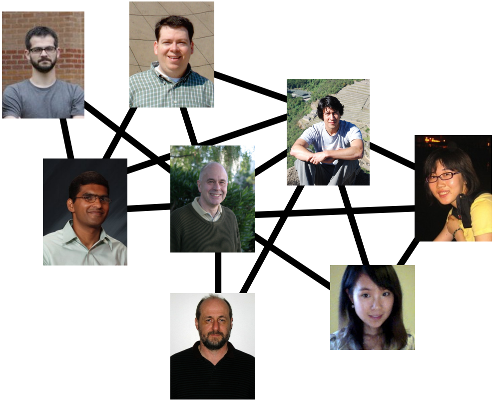
# Compressive Demodulation of Mutually Interfering Signals

Lorne Applebaum
Waheed U. Bajwa
Robert Calderbank
Yuejie Chi
Stephen Howard
Matthew Nokleby
Yao Xie



# Interference in Multiuser Communication

Why manage interference when one can avoid it?

- User Scheduling: Token ring, etc.
- Media Access Control: ALOHA, CSMA, etc.
- Interference Alignment [Cadambe and Jafar, 2007]

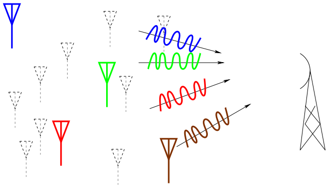
Avoiding interference might be too costly

- Cost of multiuser synchronization
- Cost of sharing channel state information
- (Hardware) cost of full duplex channels

In low-rate random access channels, it's better to manage interference

- Multiuser wireless control channels
- Ad-hoc networks neighborhood discovery

# The Random Access Channel



- Wireless control channels, presence indication
- Ad-hoc network neighborhood discovery
- Sensor network reporting

**Characteristic**

- Only a small fraction of users $\delta \ll 1$ are simultaneously active

**Goal**

- Detect active users
- Support as many total users and active users in a small signal space

# Summary of Past Results

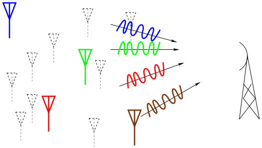
For signal dimension $M$ want to support largest total users $N$ and active users $k = \delta N$ possible

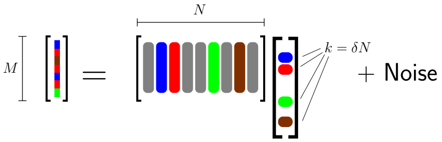
**Classical Approach**: Orthogonal signaling

$$N \lesssim M \quad k \lesssim M$$

**Recent Approaches** (Fletcher, Rangan, Goyal [2009], Guo, Luo [2010])

- Transform into sparse recovery



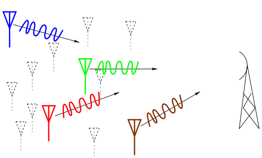


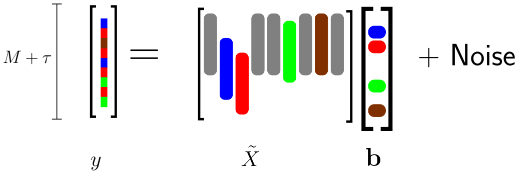

$$k \lesssim \frac{M}{\log\big(N(1-\delta)\big)}$$

- Requires perfect synchronization
- Asymptotic results in $N, M, K$, SNR
- Random codebooks

# Practical Considerations

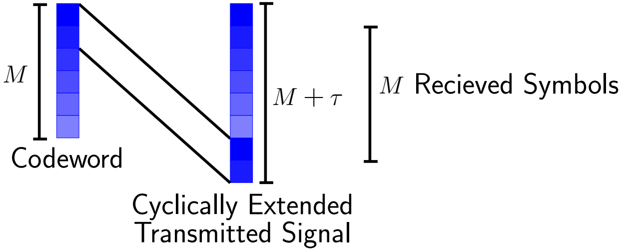
- Perfect synchronization or knowledge of individual delays extremely difficult to attain
  - Can know maximum delay $\tau$



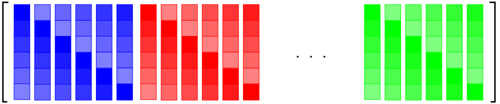



- System wide channel gains cannot be known
- Non-asymptotic theory desired for system design
- Metrics desired for codebook design

# Asynchronous Reception




Receiver matrix $\boldsymbol{A}$:



