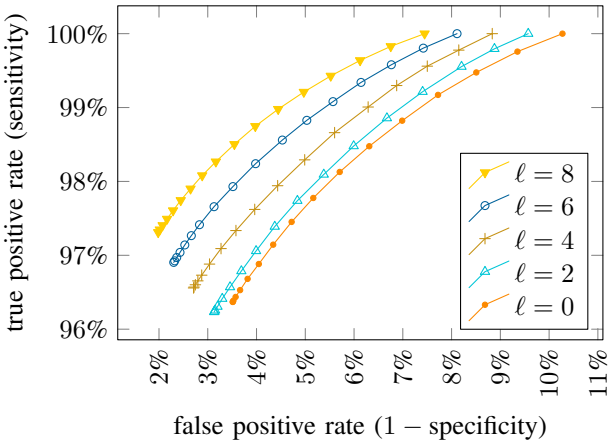

Fig. 3. The ROC tradeoff for different delay limit ℓ.

Let S be a schedule derived in the following way. Suppose $\mathcal{F}$ is a block design. Let S_{tj} be 99 if t is not in the jth block of $\mathcal{F}$. Otherwise, let S_{tj} be $\ell \cdot \text{Bernoulli}(1/2) \in \{0, \ell\}$, a random delay drawn unbiasedly and independently. Let 2^{-S} be the entry-wise exponentiation. Suppose $\boldsymbol{v}$ is the result of the matrix multiplication $\boldsymbol{v} = 2^{-S}2^{-\boldsymbol{x}}$, representing the numbers of virus particles in the tubes. Let $\boldsymbol{c}$ be $\min(40, \lfloor -\log_2(\boldsymbol{v}) \rfloor)$, where all operations are applied entry-wisely.

The decoder is given S and $\boldsymbol{c}$. It basically processes the test results in two phases. First, it looks for matches of differences $c_u - c_v = S_{uj} - S_{vj}$. If it spots such a j, the jth person is declared infected. Next, it goes over unexplained tubes and find the most probable "cause" using a variant of the classical SCOMP algorithm [13].

This decoder is highly specific for that the "certificate" $c_u - c_v = S_{uj} - S_{vj}$ is difficult to fabricate. With the second phase that aims to explain unexplained tubes, the specificity drops a little but the sensitivity is boosted. Together, $\mathcal{F}$ being a 15-by-35 Kirkman triple system is capable of achieving 97.5% sensitivity and 95% specificity when the prevalence rate is 10% and no delay is applied ($\ell = 0$). With delay ($\ell = 3$), the sensitivity reaches 98%. With greater delay ($\ell = 7$), the sensitivity reaches 99%. See Figure 3 for an ROC tradeoff. See the full version [1, Appendix D] for more plots.

VII. Conclusions

When one applies quantitative GT to data that is fuzzy and spanning a large range (including but not limited to Ct values), the fuzziness of a relatively large number erodes relatively small numbers. We propose that one might as well redefine the addition so that it is forgetful to begin with. That way, one is forced to seek for other methods to help decode, e.g., introducing delays and matching the difference thereof. The ideas we develop along this path are instances of tropical group testing.

VIII. Acknowledgment

This work was supported by NSF grants CCF-2107346 and CCF-1764104.

REFERENCES

[1] H.-P. Wang and R. Gabrys and A. Vardy. *Tropical Group Testing*. 2022. https://arxiv.org/abs/2201.05440

[2] M. Aldridge and D. Ellis. *Pooled Testing and Its Applications in the COVID-19 Pandemic*. Pandemics: Insurance and Social Protection. 2022. https://doi.org/10.1007/978-3-030-78334-1_11

[3] D.-Z. Du and F. K. Hwang. Combinatorial Group Testing and Its Applications. 1999. https://doi.org/10.1142/4252

[4] R. Gabrys et al. *AC-DC: Amplification Curve Diagnostics for Covid-19 Group Testing*. 2021. https://arxiv.org/abs/2011.05223

[5] S. Ghosh et al. *A Compressed Sensing Approach to Pooled RT-PCR Testing for COVID-19 Detection*. IEEE Open Journal of Signal Processing. 2021. https://doi.org/10.1109/OJSP.2021.3075913

[6] N. Shental et al. *Efficient high-throughput SARS-CoV-2 testing to detect asymptomatic carriers*. Science Advances. 2020. https://doi.org/10.1126/sciadv.abc5961

[7] J. Yi et al. *Error Correction Codes for COVID-19 Virus and Antibody Testing: Using Pooled Testing to Increase Test Reliability*. 2020. https://arxiv.org/abs/2007.14919

[8] A. Heidarzadeh and K. Narayanan. *Two-Stage Adaptive Pooling with RT-qPCR for Covid-19 Screening*. 2021 IEEE International Conference on Acoustics, Speech and Signal Processing. 2021. https://doi.org/10.1109/ICASSP39728.2021.9413685

[9] A. Cohen et al. *Multi-Level Group Testing with Application to One-Shot Pooled COVID-19 Tests*. 2021 IEEE International Conference on Acoustics, Speech and Signal Processing. 2021. https://doi.org/10.1109/ICASSP39728.2021.9414574

[10] H. B. Petersen et al. *Improving the Reliability of Pooled Testing with Combinatorial Decoding and Compressed Sensing*. 2021 55th Annual Conference on Information Sciences and Systems. 2021. https://doi.org/10.1109/CISS50987.2021.9400261

[11] M. Joswig. Essentials of tropical combinatorics. 2022. ISBN: 978-1-4704-6741-8

[12] A. Emad and O. Milenkovic. *Semiquantitative Group Testing*. IEEE Transactions on Information Theory. 2014. https://doi.org/10.1109/TIT.2014.2327630

[13] M. Aldridge and L. Baldassini and O. Johnson. *Group Testing Algorithms: Bounds and Simulations*. IEEE Transactions on Information Theory. 2014. https://doi.org/10.1109/TIT.2014.2314472