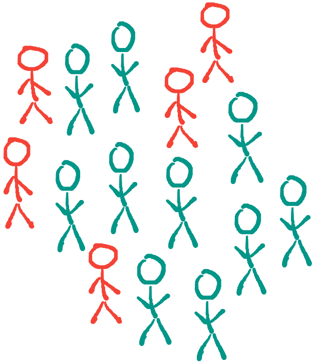
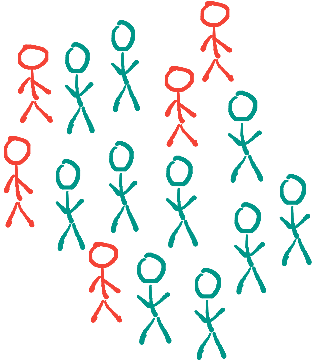
Group Testing in the High Dilution Regime

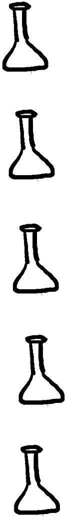
Gabriel Arpino, Nicolò Grometto, Afonso S. Bandeira

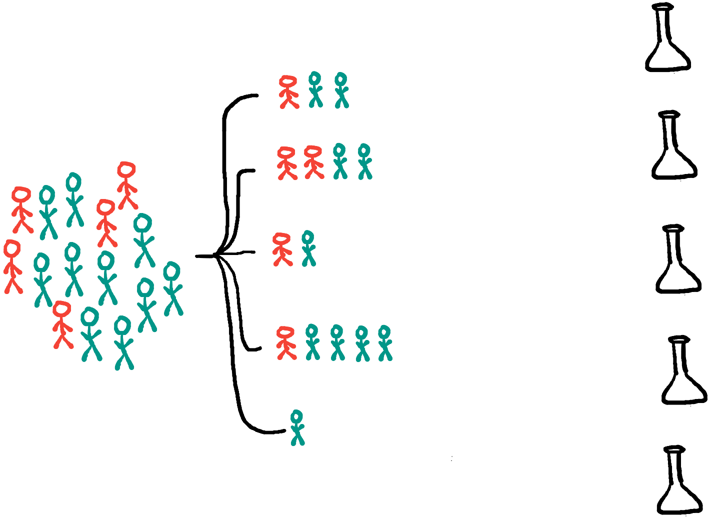
ETH Zürich

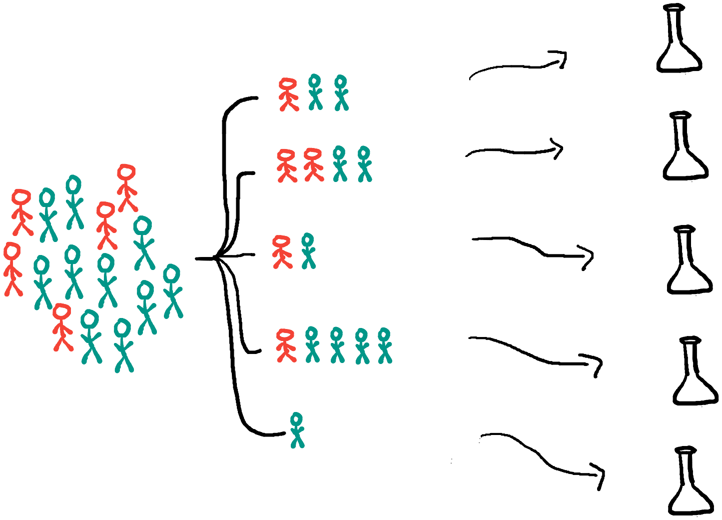
12-20 July 2021

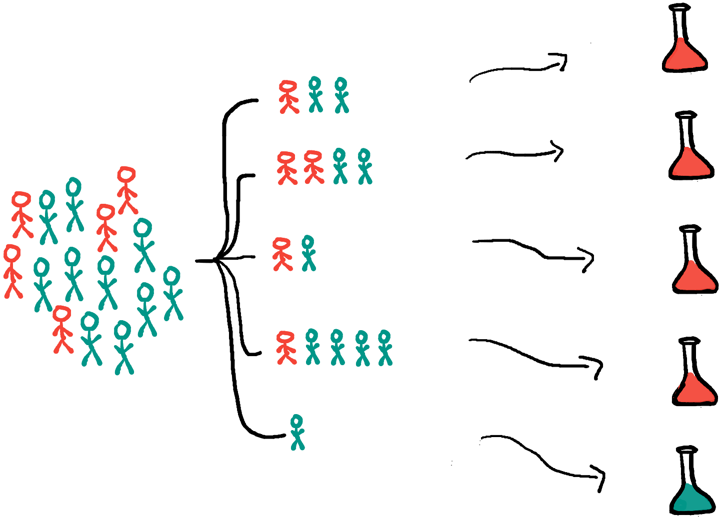








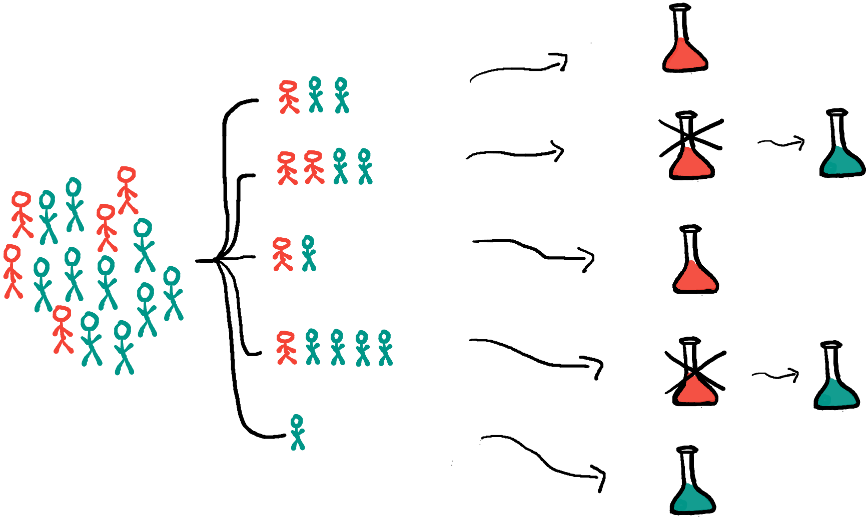


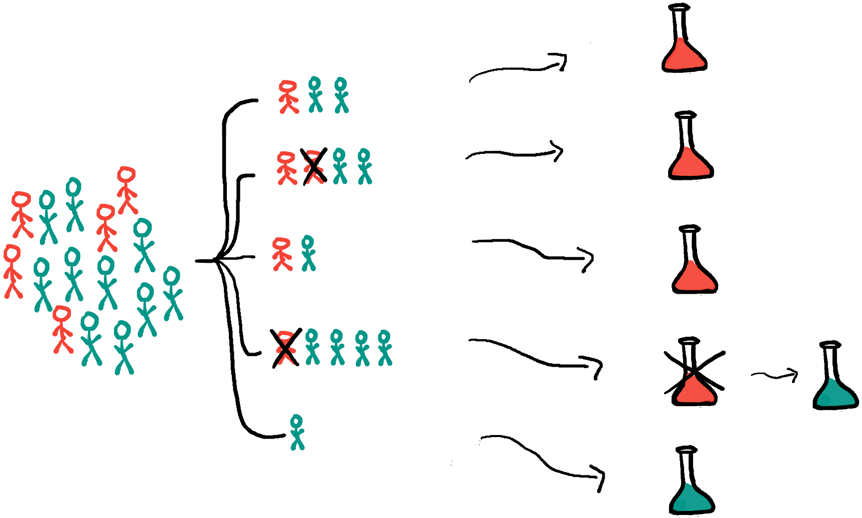


- **Objective**: given a population $[n] := \{1, ..., n\}$ and a subset $\mathcal{D} \subset [n]$ of defectives, we aim to recover $\mathcal{D}$ via the smallest possible number of pooled tests

- **Setting**:

  1. $\forall i \in [n]$: item i appears in any given test independently with probability $p = \frac{\alpha}{d}$
  2. $\frac{\alpha}{d}$ guarantees a constant probability of finding ≥ 1 defective in a given test pool
  3. Very sparse regime: $d = \Theta(n^{\theta})$, $\theta \in (0, 1)$, $n \to \infty$, $d \to \infty$
  4. Test is positive iff ≥ 1 defective is present
  5. Exact recovery





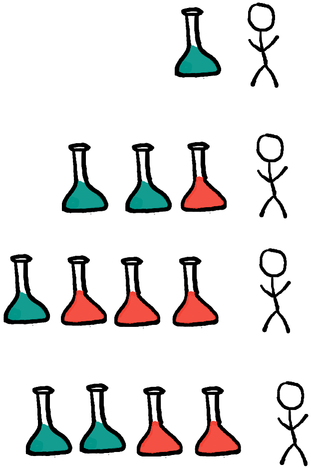
- **Thresholds**: For a decoding algorithm $\mathtt{A}$,
  $\mathsf{N}_{\mathtt{A}} :=$ min. number of tests for exact recovery w.h.p.

