PAPER • **OPEN ACCESS**

Viral load inference of high dynamic range signals using non-adaptive pooled testing

To cite this article: Mansoor Sheikh and David Saad 2025 *Phys. Scr.* **100** 035208

Physica Scripta

PAPER

Viral load inference of high dynamic range signals using non-adaptive pooled testing

Mansoor Sheikh and David Saad

Non-Linearity and Complexity Research Group, Aston University, Birmingham B4 7ET, United Kingdom

E-mail: d.saad@aston.ac.uk

Keywords: pooled testing, high dynamic range, message passing, noise models

OPEN ACCESS

RECEIVED
8 July 2024

REVISED
26 December 2024

ACCEPTED FOR PUBLICATION
16 January 2025

PUBLISHED
6 February 2025

Abstract

Medical diagnostic testing can be made significantly more efficient using pooled testing protocols. These typically require a sparse infection signal and use either binary or real-valued entries of $\mathcal{O}(1)$. However, existing methods do not allow for inferring viral loads which span many orders of magnitude. We develop a message passing algorithm coupled with a PCR (Polymerase Chain Reaction) specific noise function to allow accurate inference of realistic viral load signals. A version of the Combinatorial Orthogonal Matching Pursuit algorithm is employed to compare the obtained inference results with those of our probabilistic message passing. This work is in the non-adaptive setting and could open the possibility of efficient screening where viral load determination is clinically important.

1. Introduction

Typically the infection status of a patient is determined by carrying out a single diagnostic test, which represents a poor use of resource in the low-prevalence case where most tests return negative. A well-studied improvement is the pooled testing concept [1] which allows for the structured mixing of patient samples into groups or pools. By testing these mixtures (rather than the individual samples), the number of diagnostic tests required to determine each patient's infection status can be dramatically reduced. This can be beneficial where shortages of laboratory diagnostic equipment, raw testing materials and qualified staff occur. The two main approaches to pooled testing are the adaptive and non-adaptive protocols. In the former, tests are run sequentially, with information from the previous testing steps informing the next one [1, 2]. However during the early spread of a virulent pathogen, laboratories are typically running at capacity and the logistics of adaptive pooled testing are not always feasible. In this paper, we focus on non-adaptive pooled testing which requires only one testing procedure to infer infection status.

We consider the problem of recovering an unknown N-dimensional vector $\mathbf{x}$ representing the diagnostic status of N patients where component values can be either binary or real. We will introduce the concept using binary values. Combinations of samples are pooled according to the $M \times N$ pooling/measurement matrix, $\boldsymbol{F}$ (see figure 1) where each row specifies which patient samples are included in each test. Matrix $\boldsymbol{F}$ can be constructed as either random or structured and is a known quantity in the inference problem. Each row of $\boldsymbol{F}$ can be thought of as probing/examining the unknown signal vector by taking a linear projection via a physical pooling of patient samples. The M results are output as an M-dimensional vector, $\mathbf{y}$ and the aim is to solve the inverse problem of inferring $\mathbf{x}$ from known $\boldsymbol{F}$ and $\mathbf{y}$. A smaller measurement rate $\alpha = M/N$ corresponds to a more efficient pooled testing setup.

The judicious design of an efficient pooling regime helps minimize α. Typical approaches for real-valued variables $\mathbf{x}$ use pooling matrices with Gaussian random variable entries. Subsequent improvements include matrix designs from error-correcting codes [3] and from physics-inspired methods of crystal nucleation [4].

Individual infection status is determined from the pooled test measurements via a statistical or combinatorial inference procedure with the method used depending on the infection status representation e.g. binary status (typical in group testing) or real-valued viral loads (compressed sensing). These approaches to

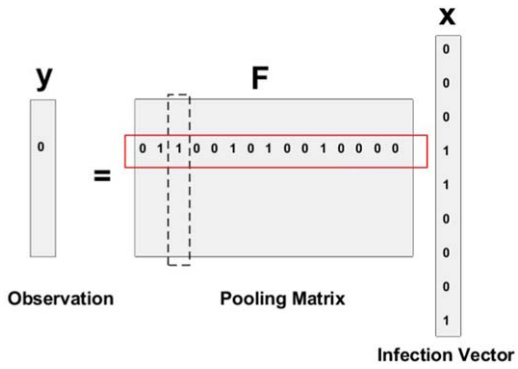

Figure 1. The inverse problem is to infer the $N \times 1$ infection vector, $\mathbf{x}$, given the $M \times 1$ observation vector, $\mathbf{y}$, and the $M \times N$ pooling matrix, $\boldsymbol{F}$.

group testing have been investigated in the fields of computer science [5], statistics [6, 7], error-correcting codes [3] and statistical physics [4, 8–10]. Group testing typically assumes a sparse infection signal i.e. low disease prevalence.

In some applications, the viral load can range from approximately $10^2 - 10^9$ copies per *mL* e.g. in the case of SARS-CoV-2 [11]. This makes inference problematic, especially for inferring real-valued variables, and, to our knowledge, there is a lack of work dealing with this case. The problem stems from each pooled measurement containing a linear combination of zero, low, medium and high real values. Informally, the lower values will necessarily be 'drowned out' by the high values in the pooling process. This *masking* phenomenon is nicely articulated in [12], which addresses the problem by concatenated experimentation. In this work, we do not assume that the signal entries are $\sim\mathcal{O}(1)$. First we ask the natural question of whether it is clinically relevant to ascertain the viral load of a sample rather than a simple absence or presence of the virus; the answer depends on the application:

- *Viral load is clinically relevant but the efficiencies of pooled testing are not required.* For chronic diseases such as HIV, a lower viral load means higher life expectancy [13]. Hence the success of treatment regimes is measured by viral load. For a previously infected patient, knowing the binary presence or absence does not matter because the virus will be there in some amount. In the case of HIV, the efficiencies of group testing are typically not required (unless it is done as community screening e.g. [14]).
- *Widespread screening required but mostly the presence or absence of the disease is relevant*: In some cases knowing the viral load is less likely to affect the treatment or isolation regime (such as COVID or Hepatitis C). Here low viral loads mean early infection stages but are less likely to affect the clinical decision. A possible *caveat* to this is evidence that patients with high viral load are more efficient at spreading infection [15, 16]. Identifying and quarantining these individuals will have an outsized effect on viral spread.
- *Both viral load and efficient screening are relevant*: Pooled testing can be used to estimate the amount of contaminants in food. Here it does matter how much salmonella there is in your chicken. There is also a clear commercial benefit for food companies to gain efficiencies from pooled testing.
- *Inferring the stage of outbreak*: The distribution of viral load estimates in a population is related to the stage of the viral outbreak [17, 18]. By inferring their values, the method described in the current paper provides a resource-efficient method of estimating the stage of infection and potentially the viral reproduction number, R_0.

A naive approach would be to run a compressed sensing algorithm to recover the viral load signal which has a high dynamic range. Compressed sensing is a theoretically and empirically accurate signal recovery scheme [6, 10, 19] where the typical error is orders of magnitude smaller than the signal magnitude. However, this results in small signal components becoming indistinguishable from the noise. Synthetic studies conventionally sample the non-zero viral loads from a uniform distribution e.g. $x \sim \mathcal{U}(2^0,\ 2^{15})$ in [7]. Here approximately 90% of the samples lie in the range $(2^{12}, 2^{15})$, which is above the typical noise level in compressing sensing algorithms, leading to flattering accuracy. The claimed high dynamic range is *probably* not that high. This argument is applicable to a lesser extent in [20] where viral loads are sampled *uniformly* from [0, 1000] but discretization results in 70% of infected samples being in the mid [300, 700) and high [700, 1000] categories. In the current paper, we sample uniformly from a logarithmic range i.e. equal number of samples from each of $2^0, 2^1, 2^2, \ldots$

which is more realistic and relevant. This ensures there is truly a high dynamic signal range but results in a materially harder problem.

The manuscript is laid out as follows: section 2 describes the model setup including combination protocols specific to PCR diagnostic testing. Section 3 outlines our specific probabilistic decoding protocol in the form of message passing equations. Section 4 develops the specific noise model and how it differs from the typical Gaussian noise approach. Section 5 reports on results obtained from numerical experiments and compares them with a modified Combinatorial Orthogonal Matching Pursuit algorithm developed specifically for inferring integer C_t values. Section 6 concludes the paper.

