Contact Tracing Enhances the Efficiency of COVID-19 Group Testing

Ritesh Goenka* [1]   Shu-Jie Cao* [2]   Chau-Wai Wong [3]   Ajit Rajwade [1]   Dror Baron [3]

[1]IIT Bombay, India   [2]ShanghaiTech University, China   [3]North Carolina State University, USA
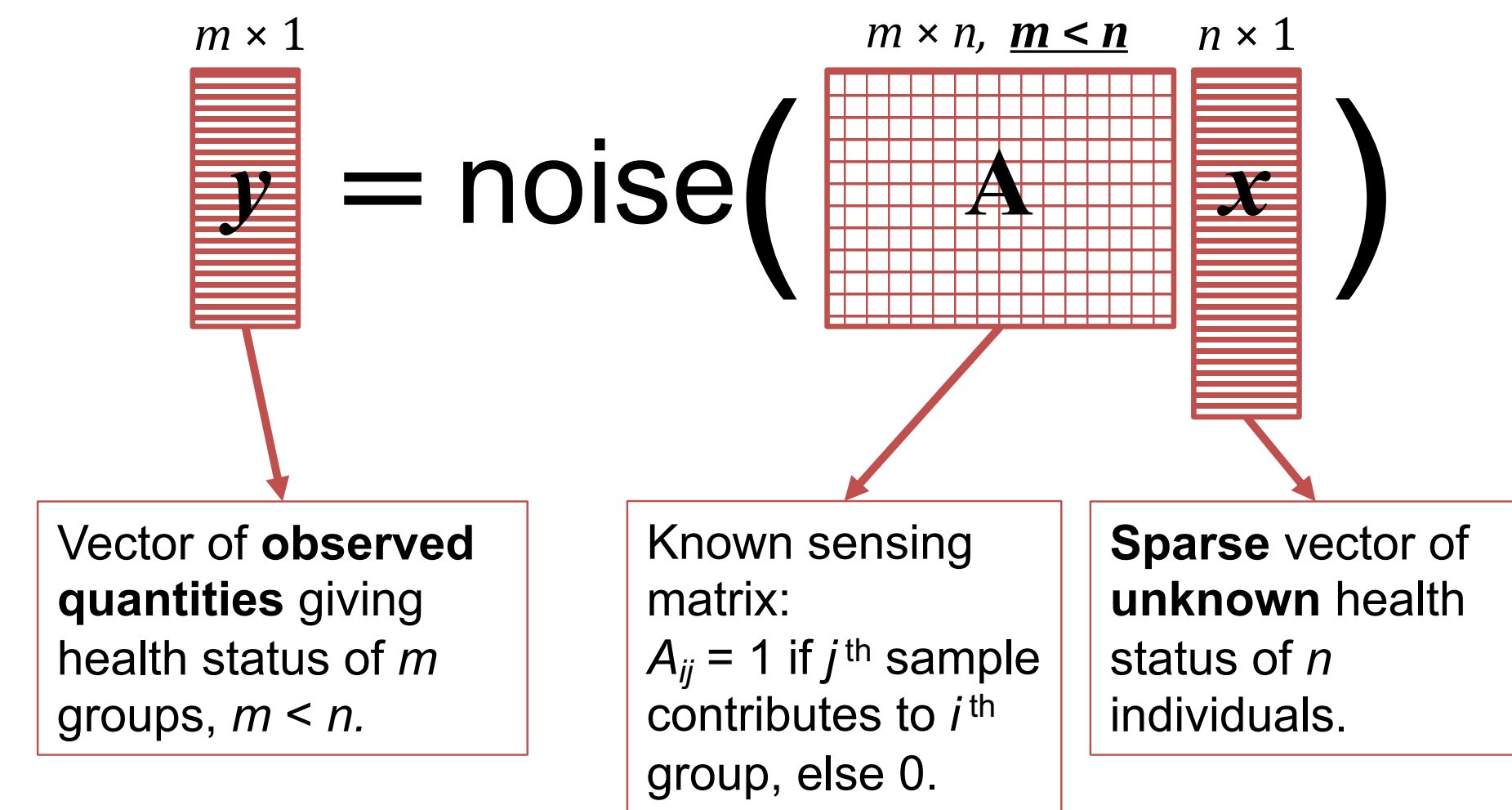
NC STATE UNIVERSITY