Users divided in cyclic equivalence classes

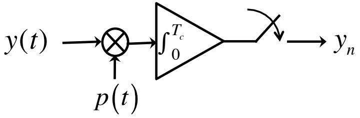
# High-rate Sampling Architecture

- Sampling pulse $p(t)$, with duration much shorter than $T_c$
- Sampling rate on the order of chip rate
- Output model

$$\boldsymbol{y} = \mathbf{H}\boldsymbol{A}\boldsymbol{R}\mathrm{b} + \boldsymbol{w}$$

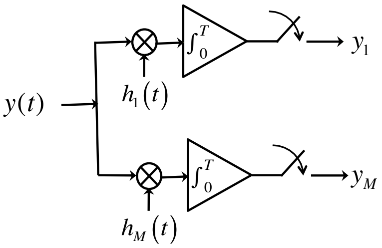
where $\mathbf{H} = \mathbf{I}_{\Omega}$,

$$\boldsymbol{w} \sim \mathcal{N}(0, \sigma^2),$$



# Mixing Signal Architecture

- Decorrelating by a bank of mixing filters
- Each multiple signal by $h_m(t)$, $m = 1, \cdots, M$
- Conventional matched-filter front-end employs user signature waveforms
  - Synchronous, $M = N$, $N = \#$ of users, $h_m(t) = s_m(t)$
  - Asynchronous, RAKE receiver, $M = N(\tau + 1)$, $\tau =$ maximum discrete delay, $h_m(t) = s_n(t - l)$, $1 \leq n \leq N$, $0 \leq l \leq \tau$
- Sampling rate $\sim 1/T$, $T$ symbol period



- Measurement

$$y_m = \int_0^{T-\tau} h_m(t) y(t+\tau) dt, \quad m = 1, \cdots, M.$$

- In general, asynchronous MUD, choice of filter signals

$$h_m(t) = \sum_{l=0}^{L-\tau-1} h_{m,l} g(t - lT_c),$$

- Define

$$\boldsymbol{h}_m = [h_{m,0} \quad \cdots \quad h_{m,L-\tau-1}]^\top, \quad m = 1, \cdots, M,$$

- Output signal model

$$y = HAR\mathrm{b} + w$$

where

$$w \sim \mathcal{N}(0, \sigma^2 HH^\top),$$

- Special choices
  - $H = I_\Omega$, mathematically, the mixing architecture and the subsampling architectures are equivalent
  - $H$ is a tight-frame, $HH^\top = I$, output noise is white

# Comparison of Two Architectures

| Architecture | Subsampling | Mixing |
|---|---|---|
| Support # of Users | $\sim N$ | $\sim N$ |
| # of Filters | 1 | $M \ (M \ll N)$ |
| # of Samples per Symbol | $M \ (M \ll N)$ | 1 |
| Sampling Rate | $\frac{N}{M} \cdot \frac{1}{T_c}$ | $1/T \ (T \gg T_c)$ |
| Signal Power | $M/N$ | 1 |
| Noise Power (tight-frame) | $\sigma^2$ | $(N/M)\sigma^2$ |
| Signal-to-Noise Ratio | $M/N$ | $M/N$ |

# Detection Algorithm: Iterative Matching Pursuit

- Input: matrices $\hat{\mathbf{H}} = \mathbf{H}\boldsymbol{A}$, signal vector $\boldsymbol{y}$, number of active users $K$
- Output: active user set $\mathcal{I}$, transmitted symbols $b_n$, $n \in \mathcal{I}$
- Iteratively detect and subtract one multi-path component, repeat $K$ times
  - Compute: $\boldsymbol{f} := \hat{\mathbf{H}}^H \boldsymbol{v}_j$
  - $i = \arg\max_{n \in \mathcal{H}_j} |f_n|$
  - Detect active users: $\mathcal{I}_{j+1} = \mathcal{I}_j \cup \{\lceil i/(\tau+1) \rceil\}$
  - Detect symbol $[\hat{\mathbf{b}}_{j+1}]_i = \mathsf{sgn}(r_i f_i)$.
  - Update residual: $\boldsymbol{v}_{j+1} = \boldsymbol{b}_j - \hat{\mathbf{H}}\boldsymbol{R}\mathbf{b}_{j+1}$
  - Update detection set: $\mathcal{H}_{j+1} = \mathcal{H}_j \backslash \{\lfloor i/(\tau+1) \rfloor (\tau+1)+1, \cdots, \lceil i/(\tau+1) \rceil (\tau+1)\}$