2. Model

2.1. Standard combination protocols

There are two typical combination protocols associated with the inverse problem of figure 1:

1. Standard matrix multiplication of $\boldsymbol{F}$ and $\mathbf{x}$ e.g. [4]

$$\tilde{y}_\mu = \sum_{i=1}^{N} F_{\mu i} x_i + \xi_\mu \tag{2.1}$$

where ξ_μ represents Gaussian noise associated with test μ, x_i the individual loads, $F_{\mu i}$ the mixing matrix and $\tilde{y}_\mu$ the noisy test result (a corrupted version of $\mathbf{y}$). This is termed the *linear estimation problem* [21]

2. The binary testing regime uses a logical sum e.g., [8]. If one patient is infected, the test output is infected.

$$\tilde{y}_\mu = C(\bigvee_{i=1}^{N} F_{\mu i} \quad x_i) \tag{2.2}$$

where x_i the individual presence/absence of the disease, $\tilde{y}_\mu$ the noisy test result and $C(\ldots)$ is a probabilistic function incorporating measurement noise such as false positive and false negative rates. Since each patient participates in multiple tests (so-called overlapping tests), the accuracy achieved can improve upon the device error settings [8].

2.2. Decoding algorithms

The objective is to solve (2.1) or (2.2) for the unknown signal vector $\mathbf{x}$ and hence individual patients' infection status. These typically rely on signal sparsity and can be categorised as greedy (matching pursuit, orthogonal matching pursuit), convex relaxation (basis pursuit, LASSO), combinatorial (Fourier sampling algorithm) or iterative pseudoposterior estimation (message passing). See [22, 23] for an overview. Some studies use additional information such as knowledge of patient interactions (*side information*), to improve recovery accuracy [24–26].

Message passing algorithms used in this paper are inherently probabilistic, have low computational complexity and the ability to run in parallel [23]. The form of the iterative equations depends on whether the signal entries are discrete or real-valued. A critical aspect that is overlooked in many of the schemes is having an appropriate noise model, which is embedded in the recovery algorithm; as the dependence of the observations (threshold PCR cycles) and the viral load is polynomial, a simple Gaussian noise assumption will not do and may lead to significant errors. There is also a link between the noise model and loss measure exercised, e.g., a quadratic loss typically implies Gaussian noise.

As discussed below, we find that recovery of a high dynamic range signal suggests using discrete valued signals. This approach coupled with a Gradient Descent algorithm was used in [27]. It relies on assumptions regarding the noise model used and its accuracy depends on the properties of the loss-landscape i.e. non-smooth or multiple local solutions. Our choice is rooted in the choice of an appropriate noise model within a principled probabilistic approximation scheme.

2.3. PCR specific notation

In this paper, we focus on amplification methods such those used in the Polymerase Chain Reaction (PCR) device. The initial patient sample is repeatedly heated and cooled to encourage a doubling of the viral RNA (see figure 2 for a schematic representation). Hence the initial viral load is amplified. A marker is added which fluoresces when attached to a specific section of the virus. Once sufficient fluorescence is detected with an optical device, the cycling is stopped and the resulting cycle number recorded. If no fluorescence is detected after an upper limit of cycles, typically 40, it is determined that no virus was originally present. The upper cycle threshold

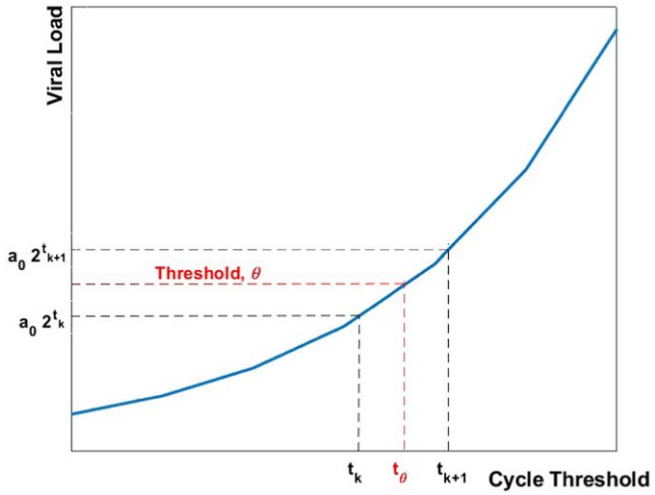

Figure 2. This schematic plot shows the doubling process central to PCR devices. Since detection is at integer values, the observed C_t reading is 'rounded-up' from the exact solution to $a_0 2^{t^\theta} = \theta$ where a_0 is a random variable which represents the initial viral load of the sample (notice that t^θ is not an integer). We do not observe a_0 directly but only via an integer valued $\lceil t^\theta \rceil$. Reading off from curve $a_0 2^{t_k} < \theta \leqslant a_0 2^{t_{k+1}}$. Re-arranging $\theta 2^{-t_{k+1}} \leqslant a_0 < \theta 2^{-t_k}$.

(C_t) limit is determined by the *limit of detection* parameter of the PCR device. A C_t value of 20 typically signifies a high viral load. We now define the notation related to the PCR protocol.

- $\theta =$ threshold for detection of fluorescence. This is typically measured in number of viral copies per ml of transport media and is assumed constant for a given PCR device.
- $a_i =$ initial viral load (number of viral copies per ml). The higher the initial viral load a_i, the lower the number of amplification cycles needed to detect the virus.
- $t_i =$ number of amplification cycles required to detect fluorescence if sample i is tested individually. Values of t_i are read off as integers rather than the 'exact' real valued solution to $\theta = a_i 2^{t_i}$.
- $t_\mu^y =$ number of amplification cycles required to detect fluorescence for the pooled test μ.

Given the nature of the PCR doubling cycle, the linear projections described in (2.1) no longer hold. This is because the viral loads are averaged in the pooling process but these viral loads are only observed via the integer cycle number. A new combination protocol is formulated in the next section.

2.4. Simple mixing example

To gain intuition for the idiosyncrasy of PCR mixing, consider two patients i and j where $\theta = a_i 2^{t_i^\theta}$ and $\theta = a_j 2^{t_j^\theta}$ represent each patient being tested individually (note that t^θ are not integers). The aim is to recover t_i and t_j from the possibly noisy measurement vector $\tilde{\mathbf{y}}$. If these samples are combined in an equal ratio in test μ, the resulting viral load, which is never directly observed, is $\frac{1}{2}(a_i + a_j) = \frac{\theta}{2}(2^{-t_i^\theta} + 2^{-t_j^\theta})$. The corresponding measurement cycle number, which we label t_μ^y is given by the relationship $2^{-t_\mu^\theta} = \frac{1}{2}(2^{-t_i^\theta} + 2^{-t_j^\theta})$ where the θ parameters cancel. To illustrate numerically, suppose the initial viral loads are $a_i = 2^5 = 32$ and $a_j = 2^6 = 64$ and the detection threshold $\theta = 2^9 = 512$. We find $t_i^\theta = t_i = 4$ and $t_j^\theta = t_j = 3$ noting the inverse relationship between t_i^θ and a_i. For samples pooled in an equal ratio, the viral load per unit volume is $\frac{1}{2}(32 + 64) = 48$ leading to $t_\mu^\theta \approx 3.42$, a non-integer value between t_i^θ and t_j^θ. The PCR device will return a measurement C_t value of $\lceil t_\mu^\theta \rceil = 4$. This measurement discretization makes the inference problem more difficult and less accurate.

2.5. General mixing example

If equal quantities of K samples are pooled (where $K < N$), the corresponding mixing relationship is

$$2^{-t_\mu^\theta} = \frac{1}{K}\sum_{k=1}^{K} 2^{-t_{\mu k}^\theta} \tag{2.3}$$

where $t_{\mu k}^\theta$ represents the real-valued threshold of element k in the mix and t_μ^θ the real-valued threshold for test μ.