Group Testing: Motivation & Compressed Sensing Formulation

- Need widespread testing for COVID-19 – pooled or **group testing** for improving efficiency.
- Initial idea by Robert Dorfman in 1943 – divide individuals into groups of fixed size; if a group tests negative, all individuals in the group are declared healthy, else proceed to individual testing.
- **Compressed sensing** approach to group testing:
  - Mix n individual samples into $m < n$ pools.
  - $y = \mathfrak{N}(Ax)$, $\mathfrak{N}$ is a stochastic noise model.
  - Estimate x given y, A and the noise model.
  - Use of different decoding algorithms.
  - Single stage (**nonadaptive**), unlike Dorfman pooling, which is more time consuming.
  - Can incorporate **side information** (SI) into the probability distribution of x.

$$y = \text{noise}\left(A \cdot x \right)$$

$m \times 1$ ... $m \times n,\ \underline{m \leq n}$ ... $n \times 1$

Vector of **observed quantities** giving health status of m groups, $m < n$.

Known sensing matrix: $A_{ij} = 1$ if jth sample contributes to ith group, else 0.

Sparse vector of **unknown** health status of n individuals.

Research Question

Can group testing efficiency be improved by utilizing side information (SI) in the following forms?
- Family structure: nonoverlapping or overlapping.
- Contact tracing data: (i) who's in contact, (ii) physical proximity, and (iii) contact duration.

Recent Progress

- Compressed sensing formulation solved using a message-passing style algorithm [1].
- Multiplicative noise model with an end-to-end study of decoding in pooled quantitative PCR [2].
- Exploited community structure with a focus on encoder design, but did not consider noise [3, 4].

[1] Zhu, Rivera, and Baron, "Noisy pooled PCR for virus testing," Apr. 2020.
[2] Ghosh et al., "A compressed sensing approach to group testing for COVID-19 detection," May 2020.
[3] Nikolopoulos, Guo, Fragouli, and Diggavi, "Community aware group testing," Jul. 2020.
[4] Nikolopoulos, Srinivasavaradhan, Guo, Fragouli, and Diggavi, "Group testing for overlapping communities," Dec. 2020.