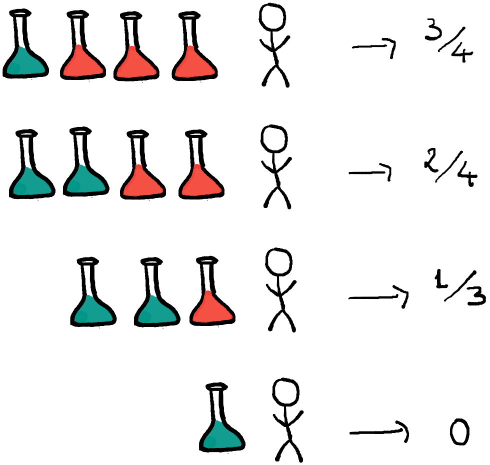
- **Performance measure**: number of bits of information on $\mathcal{D}$ retrieved per test, computed as $\mathsf{R}_{\mathtt{A}} := \frac{\log_2 \binom{n}{d}}{\mathsf{N}_{\mathtt{A}}}$

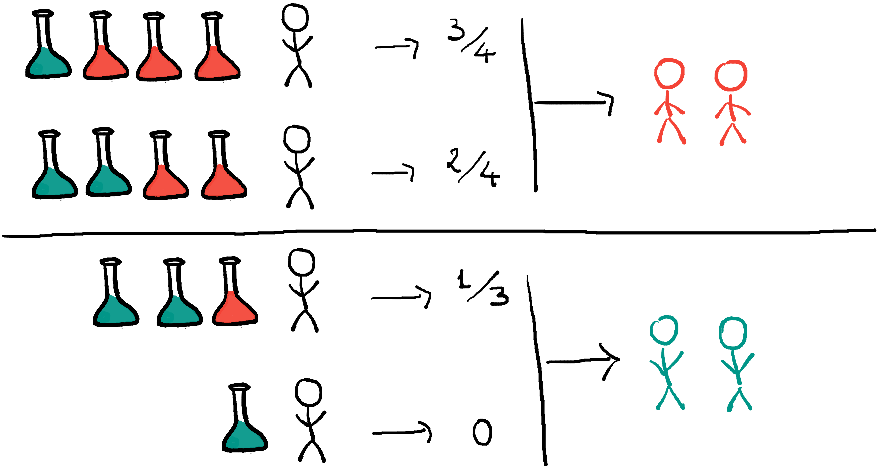
- **Design parameters**:
  1. $q \in (0,1)$ (dilution noise level)
  2. $\alpha = \frac{\log 2}{1-q}$ (optimal in high noise)

- Main results I: Achievable Bounds for the $\mathtt{NCOMP}$ algorithm

- Main results II: Information-theoretic converse bounds




3/4
2/4
1/3
0


3/4
2/4
1/3
0

Theorem I (Achievable bound for NCOMP)

For dilution noise $q \in (0,1)$, $\alpha = \frac{\log 2}{1-q}$,

$$\mathrm{N}_{\mathtt{NCOMP}} \lesssim \frac{d \log n \cdot (1 + o(1))}{(1-q)}$$

1. **Notation**: $\forall i \in [n]$, define
   $\mathcal{P}_i^+ := \#\{\text{positive tests containing i}\}, \mathcal{G}_i := \#\{\text{tests containing i}\}$
2. **Noise-sensitive decoding with slack**: $\forall i \in [n]$: $\texttt{NCOMP}(i) = 1$ if

$$\mathcal{P}_i^+ \geq \mathcal{G}_i \cdot (1 - q \cdot (1 + \Delta))$$

   for a design parameter Δ
3. **Error probabilities**: let $\mathsf{P}_-^{(\text{err})} := \mathbb{P}(\exists i \in [n] \ : \ \texttt{NCOMP}(i) = 0 \mid i \text{ def})$ (similarly for $\mathsf{P}_+^{(\text{err})}$)
4. **U.B. + concentration**: derive sufficient conditions on $\mathsf{N}_{\texttt{NCOMP}}$ s.t. $\lim_{n\to\infty} \mathsf{P}_-^{(\text{err})}, \mathsf{P}_+^{(\text{err})} = 0$
5. **Parameter tuning**: choose optimal Δ to satisfy both conditions, for $\alpha = \frac{\log 2}{1-q}$