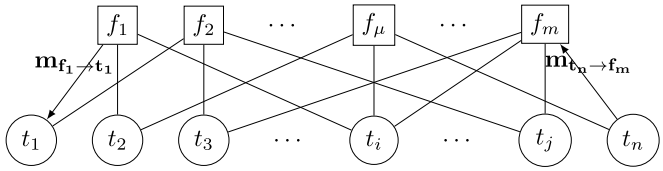

Figure 3. Random variables (circles) represent each of the N patients tested. Each factor node (squares) represents the compatibility of the test/measurement μ with the relevant patient variable values. The degree of the factors nodes is K and of the variable nodes is L.

2.6. Signal sparsity assumption

In traditional testing, we perform N diagnostic tests for N patients. Attempts to reduce the number of tests i.e. $M < N$ will lead to an ill-defined system of equations in figure 1 with an infinite number of solutions. The constraint required to solve the problem typically relies on a sparsity assumption (in some suitable basis) i.e. a majority of zero entries corresponding to non-infected patients. In the present case, uninfected patients correspond to a cycle number equal to the upper limit. In typical compressed sensing scenarios, a sparse prior is defined via the signal density ρ using a Bernoulli-Gauss distribution [9] such as $p(\mathbf{t}) = \prod_{i=1}^{N} [(1-\rho)\,\delta(t_i) + \rho\,\mathcal{N}(\mu, \sigma^2)]$ and i.i.d. signal components. In this paper, the variables used are all integers, corresponding to the PCR cycles where infection can be first detected. They range from the lowest number of cycles L to the highest U. In the absence of additional information we assume the majority of signal entries take the value U (controlled by the sparsity parameter, ρ) and the remainder are uniformly distributed values in the state space $\mathcal{S} \in \{L, L+1, \ldots, U-1\}$, according to

$$p(\mathbf{t}) = \prod_{i=1}^{N} \{(1-\rho)\delta_{t_i,U} + \rho\,\Theta[t_i - L]\,\Theta[(U-1) - t_i]\} \tag{2.4}$$

where $\Theta(x)$ is the Heaviside step function and U represents no viral load/not infected.

An alternative approach, used in [5], is to discretize the entire signal range $\mathcal{S} \in \{L, \ldots, U\}$ into low, medium, high and non-infected ranges corresponding to clinically relevant ranges of low, mild and high infection status. This simplification simplifies the inference problem at the expense of accuracy.

3. Message passing

We aim to recover the N-dimensional vector $\mathbf{t}$ where each component $t_i \in \mathcal{S}$. The viral load variables only participate through the mixing relationship described in section 2.5. The inference problem can be represented by a bipartite factor graph (see figure 3) and since the graph is sparsely connected, inference can be efficiently achieved using a message passing algorithm, whereby conditional probabilities are iteratively exchanged between factors and variables until they converge to provide pseudo marginal posterior probabilities for the individual variables. Message passing methods are exact on trees but offer approximate solutions on loopy graphs [28]. Their use in the context of group testing has been studied previously but mostly in the case of binary infection status to our knowledge [8, 29]. While preparing the manuscript, we came across the paper [5] which utilizes a message passing algorithm for inference of a real-valued signal. However, there is a distinct difference between our approach and the one of [5], that is based on the iterative removal of values found and uses a very small number of categories, arguably making the problem easier.

Messages exchanged between factors and variables represent a closed set of equations of messages from factors to variables $p(f_\mu|t_i) \equiv m_{\mu\to i}(t_i)$ and from variables to factors $p(t_i|\{\mathbf{f}\}_{\setminus\mu}) \equiv m_{i\to\mu}(t_i)$, detailed below.

3.1. Factor-to-variable messages (PCR noise)

Here we denote the noise model as $\phi(\mathbf{t}, \mathbf{t}^y)$ where $\mathbf{t} \in \{t_1, t_2, \ldots, t_K\}$. The mechanism needed for the discretization inherent in the PCR measurement process is developed in section 4.

$$m_{\mu\to i}^{(\tau+1)}(t_i) \propto \sum_{\mathbf{t}\setminus t_i} \phi(\mathbf{t},\, \mathbf{t}^y) \prod_{j\in\partial\mu\setminus i}^{N} m_{j\to\mu}^{(\tau)}(t_j) \tag{3.1}$$

The superscript τ denotes the iteration step and the notation $\partial\mu\setminus i$ refers to all variables connected to factor μ except i. Each message is an $|\mathcal{S}|$-dimensional column vector with entries being real numbers between 0 and 1 representing the state probability. The summation is over all possible states of the vector $\mathbf{t}$ excluding t_i. This is the main computational bottleneck and does not scale well with $|\mathcal{S}|$ (state space size) or K (factor node degree). This difficulty is reduced in [5] by assuming a state space of none, low, medium and high viral loads. For real-valued signals, the messages can be approximated by their means and variances leading to approximate message passing protocols [10, 19].

3.2. Variable-to-factor messages

$$m_{i\to\mu}^{(\tau+1)}(t_i) \propto \{(1-\rho)\delta_{t_i,U} + \rho\,\Theta[t_i - L]\,\Theta[(U-1) - t_i]\} \prod_{\gamma\in\partial i\setminus\mu} m_{\gamma\to i}^{(\tau)}(t_i) \tag{3.2}$$

where δ_{ij} is the Kronecker delta function and the notation $\partial i\setminus\mu$ refers to all factors connected to variable i except μ. For both factor-to-variable and variable-to-factor messages, the constants of proportionality can be calculated by normalisation e.g. $\sum_{t_i\in\mathcal{S}} m_{i\to\mu}(t_i) = 1$.

3.3. Marginal probabilities

Once the messages (3.1) (3.2) have been iterated to convergence, the marginal probabilities for each variable/patient can be calculated:

$$p(t_i) \propto \{(1-\rho)\delta_{t_i,U} + \rho\,\Theta[t_i - L]\,\Theta[(U-1) - t_i]\} \prod_{\mu\in\partial i} m_{\mu\to i}(t_i) \tag{3.3}$$

The inferred C_t value corresponds to the message component with highest probability. It is an approximation in our case since the bipartite graph contains loops. Hence, the time complexity of message passing algorithms is linear in the number of free parameters $O(NK\mathcal{S})$, where N is the number of patients, K the number of tests and $\mathcal{S}$ the number of time intervals. Iterative probabilistic estimation methods of the type used here do not come with a clear estimate of the convergence speed. From our experience this number is small, ranging between 10 and 100 iterations depending on system size; the current system is relatively small.

3.4. Pseudocode

The bipartite graph, G, described in figure 3, initialises the pooling matrix. The two messages associated with each edge (factor-to-variable and variable-to-factor) are initialised using random draws from the uniform distribution in the interval (0, 1). Other practical considerations for initialising parameters are described in section 5.

The iterative solution process is contained in the while loop where the *conv* variable is sum of squared differences between current and old messages. The stopping condition is met when this squared error term is below a user-defined threshold, *thresh* variable.

Algorithm 1. Message Passing Algorithm

Input: Graph G, convergence threshold $=$ *thresh*
Output: Final beliefs
for each edge $(i, \mu) \in G$ **do**
  $m_{i\to\mu}(t_i) \leftarrow$ *initial_message*
  $m_{\mu\to i}^{(\tau)}(t_i) \leftarrow$ *initial_message*
end for
while *conv* $<$ *thresh* **do**
  for each node $i \in G$ **do**
    for each neighbor μ of i **do**
      $m_{\mu\to i}(t_i) \leftarrow$ update_message using (3.1)
      $m_{i\to\mu}(t_i) \leftarrow$ update_message using (3.2)
    end for
  end for
  Calculate *conv* using squared error between old and new messages
end for
Return marginal probabilities: belief for each node i using (3.3) $= 0$

Our algorithm and data will be available from the Aston open access archive.

4. Developing the noise model

The two types of noise present in the PCR mechanism are discretization noise and physical measurement noise. We will address the former in this section and the latter in section 5.6. Discretization noise is present since C_t values are recorded as integers rather than decimals and requires careful treatment in the mathematical modelling setup.