# Performance Guarantee

- Performance guarantee stated in terms of worst-case and average coherence property of $\hat{\boldsymbol{H}}$
- *Worse-case coherence* $\mu(\hat{\boldsymbol{H}}) \triangleq \max_{n \neq m} |\hat{\mathbf{h}}_n^H \hat{\mathbf{h}}_m|$
- *Average coherence* $\nu(\hat{\boldsymbol{H}}) \triangleq \frac{1}{N(\tau+1)-1} \max_n \left| \sum_{m \neq n} \hat{\mathbf{h}}_n^H \hat{\mathbf{h}}_m \right|$
- *Coherence property*

$$\mu(\hat{\mathbf{H}}) \leq 0.1/\sqrt{2 \log N(\tau+1)}, \quad \nu(\hat{\mathbf{H}}) \leq \mu/\sqrt{M}. \qquad (1)$$

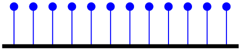
- Performance also depends on
  - $n$-th largest signal-to-noise ratio

$$\mathrm{SNR}_n = \frac{|r|^2_{(n)}}{\mathbb{E}\{\|\boldsymbol{w}\|_2^2\}/K}, \tag{2}$$

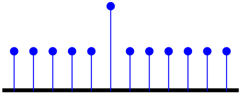
  - and $n$-th largest user relative to the rest of the weaker users

$$\mathrm{LAR}_n = \frac{|r|^2_{(n)}}{\|\boldsymbol{r}_{\mathcal{I}_n}\|_2^2/K}, \tag{3}$$

  where $\|\boldsymbol{r}_{\mathcal{I}_n}\|_2^2 = \sum_{i=n}^{k} |r|^2_{(i)}/K$.



Low LAR



High LAR

# Performance guarantees

## Theorem

*Suppose that the matrix $\hat{\boldsymbol{H}}$ satisfies the coherence property and $\boldsymbol{H}$ is a tight-frame. Let the noise $\boldsymbol{w}$ be distributed as $\mathcal{N}(\mathbf{0}, \sigma^2 \boldsymbol{I}_{M\times M})$. We write $\mu(\hat{\mathbf{H}})$ as $\mu = c_1 M^{-1/\gamma}$ for some $c_1 > 0$ ($c_1$ may depend on $N(\tau+1)$ and $\gamma \in \{0\} \cup [2, \infty)$). Then the detection algorithm satisfies $\mathbb{P}\{\hat{\mathbf{b}} \neq \mathbf{b}\} \leq 6N^{-1}(\tau+1)^{-1}$ as long as $N(\tau+1) \geq 128$ and the number of measurements $M$ satisfies*

$$M > \max\left\{2K \log N(\tau+1), \frac{c_2 K \log(N(N-n)(\tau+1)^2)}{\mathrm{SNR}_n}, \left(\frac{c_3 K \log(N(\tau+1))}{\mathrm{LAR}_n}\right)^{\gamma/2}\right\}_{n=1}^{K},$$

*where $c_2 = 8(1-t)^{-2}$ and $c_3 = 800 c_1^2 t^{-2}$.*

# Random Subsampling for Gabor Frames

Subsampling from random locations in $\Omega$ in Gabor Frames:

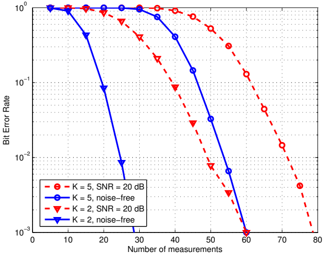
$$\mathcal{G} = I_\Omega \cdot [W_0 T(g),\ W_1 T(g),\ \ldots,\ W_{n-1} T(g)]$$

where $|\Omega| = M$.

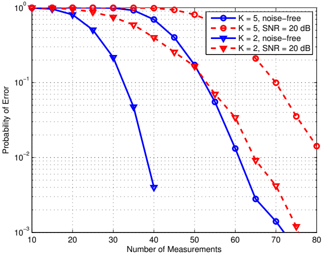
Coherence Properties:

- The worst case coherence $\mu_\Omega \leq \mu + 4\sqrt{\frac{\log N + \epsilon}{M}}$.
- The average case coherence $\nu_\Omega = \frac{N^2+M}{N^2-1} \cdot \frac{1}{M} \approx \frac{1}{M}$.