Infection Model

- Use SEIR style dynamical infection model incorporating contact tracing SI for data simulation.
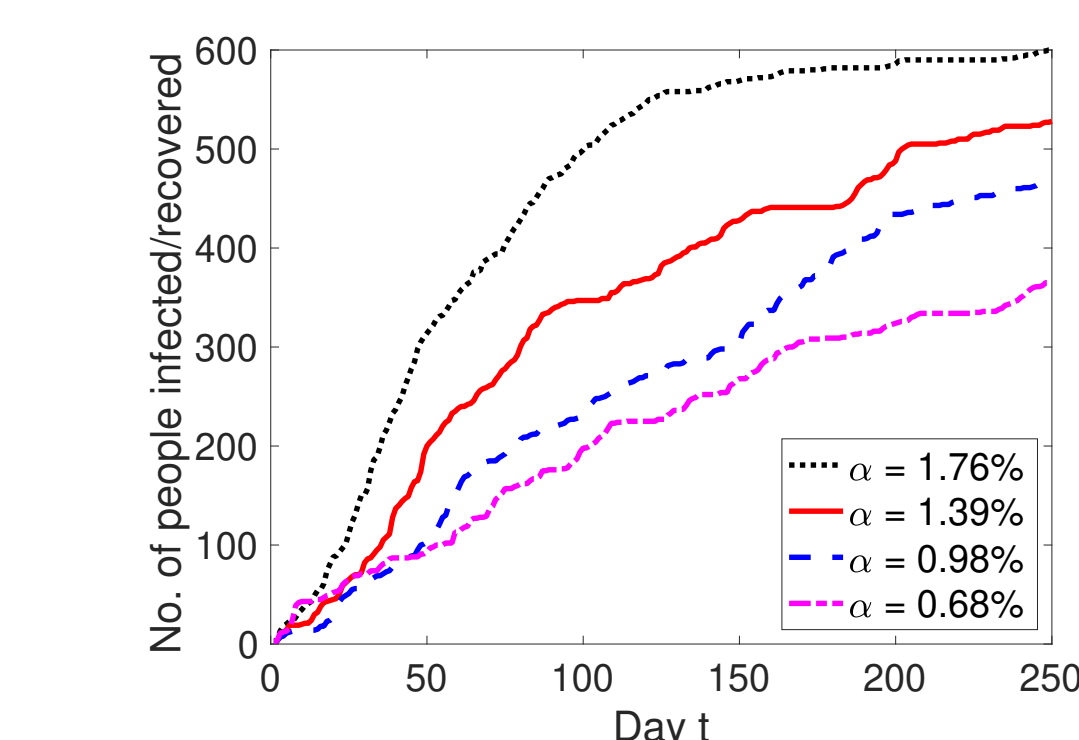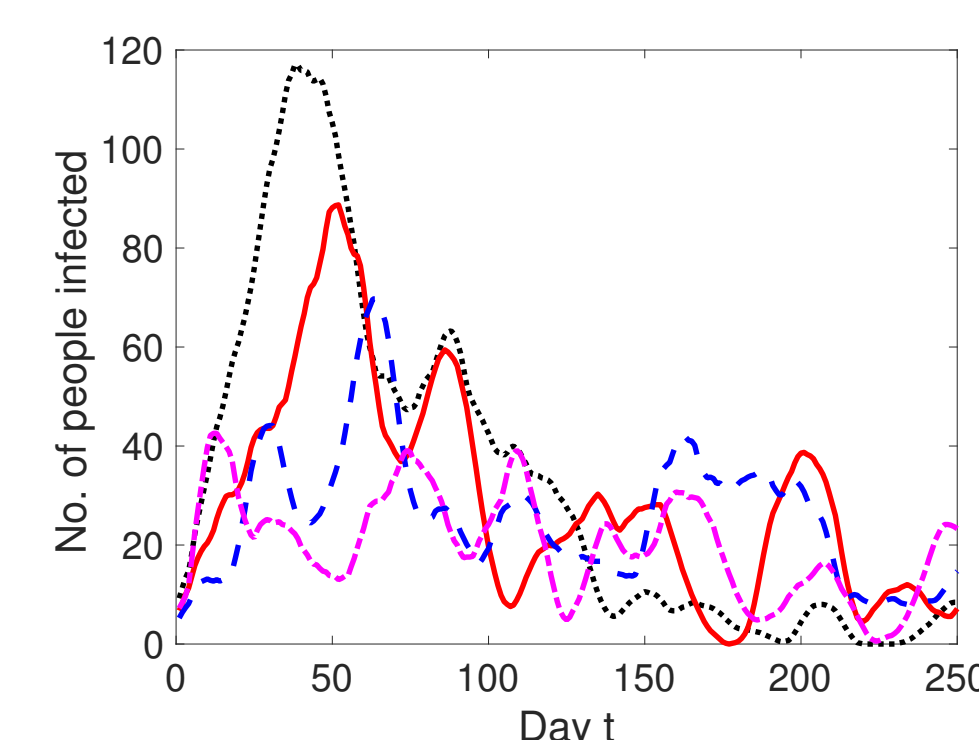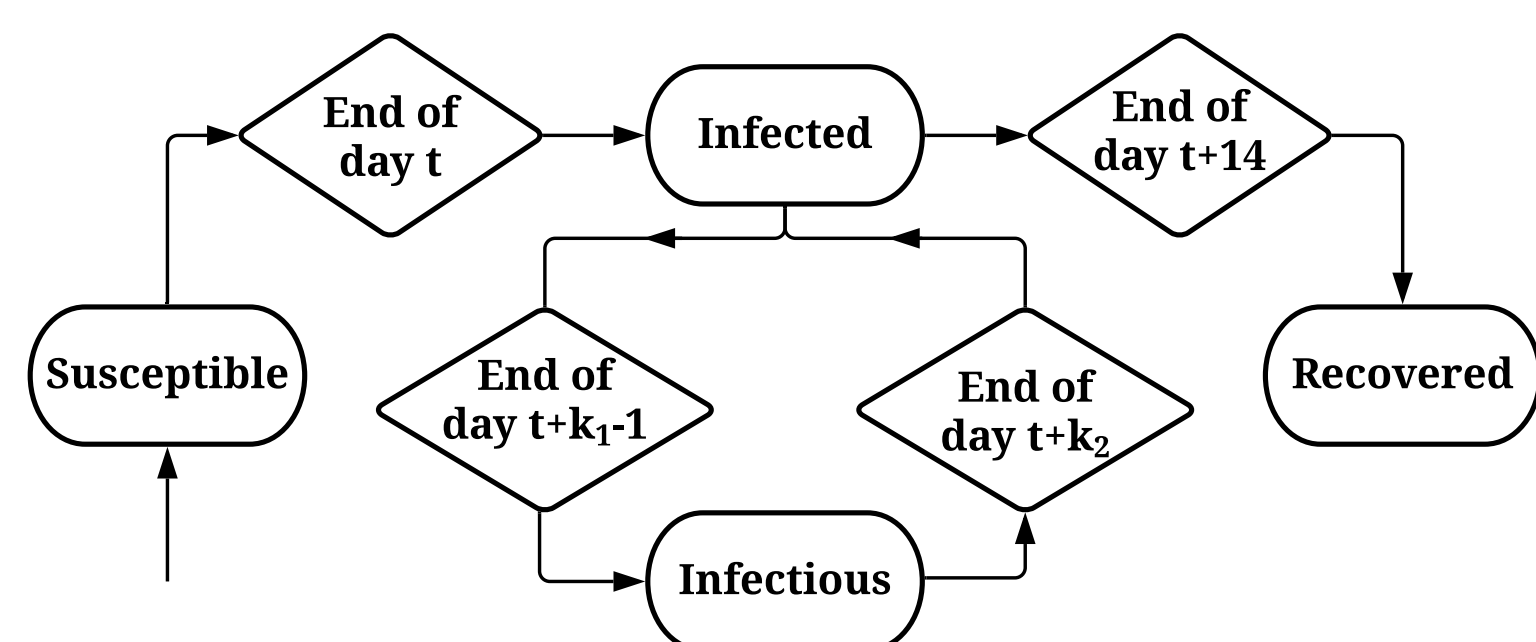
(a) State transition diagram for an individual