4.1. No measurement noise

In the case where the real cycle values are applied (denoted by the θ superscript) equation (2.3) holds, and consequently

$$\phi(\mathbf{t},\,\mathbf{t}^y)=\prod_{\mu=1}^{M}\;\delta\left(\frac{1}{K}\sum_{k=1}^{K}2^{t_\mu^\theta-t_{\mu k}^\theta}-1\right) \tag{4.1}$$

However, discretization to the next higher integer cycle value creates a difference between the test cycle integer value and those of the individual samples. The modelling of this difference is a challenge and an appropriate noise model has to be used.

4.2. Gaussian measurement noise

One of the simplest models to accommodate these differences is to assume measurement errors follow a Gaussian distribution incorporating (2.3) and the discretization (see [7]). We will replace the 'true' but unknown real cycle values t_μ^θ and $t_{\mu k}^\theta$ by the integer variables t_μ^y and $t_{\mu k}$, respectively. We expect that $t_\mu^y=\lceil t_\mu^\theta\rceil$ and $t_{\mu k}=\lceil t_{\mu k}^\theta\rceil$.

$$\phi(\mathbf{t},\,\mathbf{t}^y)=\prod_{\mu=1}^{M}\;\frac{1}{\sqrt{2\pi\Delta_\mu}}\exp\left[-\frac{1}{2\Delta_\mu}\left(\frac{1}{K}\sum_{k=1}^{K}2^{t_\mu^y-t_{\mu k}}-1\right)^2\right] \tag{4.2}$$

where Δ_μ is the variance in the noise measurement.

We will consider two different noise distributions $\phi(\mathbf{t},\mathbf{t}^y)$ to 'filter out' unreasonable combinations of $\mathbf{t}$ and $\mathbf{t}^y$ from the factor-to-variable messages in (3.1). The first is a simple binary function in section 4.4 and the second uses the overlap between distributions to weight message products in section 4.5.

4.3. Comparison with optimisation classification approaches

It is worthwhile to briefly compare generic algorithmic optimisation methods with our probabilistic message passing approach. The probabilistic approach is based on likelihood which corresponds to the probability of the model producing the data, which includes an explicit noise model, and a prior. In our framework, the highly nonlinear and non-symmetric noise model, comparing data and model prediction plays a crucial role in determining the messages and resulting marginal probabilities. The model itself has very few parameters and aims to achieve the typical case result. Optimisation-based classification methods are typically based on minimising a quadratic error under weighted constraints using some optimisation algorithm, such as gradient descent, e.g. [27]. The quadratic loss function translates into an implicit log-normal noise assumption in the probabilistic approach, which does not necessarily represent an accurate noise model. A couple of other detrimental aspects of the optimisation approach are the arbitrary weight given to the constraint, the inability of gradient descent method to deal with non-smooth constraints, which requires a modification of the norm used in the constraints, and the strong dependence of their performance on the parameters used, e.g. the step-size coefficient. We are fully aware that these are commonly used practices and one can usually find sensible parameters that work in most cases. Nevertheless, we would like to point out that the different methods have different strengths and weaknesses.

4.4. Step-function distribution

First consider the measure d inspired by (2.3) but using integer rather than real values.

$$d=\left(\frac{1}{K}\sum_{k=1}^{K}2^{-t_k}\right)-2^{-t^y}=X-Y \tag{4.3}$$

where we will use short-hands $X\equiv\frac{1}{K}\sum_{k=1}^{K}2^{-t_k}$ and $Y\equiv 2^{-t^y}$ which are fixed for a given combination. The uncertainty in both random variables X and Y can be represented by uniform distributions $\mathcal{U}(X,\,2X)$ and $\mathcal{U}(Y,\,2Y)$ leading to an inequality:

$$\frac{1}{2}2^{-t^y}\leqslant\frac{1}{K}\sum_{k=1}^{K}2^{-t_k}\leqslant 2\cdot 2^{-t^y} \tag{4.4}$$

Hence our first approximation for $\phi(\mathbf{t},\mathbf{t}^y)$ can be written:

$$\phi(\mathbf{t},\,\mathbf{t}^y)=\begin{cases}1 & \text{if } -\frac{1}{2}2^{-t^y}\leqslant d\leqslant 2^{-t^y}\\ 0 & \text{otherwise.}\end{cases} \tag{4.5}$$

Notice this function does not need to be normalized since the messages are normalized in the message passing iterations. Clearly $d=0$ is in the interval as expected. The message passing summation (3.1) includes all

combinations of $\{t_k\}$ for a given t_y and the role of $\phi(\mathbf{t}, \mathbf{t}^y)$ is to exclude implausible combinations. This simple binary approach has the benefit of materially reducing the number of summands in each message passing iteration speeding up implementation.

4.5. Distribution overlap

Rather than a binary function, consider a function $\phi(\mathbf{t}, \mathbf{t}^y)$ which weights different combinations of $\{t_k\}$ and t_y depending on the overlap of the distributions $\mathcal{U}(X,\ 2X)$ and $\mathcal{U}(Y,\ 2Y)$. For the case $X < Y$:

$$\phi(\mathbf{t},\ \mathbf{t}^y) = \begin{cases} 0 & \text{if } X \leqslant \frac{1}{2}Y \\ \frac{2X - Y}{Y} & \text{if } \frac{1}{2}Y \leqslant X \leqslant Y \\ \frac{2Y - X}{Y} & \text{if } Y \leqslant X \leqslant 2Y \\ 0 & \text{if } X > 2Y \end{cases} \tag{4.6}$$

This noise distribution, which has the same support as (4.4), is used in the numerical experiments of section 5.

4.6. Further refinement

Using a uniform distribution for 2^{-t^y} is plausible but the distribution for $X = \frac{1}{K}\sum_k 2^{-t_k}$ should properly account for each t_k chosen at random from $\{L, \ldots, U\}$. The probability density function has been derived in [30, 31]. The cumulative distribution function for X, derived in [32], can be used to calculate the overlap with $\mathcal{U}(2^{-t^y},\ 2 \cdot 2^{-t^y})$. This calculation is not implemented due to the increased computational cost required due to the nested summations.

4.7. Alternative treatments

Previous research has considered the PCR noise function but primarily in the case of binary variables (infected/non-infected). The traditional noise measures of false positive and false negative rates are comprehensively treated in [8] where it is shown that pooled testing can overcome technical device limitations since each patient participates in multiple tests. Noise in C_t values and fluoresence thresholds are dealt with in [7] but with a different signal reconstruction algorithm and the discretization effect is modelled as Gaussian rather than uniform. We have avoided the need for calibration of the threshold θ and mapping between amplified viral load and fluoresence in [7] with our functional form of $\phi(\mathbf{t}, \mathbf{t}^y)$. An alternative approach is to learn the translation function between viral loads and discrete measurement classes using supervised learning methods such as a neural network [5]. While our approach splits this up into mixing (2.3) and noise probability functions, [5] may be able to learn from previous PCR results *if the ground truth properties are known*.

5. Numerical Experiments

5.1. Implementation

The problem is defined by initializing parameters ρ, M, N, K and state space $\mathcal{S}$. Of course, the signal sparsity ρ is not a priori known for a given batch of samples but could be estimated by either the Expectation-Maximization [8, 9] or the Expectation Propagation [33] methods. An approximate but practical approach is to take a recent average disease prevalence value as an estimate of ρ. This parameter is used to set the number of observations M in the pooling matrix.

We now carefully describe the protocol for generating the ground truth signal $\mathbf{t}$ in our numerical experiments. This is important as it influences our subsequent choice of noise function boundary. *Decimal* values are sampled for infected and non-infected patients from uniform distributions $\mathcal{U}(L - 1,\ U - 1)$ and $\mathcal{U}(U - 1,\ U)$ respectively and combined using (2.4). The *decimal* measurement vector t_y^θ is calculated *in silico* using the PCR mixing protocol (2.3). Finally an *integer* measurement value is obtained by rounding up t_y^θ to the nearest integer. Additional noise is added (see section 5.6) in some scenarios as described below. The discretized version of $\mathbf{t}$ is stored for calibration purposes once the inference is complete. A different (unphysical) protocol could be considered which starts from *integer* values of infected patients. This would reduce uncertainty in the noise function thus providing more flattering accuracy statistics but is not used in our experiments since it unrealistic.