# Numerical Examples



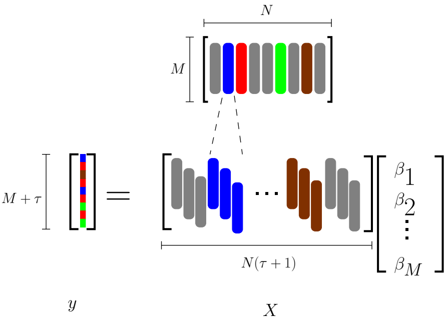

(a) Alltop Gabor Frame




(b) Random Gabor Frame

Figure: Multi-user detection error rate with respect to the number of measurements using a Gabor frame generated from (a) an Alltop sequence with length $P = 127$, and (b) a random uniform vector with length $P = 128$ for different active users and SNR, where the maximum chip delay is $P$.

# Detection Algorithm: Lasso




**Algorithm: MUD via Lasso**

**Inputs**

1. The chip-rate sampled vector $\mathbf{y}$
2. Set of $M$-dimensional codewords $\{\mathbf{x}_i\}_{i=1}^{N}$
3. Maximum discrete delay $\tau$ in the system
4. A regularization parameter $\lambda$ for the lasso

**Compute** the matrix $\mathbf{X}$ from the codewords $\{\mathbf{x}_i\}$

$$\widehat{\boldsymbol{\beta}} \leftarrow \underset{\mathbf{b}\in^{N(\tau+1)}}{\arg\min} \ \frac{1}{2}\|\mathbf{y}-\mathbf{X}\mathbf{b}\|_2^2 + \lambda\|\mathbf{b}\|_1$$

$$\widehat{\mathcal{I}} \leftarrow \{i : \|\widehat{\boldsymbol{\beta}}_i\|_0 > 0\}$$

**Return** $\widehat{\mathcal{I}}$ as an estimate of the set of active users $\mathcal{I}$

System Assumptions:

- AWGN at receiver distributed as $\mathcal{N}(\mathbf{0}_{M+\tau}, \mathbf{I}_{M+\tau})$
- Active users can estimate and invert channel coefficient magnitudes

# User Detection from Asynchronous Codes

## Theorem - Arbitrary Delays (ABDC [2010])

Assuming,

- User activity random Bernoulli with probability $\delta$
- User delays are arbitrary
- Minimum received SNR $|\beta_i|^2 > 128 \log\left(N\sqrt{\tau+1}\,\right)$

then the MUD via Lasso algorithm is successful with high probability $O(N^{-1})$ when

$$N \leq \frac{\exp((c_1(\tau+1)\mu(\mathbf{X}))^{-1})}{\tau+1} \quad \text{and} \quad \delta \leq \frac{1}{\|\mathbf{X}\|_2^2 \log\big(N(\tau+1)\big)}$$

**Important Metrics**

- $\mu(\mathbf{X}) \overset{def}{=} \max_{(i,j)\neq(i',j')} \left|\langle \mathbf{x}_{i,j}, \mathbf{x}_{i',j'} \rangle\right|$
- $\|\mathbf{X}\|_2$

# User Detection from Asynchronous Codes

**Theorem - Random Delays (ABDC [2011])**

Assuming,

- User activity random Bernoulli with probability $\delta$
- User delays are uniformly distributed in $[0, \tau]$
- Minimum received SNR $|\beta_i|^2 > 128 \log\left(N\sqrt{\tau+1}\,\right)$

then the MUD via Lasso algorithm is successful with high probability $O(N^{-1})$ when

$$N \leq \frac{\exp((c_1(\tau+1)\mu(\mathbf{X}))^{-1})}{\tau+1} \quad \text{and} \quad \delta \leq \frac{\tau+1}{\|\mathbf{X}\|_2^2 \log\big(N(\tau+1)\big)}$$

- Allowable activation rate increases by factor of $\tau + 1$
- Probability measure now over random user activation and delays

# Implications

Assume:

- $\mu(\mathbf{X}) \approx \frac{1}{\sqrt{M}}$ (Welch-like)
- $\|\mathbf{X}\|_2 \approx \sqrt{\frac{N(\tau+1)}{M}}$ (Tight frame)

| Classical | Proposed |
|---|---|
| $N \lesssim M$ | $N \lesssim e^{M^{1/3}}$ |
| $k \lesssim M$ | $k \lesssim \frac{M}{\log(N)}$ |

**Important Differences from Earlier Work**

- Asynchronously received columns. True for all choices $\tau_i \leq \tau$
- Non-asymptotic in problem dimensions or SNR
- Stochastic active user set, suitable framework for deterministic matrix construction

# Proof Sketch

Model Selection with Lasso (Candès & Plan [2009]):