(b) Number of active infections

(c) Cumulative number of infections

Proposed Methods

Binary Noise: Model M1

- Pooling: $w = Ax$, where $A \in \{0,1\}^{m \times n}$ and $x \in \{0,1\}^n$.
- Noise: $y_i \in \{0,1\}$ with $\Pr(y_i = 1 | w_i = 0) = 0.1\%$ and $\Pr(y_i = 0 | w_i > 0) = 2\%$.
- Decoding: **Generalized Approximate Message Passing** (GAMP) framework.
  - The key is to design a *denoiser* g_{in} with $\hat{x}_i = g_{\text{in}}(v) = \mathbb{E}[X_i | V = v]$, where v is pseudo data.
  - Design two denoisers – family denoiser and contact tracing denoiser.

Multiplicative Noise: Model M2

- Pooling: $w = Ax$, where $A \in \{0,1\}^{m \times n}$ and $x \in [0,\infty)^n$.
- Noise: $y = w \circ (1+q)^{\mathcal{N}(0,\sigma^2 \mathbb{I})}$, where $q \in (0,1]$ and $\sigma^2 \ll 1$ are known parameters.
- Decoding: Least Absolute Shrinkage and Selection Operator (LASSO) based algorithms.
  - COMP followed by group square-root LASSO for family structure.
  - COMP followed by **overlapping group square-root LASSO** for contact tracing structure.