The random binary measurement matrix $\boldsymbol{F}$ defines which samples are included in each test and is typically defined by the degree distribution of the variable and factor nodes of the bipartite graph (figure 3). Choices include random-random or random-Poisson [29]. In this paper, we follow [8] in using random-random matrices such that each row has K non-zero entries and each column has $L = KN/M$ non-zero entries. In other

words, the factor nodes have a constant degree K and the variable nodes a constant degree L. The values used in our setup result in a sparse $\boldsymbol{F}$ matrix. Our message passing algorithm could also be combined with the structured matrix approach [4, 9, 34].

Given knowledge of $\boldsymbol{F}$ and $\mathbf{y}$, the message passing equations (3.1), (3.2) are iterated until convergence. The message passing equations search for a fixed point to the dynamical system of messages, from which posterior variable values can be inferred using (3.3). Typical of iterative problems, our implementation utilises a damping factor (in our case, set to 0.01) to help convergence. In addition, the sum of squared differences between all messages is chosen as a metric to determine convergence. The threshold was chosen empirically as 10^{-5} for the noiseless cases. This synthetic setup can therefore be used to test the accuracy of signal recovery.

5.2. Easy algorithm

In problems where only binary infection status is considered, *sure variables* of certain states can be determined using simple logical arguments without resorting to more complex inference methods. These are termed *Combinatorial Basis Pursuit (CBP)* or *Combinatorial Orthogonal Matching Pursuit (COMP)* in computer science implementations of group testing [35] and are described in an *Easy Algorithm* [29]. Briefly, they identify certain (sure) negative patients since they participate in a negative test result. Further, sure positive samples can be found as those who participate in positive tests alongside sure negative samples. These variables are then removed reducing the problem size. One drawback of this approach is the inability to infer *undetermined variables* [29], the inability to use prior knowledge and to exploit the power of probabilistic inference in the inevitable presence of measurement noise.

In the present case, where the state space $\mathcal{S}$ represents C_t values, a non-infected test measurement result, $t^y_\mu = U$, does not guarantee all samples are non-infected (even in the noiseless case) e.g. true C_t values of $\{U-1, U, U, U, U\}$ will result in $t^y_\mu = U$. This is due to the peculiarity of the PCR mixing protocol (2.3) and the discretization of C_t values. The issue can be seen clearly for the low viral load values (equivalently high C_t values) in figure 4. Hence the *Easy Algorithm* cannot be used in our work, either as a pre-processing step or as a comparison baseline for our message passing results. In fact it is arguably impractical to any scenario that includes a measurement noise.

5.3. Uncertain inference

Recall that each message consists of $|\mathcal{S}|$ components representing the probability of each state from L to U. To be clear, each state of $\mathcal{S}$ represents a doubling of the viral load from the preceeding state on a logarithmic scale (base 2). Our Maximum A Posteriori *(MAP)* inference protocol uses the highest probability state as our diagnosis. However, for some samples, two states could have similar high probabilities implying uncertain inference. A *computational* improvement could be envisaged, using decimation, where, upon convergence, the less ambiguous results are fixed to a standard basis vector e.g. $(0, 1, 0, \ldots, 0)^T$ and the message passing algorithm continues with a smaller convergence threshold. Note, the 'certain' samples are simply the complement of the uncertain samples. This intervention in the algorithm dynamics did not meaningfully change the results in our test implementations. A possible *clinical* protocol would be to physically re-test this small number of individuals with ambiguous results, but this has its own operational and medical implications.

5.4. Motivating example

To illustrate the discretization problem central to PCR testing, figure 4 shows the confusion matrix for the noiseless inference problem of $N = 2400$ individual samples, $M = 800$ observations, $K = 6$ patients in each group, $\rho = 0.01$ prevalence and $\mathcal{S} \in \{20,\ldots,30\}$. As a guideline, the expected number of tests required in adaptive *binary (yes/no only)* Dorfmann tests is approximately $2\sqrt{\rho}N = 480$ (see [29]).

Predicting a C_t value one away from its true value is clinically acceptable leading to the formulation of two accuracy measures C_0 and $C_{\pm 1}$ corresponding to a tolerance of zero and $\pm$ 1 C_t value, respectively. However, given the imbalanced classes in the signal, this metric is found not to be suitable to our task since simply predicting no infection results in high accuracy.

For the remainder of this paper, we will focus on the clinically relevant metric of sensitivity (the ratio of true positives to positive samples) and use a state space comprising $\mathcal{S} \in \{20,\ldots,30\}$. The relationship between sensitivity and $|\mathcal{S}|$ is investigated in figure 5.

5.5. Accuracy of inference procedure

We carry out numerical experiments with a fixed measurement ratio $\alpha = M/N = 800/2400 = 0.33$ and vary the signal sparsity in the range $\rho = [0.01, 0.05]$ to represent typical infectious disease values. 30 simulations are run for each ρ value with C_0 and $C_{\pm 1}$ accuracy, true/false positive/negative measures recorded. The high accuracy

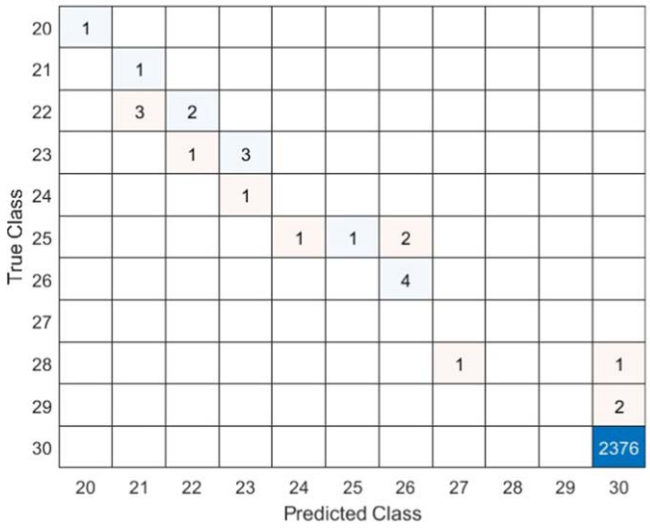

Figure 4. Confusion matrix representing a single problem instance, described in section 5.4, using synthetic data: $M = 800, N = 2400, K = 6, \rho = 0.01, \eta = 0$ and $\mathcal{S} \in \{20,\ldots,30\}$. The noise function $\phi(\mathbf{t}, t_y)$ is taken as described in section 4.5, producing an accuracy of 2388/2400 or 99.5% (compared to 99.0% for a always-negative classifier). The false negative rate of $3/24 = 12.5\%$ reduces as the state space $\mathcal{S}$ increases (see figure 5).

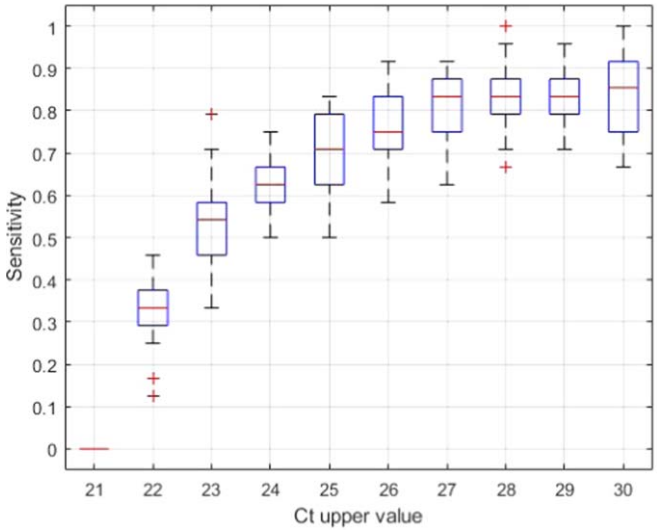

Figure 5. Plot of Sensitivity (True Positive Rate = TP/P) against the largest C_t value. The lower C_t value is fixed at 20. The intuition from our motivating example in figure 4 is false negatives occur for C_t values close to the maximum. This plot confirms false negatives decrease as the range size become larger the number of values per variable is higher. The zero sensitivity value at $C_t = 21$ corresponds to all 24 positive samples being falsely classified as negative. Parameters $\rho = 0.01$, $\alpha = M/N = 800/2400$ are fixed. Changing a true to a false positive results in a sensitivity change of $1/24 \approx 4\%$. Simulations are run 30 times. Red lines display the median value, the boxes cover the interquartile range and the whiskers show the extreme values.

values shown in figure 6 mask an underlying issue with the false negative rates hence our focus on sensitivity. Although an error of $\pm 1 C_t$ value is an order of 2 error, it is unlikely to affect the clinical decision process.