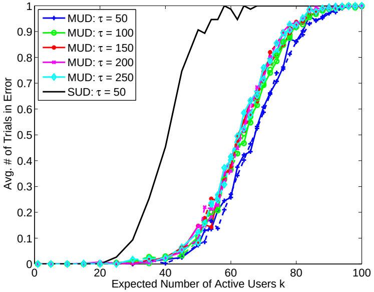
- Five conditions on Lasso solution derived from sub-gradient requirement

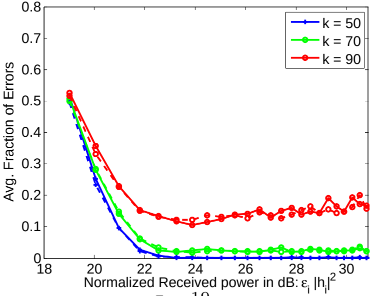
$$\text{eg:} \quad \|(\mathbf{X}_{\mathcal{S}}^{\mathrm{H}}\mathbf{X}_{\mathcal{S}})^{-1}\|_2 \leq 2$$

- Exploiting the structure of $\mathbf{X}$, we prove each occurs with high probability
- Structured matrix generalization of random matrix sub-selection moment calculation (BCD[2010],Tropp [2008])

# Numerical Results

$$N = 3072, M = 1023$$




$\tau = 19$

- SNR=31dB prescribed in theory.
- Worst case analysis on shifts in theory. Average in simulation.
- In average case: recovery not strongly dependent on $\tau$
- Theory overly restrictive in power: $\epsilon_i |h_i|^2 \geq 31\text{dB}$
- Theory correct in predicting required power not a function of $k$

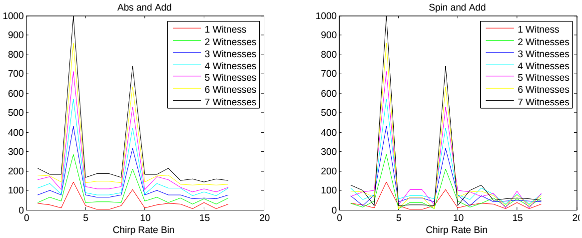
# Managing Complexity: Chirp Signaling for MUD

Columns of $\boldsymbol{A}$:

$$\psi_{x,y}(t) = e^{\frac{2\pi i x t}{M}} e^{\frac{\pi i y t(t-M)}{M}}$$

Properties:

- Cyclic equivalent classes: $\psi_{x,y}(t+s) = \psi_{x,y}(s)\psi_{x+sy,y}(t)$
- "Wiggling equivalent" to Alltop Gabor frames (column permutation + phase)
- When $\boldsymbol{H} = \mathbf{I}$ fast chirp reconstruction algorithm.



# Subsampling From a Subspace for Kerdock Codes

Subsampling from a subspace $\mathcal{S}$ in $\mathbb{F}_2^m$ in Kerdock Codes:

$$\psi_{P,a}^{\mathcal{S}} = \frac{1}{\sqrt{M}} i^{xPx^T + 2ax^T}, \quad x \in \mathcal{S}$$

where $|\mathcal{S}| = M$.

Coherence Properties:

- The worst case coherence $\mu_{\mathcal{S}} = \frac{1}{\sqrt{M}}$.
- The average case coherence $\nu_{\mathcal{S}} = \frac{N^2 - M}{(N^2-1)M} \approx \frac{1}{M}$.

# Fundamental Limits

Precoder Design in MIMO Gaussian Channels:

$$y = \sqrt{\mathsf{snr}} H P x + w$$

How to design $P$ to maximize the mutual information?

$$\begin{aligned} \text{max} \quad & I(b; y = \sqrt{\mathsf{snr}} H P x + w) \\ \text{subject to} \quad & \mathrm{Tr}(P P^\dagger) \le 1 \end{aligned}$$

The solution is given by

$$P^* = \gamma^{-1} H^\dagger H P^* E,$$

where $E$ is the mmse matrix of $x$, and $\gamma = \|H^\dagger H P^* E\|$.

Question: Can we leverage mutual information for MUD?

# Summary

In many applications, it is better to manage interference

- Two code flavors:
  - random
  - deterministic (why take a chance?)
- Codebook metrics to evaluate potential code design
- Multiple widely-applicable recovery methods (Lasso, OST)
- Obtained non-asymptotic performance guarantees
- Fairness between users is guaranteed
  - Near-far problem mitigated in MUD
  - Receivers solve equivalent recovery problems in ad-hoc network
- Joint detection provides favorable system parameter scaling