Theorem I (Achievable bound for NCOMP)

For dilution noise $q \in (0,1)$, $\alpha = \frac{\log 2}{1-q}$,

$$\mathrm{N}_{\mathtt{NCOMP}} \lesssim \frac{d \log n \cdot (1 + o(1))}{(1-q)}$$

*Previous achievable bounds in dilution noise [2], [3]:

$$\mathrm{N}_{\mathtt{A}} \lesssim \frac{d \log n \cdot (1 + o(1))}{(1-q)^2}$$

Theorem II (Algorithm-independent converse bounds)

For dilution noise $q \in (0,1)$, $\alpha = \frac{\log 2}{1-q}$, *and for any decoder* $\mathtt{A}$,

$$\mathsf{N}_{\mathtt{A}} \gtrsim \frac{\log_2 \binom{n}{d} \cdot (1 + o(1))}{1 - \mathbb{E}_{\mathsf{Z}\sim \mathsf{P}(\alpha)} \left[\mathsf{H}_b \left(q^{\mathsf{Z}}\right)\right]}$$

where $H_b(\rho) = \rho \log_2 \left(\frac{1}{\rho}\right) + (1-\rho) \cdot \log_2 \left(\frac{1}{1-\rho}\right)$ *is the binary entropy function and* $P(\alpha)$ *denotes a Poisson*(α) *distribution.*

1. **Prior results**: from [1], $\forall \mathtt{A}$, with $\lim_{n\to\infty} \mathsf{P}_{\mathtt{A}}^{(\mathrm{err})} = 0$ requires

$$\mathsf{N}_{\mathtt{A}} \geq \frac{\log_2 \binom{n}{d} \cdot (1 + o(1))}{\mathcal{I}(\mathbf{S}; y)}$$

   where **S** denotes the set of items included in a test with outcome y

2. **Computing the mutual information**: for $\alpha = \frac{\log 2}{1-q}$ rewrite as $\mathcal{I}(\mathbf{S}; y) = H(y) - H(y \mid \mathbf{S}, y)$, for

$$\mathsf{H}(y) \xrightarrow{n\to\infty} \mathsf{H}_b\left(e^{-\alpha(1-q)}\right) = 1$$

$$\mathsf{H}(y \mid \mathbf{S}) \xrightarrow{n\to\infty} \mathbb{E}_{\mathsf{Z}\sim\mathsf{P}(\alpha)}\left[\mathsf{H}_b\left(q^{\mathsf{Z}}\right)\right]$$

- **Rate computations**:
  1. Achievable rates: $R_{\texttt{NCOMP}}(q) \sim (1-\theta)\cdot(1-q)$
  2. Converse rates: $R_{\texttt{CONV}}(q) \sim 1 - \mathbb{E}_{Z\sim P(\alpha)}\left[H_b\left(q^Z\right)\right]$

- **Optimality in high noise**: for $\alpha = \frac{\log 2}{1-q}$

$$\lim_{q\to 1} R_{\texttt{CONV}}(q) \in \mathcal{O}(1-q)$$
$$\lim_{q\to 1} R_{\texttt{NCOMP}}(q) \in \Omega(1-q)$$

- **Intuition**: increasing the number of items in a test according to α offsets the noise, with optimal scaling $(1-q)^{-1}$

- **Main results**: in the high noise regime,

  1. Matching achievable and converse rates up to order
  2. Optimal scaling for the average test sparsity α w.r.t. the noise level

- **Future directions**:

  1. Fully characterising the information-theoretic consequences of dilution noise
  2. Thought experiment: consider the case in which a test is repeated multiple times in dilution noise

Thank you!

References

Matthew Aldridge, Oliver Johnson, Jonathan Scarlett, et al. "Group testing: an information theory perspective". In: *Foundations and Trends® in Communications and Information Theory* 15.3-4 (2019), pp. 196–392.

George K Atia and Venkatesh Saligrama. "Boolean compressed sensing and noisy group testing". In: *IEEE Transactions on Information Theory* 58.3 (2012), pp. 1880–1901.

Chun Lam Chan et al. "Non-adaptive probabilistic group testing with noisy measurements: Near-optimal bounds with efficient algorithms". In: *2011 49th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2011, pp. 1832–1839.