5.6. Accuracy versus measurement noise

In previous sections, we have accounted for the discretization noise arising from the PCR doubling mechanism. Physical noise sources, also inherent in recording t^y values, include fluorescence not being distinguished from background levels e.g. light leaks into the sample well and imperfect doubling during the amplification phase e.g. $(1+q)^t$ where $q \in (0, 1)$ rather than 2^t [7]. These tend to lower and raise the measured C_t value from its true reading respectively. This PCR efficiency factor can be incorporated into our analysis by replacing the factor of 2 with $1 + q$. The issue of estimating q is addressed in [27]. In the absence of accurate statistics, we naively assume under- and over- C_t estimation is equiprobable leading to a noisy measurement $\tilde{t}^y$:

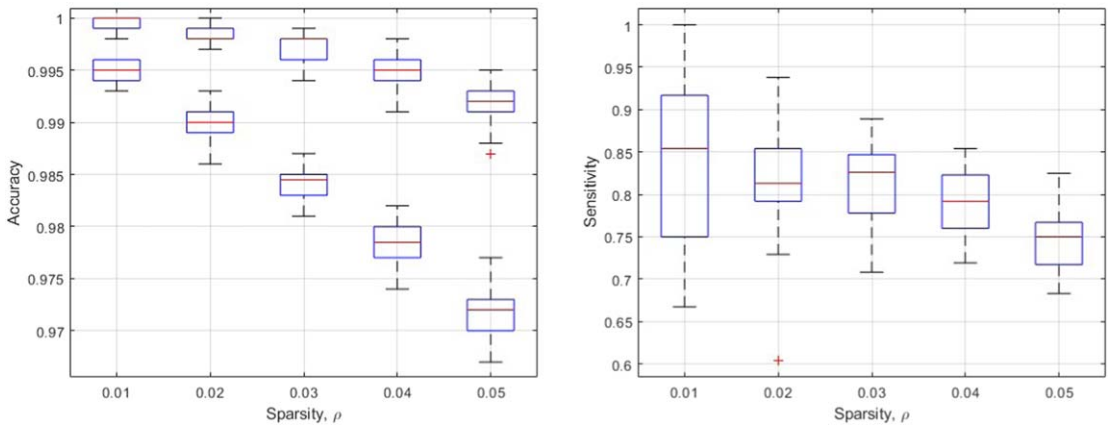

Figure 6. Left: Plot of C_0 (lower) and $C_{\pm 1}$ (upper) accuracy measures against signal sparsity ρ. Right: Plot of Sensitivity against ρ. Note that specificity (True Negative Rate = TN/N) is $\geqslant$0.999 for $\rho \in [0.01, 0.05]$ so is not plotted. The state space is $\mathcal{S} \in \{20,\ldots,30\}$. Simulations are run 30 times and the red lines display the median value, the boxes cover the interquartile range, the whiskers show the extreme values and outliers are shown with red '+' symbols. The signal sparsity/prevalence range (x-axis) is chosen to represent typical infectious diseases values. The measurement ratio α is constant at $M/N = 800/2400$ and no measurement noise is added ($\eta = 0$, see discussion in section 5.6). These plots correspond to the experiment described in section 5.5.

$$p(\tilde{t}^y) = (1-\eta)\,\delta(\tilde{t}^y - t^y) + \frac{1}{2}\eta\,\delta[(\tilde{t}^y - t^y) - 1] + \frac{1}{2}\eta\,\delta[(\tilde{t}^y - t^y) + 1] \tag{5.1}$$

where $\eta \ll 1$. Setting $\eta = 0$ represents no physical measurement noise (which was the case for our previous experiments). Equation (5.1) is used to add noise to our synthetically generated signals. After convergence of our message passing equations, the inferred C_t values are converted into binary variables where negative corresponds to $C_t = U$ and positive otherwise. The resulting sensitivity values, plotted in figure 7, assess how robust the inference is to increasing η. The presence of overlapping tests i.e. each patient participating in multiple tests ($L = KN/M > 1$) was considered in [8] and was found to overcome typical noise levels.

In addition to device-specific measurement noise, modelled by (5.1), laboratory sample handling errors can occur [36]. These include cross-contamination of samples and errors pipetting samples into physical groups/pools. This source of error is not accounted for in our model. Another source of noise inherent in the PCR protocol is introduced by the *limit of detection*. This can be modelled using the *stochastic threshold group testing* method where the result is negative when the number of infected samples in the pool is no larger than l, positive if the pool contains more than u infected samples, and stochastic in the interval $[l, u]$ (see [37] for a detailed description). This setup could be incorporated into our message passing equations but instead we focus on our primary objective of recovering viral load signals with an extremely large dynamic range.

5.7. Comparison to other decoding protocols

5.7.1. Qualitative comparison

To our knowledge, there are three studies describing non-adaptive pooled testing with ranges of viral loads [5, 20] or C_t values [27]. Table 1 qualitatively compares these to the current work. Pooled testing with binary values [4, 14] is not comparable to the current method.

5.7.2. Quantitative comparison

It is difficult to reproduce the results reported in [20, 27, 38] due to the lack of available code, data and detailed protocols. Therefore, to compare the results obtained against a generic commonly used approach we have modified the Combinatorial Orthogonal Matching Pursuit (COMP) algorithm to account for discrete C_t values in the noiseless non-adaptive setting.

The COMP algorithm (described in section 5.2). All items in a negative test are *Definitely Non-Defective (DND)* and all other items are *Possibly Defective (PD)*. If a test contains only one *PD* item, then that item is *definitely defective (DD)*. The COMP algorithm declares *DND* items to be non-defective and the remainder defective. The *DD* algorithm finds all *DD* items are defective and the remainder non-defective. Here *defective* means infected/presence of viral load and *Non-Defective* means not infected/absence of viral load. This algorithm works in the noiseless setting.

The modified COMP algorithm. The N samples are divided into $N/3$ distinct non-overlapping sets i.e. exactly 3 samples per test resulting in a pooling matrix with components

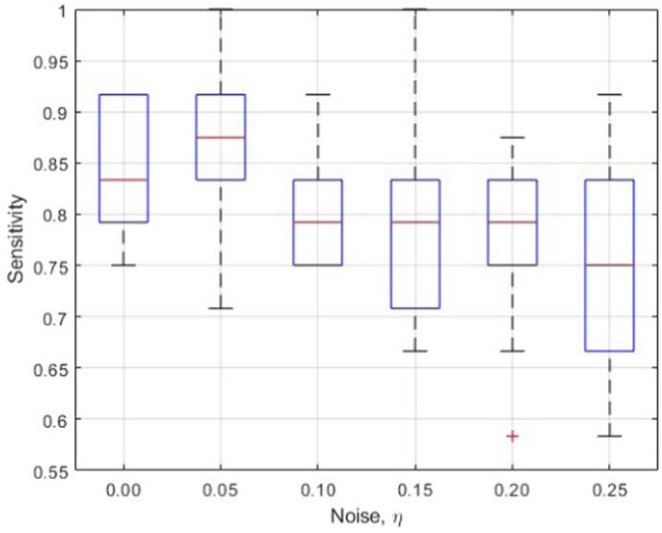

Figure 7. Sensitivity against noise level η defined in (5.1). This corresponds to experiment described in section 5.6. Parameters $\rho = 0.01, \alpha = M/N = 800/2400$ are fixed. Simulations are run 30 times. The necessity of more iterations for the noisy signals is well known [29] so it was necessary to relax the convergence threshold to 10^{-3}.

Table 1. Comparison of inference algorithm protocols.