Experimental Results

- Figure below shows performance of methods M1 and M2 for a population of $n = 1000$ individuals.
- M1: FNR & FPR $< 5\%$ for up to 6% sparsity; contact tracing data helps more than family structure.
- M2: Significant improvement using SI in decoding; allows estimation of individual viral loads.

Sparsity: 2.12%   3.98%   6.01%   8.86%
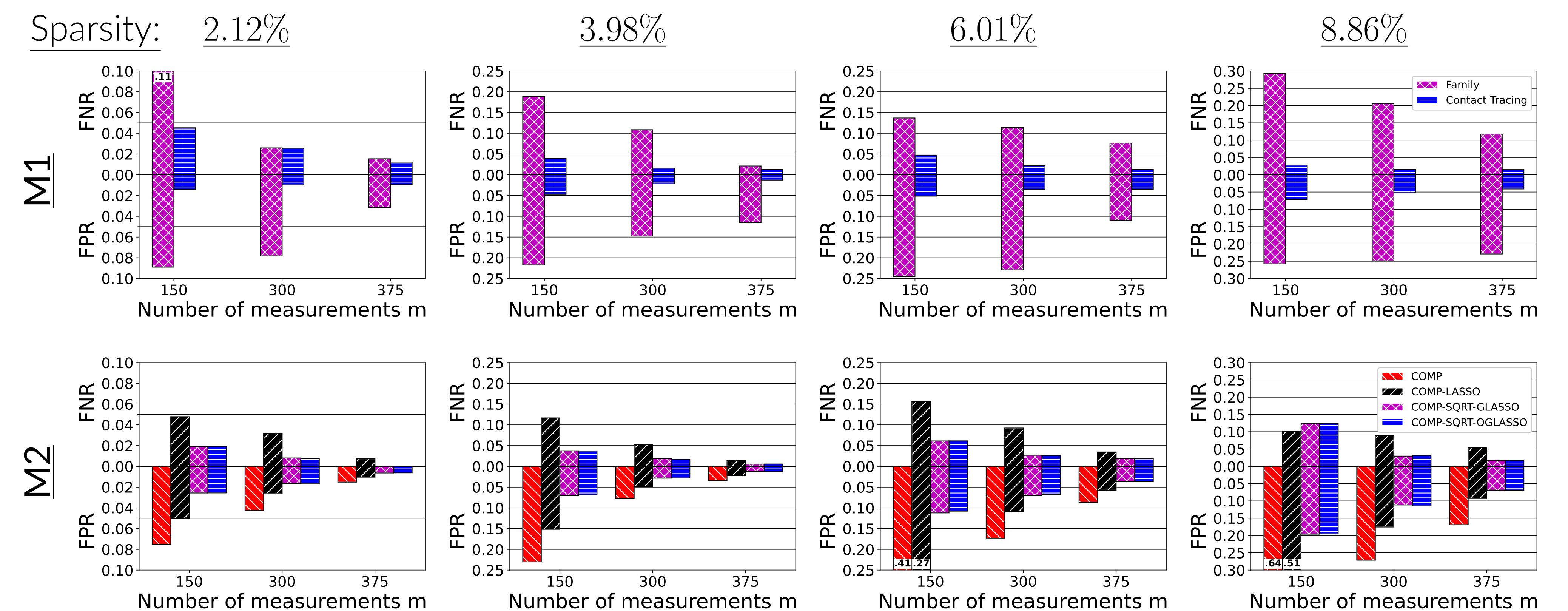
Conclusion and Future Directions

- Improved efficiency by exploiting family structure and contact tracing data – achieved $\sim 5\%$ FNR & FPR at 4% prevalence with need for only 15% of tests required in a conventional testing scenario.
- Could incorporate SI into encoder design and generate better sensing matrices with flexible sizes.