Reference	Signal type	Decoding	Noise model
[5]	Discrete - none, low, mid, high viral load	Loopy Belief Propagation	Neural network classifier
[12]	C_t values	*Underestimate, Match, Explain* and *Dark Room*	Dependent on *Schedule Matrix, S*
[20]	Discrete - none, low, mid, high viral load	*Definitely not defective (DND)* [38], combinatorial *Maximum Likelihood (ML)* [39], followed by viral load estimation	Some stochastic mapping f: $\mathbb{R}^+ \rightarrow \mathbb{R}^+$"
[27]	C_t values	Optimisation (gradient descent and iterative mirrored hard thresholding methods)	log-normal noise
[40]	C_t values	*Tropical DD, COMP* and *SCOMP*	Similar to classical algorithm counterparts
Current paper	C_t values	Probabilistic (Bayesian message passing)	PCR-specific

$$F_{\mu i} = \delta_{i,3\mu-2} + \delta_{i,3\mu-1} + \delta_{i,3\mu} \qquad 1 \leqslant \mu \leqslant N/3, \quad 1 \leqslant i \leqslant N \tag{5.2}$$

where $\delta_{i,j}$ is the Kronecker delta function. Each sample only participates in one test and so no information about the samples is shared between tests. This differs from most pooled testing set ups. The following algorithm was developed by the authors but reviewers of this manuscript have kindly pointed out the similar approaches of [12] and [40] using *tropical arithmetic*.

Assume the state space $\mathcal{S} \in \{L, L+1, \ldots, U\}$. The difference with the traditional COMP algorithm is how outputs $\{L, \ldots, U-1\}$ are treated. The probability of having k = 1, 2 or 3 defective ($C_t < U$) members can be calculated from the binomial distribution $\binom{3}{k}\rho^k(1-\rho)^{3-k}$ as 0.03, 0.0003 and 0.000001 respectively. Hence we make the simplifying assumption that each test contains exactly one defective. This infected C_t value, t, can be calculated from the pooled measurement output, t_{out}, and the mixing equation (2.3) as

$$t = \lceil (U-1) - \log_2(3 \times 2^{(U-1)-t_{out}} - 1) \rceil \tag{5.3}$$

The *location* of the defective cannot be determined amongst the three slots (see (5.2)) so all three samples are labelled *Possibly Defective (PD)*. By construction of our non-overlapping tests, we cannot determine *Definitely Defectives (DD)*.

We make the assumption that it is clinically preferable to label a sample infected when it is not, than not infected when it is infected i.e. reducing the number of false negatives (at the expensive of more false positives).

Therefore all three samples in the test are labelled with the C_t value calculated from (5.3), increasing the sensitivity at the expense of the specificity.

Algorithm 2. Modified COMP Algorithm

Input: Number of samples per test = 3, sparsity, $\rho = 0.01$
Output: Inferred C_t values
$\mathbf{x} \leftarrow$ initial $N \times 1$ signal vector with sparsity, ρ
$\hat{\mathbf{x}} \leftarrow$ inferred $N \times 1$ signal vector
for each test μ **do**
$F_{\mu i} \leftarrow \delta_{i,3\mu-2} + \delta_{i,3\mu-1} + \delta_{i,3\mu}$ from (5.2)
end for
for each test μ **do**
$\hat{\mathbf{x}}_{3\mu-2}, \quad \hat{\mathbf{x}}_{3\mu-1}, \quad \hat{\mathbf{x}}_{3\mu} \leftarrow$ update using (5.3)
end for
Return Inferred signal vector, $\hat{\mathbf{x}} = 0$

To allow direct comparison with our numerical simulations (section 5), model parameters are set to $M = 800$, $N = 2400$, $\rho = 0.01$, $\eta = 0$ and $\mathcal{S} \in \{20, 21, \ldots, 30\}$. With this sparsity value, approximately 97% of the 800 tests have a value of 30 (using the Binomial distribution in the noiseless case) and approximately $0.97 \times 800 \times 3 = 2,328$ samples are *Definitely Non-Defective (DND)*. As discussed above, the remaining tests will have one expected defective sample with inferred C_t:

$$t = \lceil 29 - \log_2(3 \times 2^{29 - t_{out}} - 1) \rceil \tag{5.4}$$

This simple modification to the COMP algorithm has been constructed to account for C_t values and to allow for a comparison to section 5. In this framework, it may not be practically possible to increase the number of samples per test much above three. The confusion matrix for a single simulation is shown in figure 8 with roughly the expected number of false positives (46, close to the expected 48). Despite the simple algorithm, most infected samples are either predicted correctly or one C_t value away.

To allow comparison with Algorithm 1, 30 simulations of Algorithm 2 were run and mean accuracy metrics recorded in table 2. Metrics C_0 and $C_{\pm 1}$ are defined in section 5.4. As a sanity check, a naive classifier, where all predictions are negative, is included (sensitivity $= TP/P = 0/24 = 0$, specificity $= TN/N = 2376/2376 = 1$).

Some studies include the COMP algorithm as a pre-processing step to the decoding protocol. Since the COMP logic only works in the noiseless case and our noise model (section 4) is an integral part of our investigation, we do not combine the modified COMP and message passing.

6. Discussion

Pooled testing was originally designed for the efficient diagnosis of infectious diseases [1]. There are numerous explanations for the lack of widespread adoption in public health including perceived complexity, requirement for sophisticated laboratory management or conflicting economic incentives. By combining pooled testing efficiencies with accurate estimation of a realistic range of viral loads, the methods presented here may lead to wider adoption into applications such as food contaminant testing and community screening for infectious diseases.

It is known that the viral load present in infectious disease testing samples can vary by many orders of magnitude [11]. This concept has typically not been taken into account in theoretical investigations of pooled testing. We estimate the viral load by mapping the inference task to a message passing problem where the factor nodes incorporate the specific PCR mixing protocol. Numerical experiments explicitly highlight the source of the error originating from mixing patient samples and we have dealt with this through a modified noise function $\phi(\mathbf{t}, \mathbf{t}^y)$. This filter, used in section 4.4, approximates the distributions of random variables $X \equiv \frac{1}{K}\sum_{k=1}^{K} 2^{-t_k}$ and $Y \equiv 2^{-t^y}$ as step functions with the overlap providing weights in the message passing equations. For completeness, we note that for applications where the viral load actually has a narrow range of values, the well-developed theory of compressed sensing [9, 10] can be applied to infer signal values, accurately and efficiently, using either random or structured pooling matrices.

It is clearly important to compare our results to other algorithms. However, the scarcity of comparable approaches led us to create a modified version of the COMP algorithm to account for C_t values. The lack of comparable algorithms also suggests that our probabilistic approach is a valid addition to the existing optimisation literature.

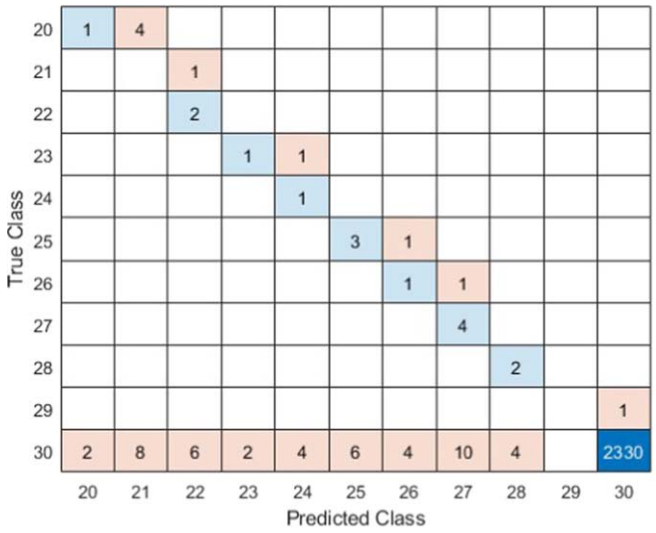

Figure 8. Confusion matrix for a single instance of the modified COMP algorithm described in Algorithm 2. Since the exact location of infected samples cannot be determined within the test, all three samples in the *defective* test are labelled *defective*. Hence there are expected to be 2 false positives per *defective* test i.e. a total of approximately 48 false positives. In this instance, there are 46.

Table 2. Comparison of inference algorithms. Accuracy figures are mean values calculated from 30 numerical simulations. The two accuracy measures C_0 and $C_{\pm 1}$ correspond to a tolerance of zero and $\pm$ 1 from the true C_t value.

	C_0	$C_{\pm 1}$	Sensitivity	Specificity
Message Passing (Algorithm 1)	0.995	0.999	0.854	1.000
Modified COMP (Algorithm 2)	0.977	0.984	0.946	0.982
Naive classifier	0.990	$\approx$ 0.990	0.000	1.000

The main focus of future work is to improve the scalability of algorithm. While real-valued signals can be approximated via the first two moments of the message distributions, the approximation for our integer-valued state space problem is not clear. The resulting summation in the factor-to-variable messages (3.1) does not scale well with $|\mathcal{S}|$ (increasing C_t range) or K (number of patients in each pool). Our choice of noise function mitigates this effect by materially reducing the number of summands in the factor-to-variable messages. An alternative approach is to translate the integer C_t values to broad classes of infection status e.g. [5]. Further code efficiencies or parallelization may also help.

Other avenues of future work include a wet laboratory experiment to test our modelling assumptions. The methods described in this paper can be applied to other amplification protocols such as Loop-mediated isothermal amplification (LAMP).

Acknowledgments

This work was funded by the Engineering and Physical Sciences Research Council through grant EP/W015412/1. The authors would like Professor Andrew Beggs for an earlier collaboration on group testing for the SARS-CoV-2 diagnosis.

Data availability statement

All data that support the findings of this study are included within the article (and any supplementary files).

ORCID iDs

Mansoor Sheikh https://orcid.org/0000-0001-6566-5562
David Saad https://orcid.org/0000-0001-9821-2623

References

[1] Dorfman R 1943 The detection of defective members of large populations *Ann. Math. Stat.* **14** 436–40
[2] Mézard M and Toninelli C 2011 Group testing with random pools: Optimal two-stage algorithms *IEEE T. Inform Theory* **57** 1736–45
[3] Shental N *et al* 2020 Efficient high-throughput sars-cov-2 testing to detect asymptomatic carriers *Science Advances* **6** eabc5961
[4] Zhang P *et al* 2013 Non-adaptive pooling strategies for detection of rare faulty items (2013 IEEE Int Conf Comm) 1409–14 (IEEE)
[5] Ben-Knaan *et al* Recovery of noisy pooled tests via learned factor graphs with application to covid-19 testing *ICASSP 2022-2022 Int Conf. Acoust Spee.* 4518–22 (IEEE) 2022
[6] Donoho D L *et al* 2009 Message-passing algorithms for compressed sensing *P Natl. Acad. Sci.* **106** 18914–9
[7] Ghosh S *et al* 2021 A compressed sensing approach to pooled rt-pcr testing for covid-19 detection *IEEE Open Journal of Signal Processing* **2** 248–64
[8] Sakata A 2020 Bayesian inference of infected patients in group testing with prevalence estimation *J. Phys. Soc. Jpn.* **89** 084001
[9] Krzakala F *et al* 2012 Probabilistic reconstruction in compressed sensing: algorithms, phase diagrams, and threshold achieving matrices *J. Stat. Mech. - Theory E.* **2012** P08009
[10] Krzakala F *et al* 2012 Statistical-physics-based reconstruction in compressed sensing *Phys. Rev.* X **2** 021005
[11] Pan Y *et al* 2020 Viral load of sars-cov-2 in clinical samples *Lancet Infect Dis.* **20** 411–2
[12] Wang H-P, Gabrys R and Vardy A 2023 Tropical group testing *IEEE Trans. Inf. Theory* **69** 6098–120
[13] Kiplagat J and Justice A 2022 Hiv viral suppression is key to healthy longevity *Lancet HIV* **9** e672–3
[14] Krajden M *et al* 2014 Pooled nucleic acid testing increases the diagnostic yield of acute hiv infections in a high-risk population compared to 3rd and 4th generation hiv enzyme immunoassays *J. Clin. Virol.* **61** 132–7
[15] Beldomenico P M 2020 Do superspreaders generate new superspreaders? a hypothesis to explain the propagation pattern of COVID-19 *Int. J. Infect. Dis.* **96** 461–3
[16] Liu Y *et al* 2020 Viral dynamics in mild and severe cases of COVID-19 *Lancet Infect Dis.* **20** 656–7
[17] Hay J A *et al* 2021 Estimating epidemiologic dynamics from cross-sectional viral load distributions *Science* **373** eabh0635
[18] Singanayagam A *et al* 2022 Community transmission and viral load kinetics of the sars-cov-2 delta (b. 1.617. 2) variant in vaccinated and unvaccinated individuals in the uk: a prospective, longitudinal, cohort study *Lancet Infect Dis.* **22** 183–95
[19] Kabashima Y 2003 A statistical mechanical approach to cdma multiuser detection: propagating beliefs in a dense graph *In IEEE International Symposium on Information Theory, 2003. Proceedings.* 329–329
[20] Cohen A *et al* 2021 Multi-level group testing with application to one-shot pooled covid-19 tests *In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* 1030–4 (IEEE)
[21] Rangan S 2011 Generalized approximate message passing for estimation with random linear mixing *2011 IEEE International Symposium on Information Theory Proceedings* 2168–72 (IEEE)
[22] Crespo E M *et al* 2018 A review of sparse recovery algorithms *IEEE Access* **7** 1300–22
[23] Manchanda R and Sharma K 2020 A Review of Reconstruction Algorithms in Compressive Sensing *2020 International Conference on Advances in Computing, Communication & Materials (ICACCM (Dehradun, India, 21-22 August 2020)* (IEEE) pp 322–5
[24] Cao S-J *et al* 2023 Group testing with side information via generalized approximate message passing *IEEE Trans. Signal Process.* **71** 2366–75
[25] Goenka R *et al* 2021 Contact tracing enhances the efficiency of covid-19 group testing *In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE) pp 8168–72
[26] Nikolopoulos P *et al* 2021 Group testing for connected communities *In International Conference on Artificial Intelligence and Statistics* 2341–9 (PMLR)
[27] Bharadwaja S and Murthy C R 2022 Recovery algorithms for pooled rt-qpcr based covid-19 screening *IEEE Trans. Signal Process.* **70** 4353–68
[28] Opper M and Saad D 2001 Advanced mean field methods: theory and practice *Neural Information Processing* (MIT)
[29] Mézard M *et al* 2008 Group testing with random pools: Phase transitions and optimal strategy *J. Stat. Phys.* **131** 783–801
[30] Bradley D M and Gupta R C 2002 On the distribution of the sum of n non-identically distributed uniform random variables *Ann. I. Stat. Math.* **54** 689–700
[31] Sadooghi-Alvandi S M *et al* 2009 On the distribution of the sum of independent uniform random variables *Stat. Pap.* **50** 171–5
[32] Buonocore A *et al* 2009 A note on the sum of uniform random variables *Stat. Probabil. Lett.* **79** 2092–7
[33] Braunstein A *et al* 2020 Compressed sensing reconstruction using expectation propagation *J. Phys. A: Math. Theor.* **53** 184001
[34] Angelini M C *et al* 2012 Compressed sensing with sparse, structured matrices *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 808–14 (IEEE)
[35] Chan C L *et al* 2011 Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)* 1832–9 (IEEE)
[36] Courtney P and Royall P G 2021 Using robotics in laboratories during the covid-19 outbreak: a review *IEEE Robot Autom. Mag.* **28** 28–39
[37] Chan C L *et al* 2013 Stochastic threshold group testing *2013 IEEE Information Theory Workshop (ITW)* 1–5 (IEEE)
[38] Chan C L, Jaggi S, Saligrama V and Agnihotri S 2014 Non-adaptive group testing: explicit bounds and novel algorithms *IEEE Trans. Inf. Theory* **60** 3019–35
[39] Atia G K and Saligrama V 2012 Boolean compressed sensing and noisy group testing *IEEE Trans. Inf. Theory* **58** 1880–901
[40] Paligadu V, Johnson O and Aldridge M 2024 Small error algorithms for tropical group testing *IEEE Trans. Inf. Theory* **70**