Compressive Sampling of Ensembles of Correlated Signals

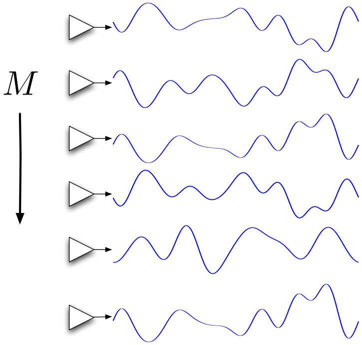
Justin Romberg

Collaborator: Ali Ahmed

Georgia Tech, School of ECE

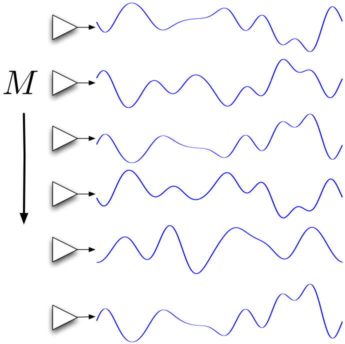
Duke Workshop on Sensing and Analysis of High-Dimensional Data
July 26, 2011
Durham, North Carolina

Sampling correlated signals

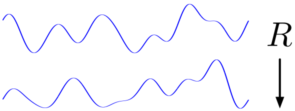


- Goal: acquire an *ensemble* of M signals
- Bandlimited to $W/2$
- "Correlated" $\rightarrow$ M signals are $\approx$ linear combinations of R signals

Sampling correlated signals

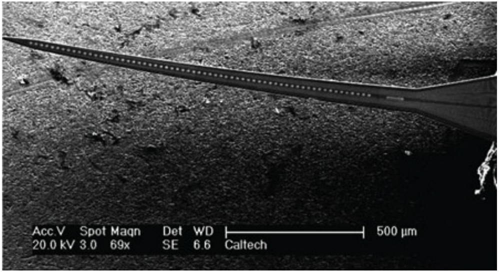


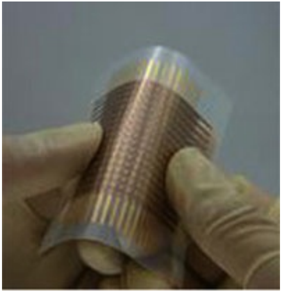
$$= \begin{bmatrix} -0.82 & -1.31 \\ 1.09 & 0.27 \\ 1.05 & 1.81 \\ -0.74 & -0.31 \\ -0.97 & 0.94 \\ 1.19 & 2.19 \end{bmatrix}$$

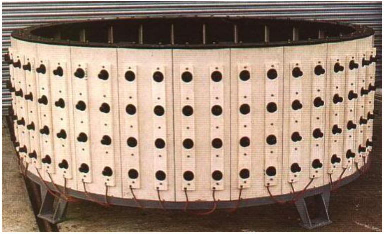


- Goal: acquire an *ensemble* of M signals
- Bandlimited to $W/2$
- "Correlated" $\rightarrow$ M signals are $\approx$ linear combinations of R signals

Sensor arrays



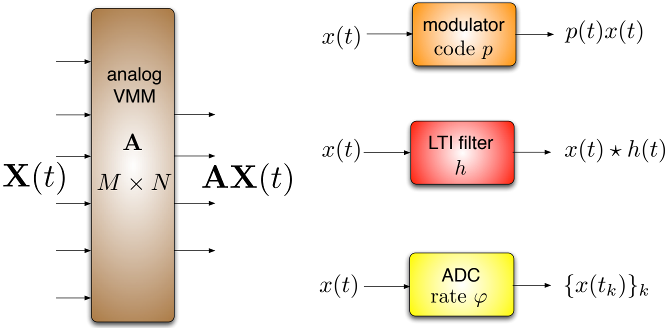




Framework

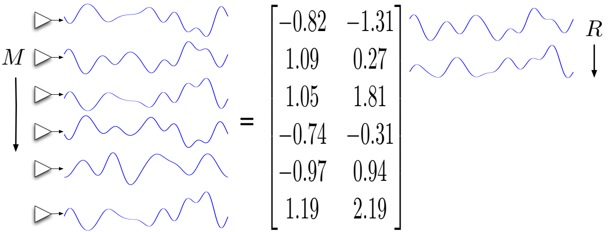
- “Wired” local arrays that may or may not share a (multiplexed) ADC
- Sparsity has nothing to do with it (but makes a guest appearance...)
- Correlation structure is unknown (low-rank recovery problem)
- Interested in systems with clear “implementation potential”

Components



- Analog vector-matrix multiplier spreads energy across channels
- Modulators spread energy across frequency
- Filters spread energy in one channel across time
- We will use both uniform and non-uniform ADCs

Known correlation structure $\rightarrow$ whiten then sample

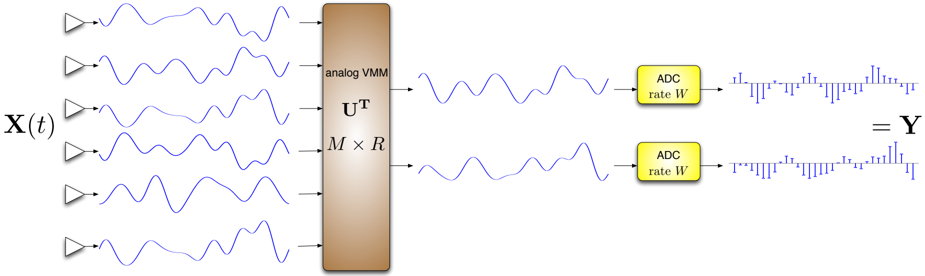


- Suppose the "mixing matrix" $\mathbf{A}$ is known and has SVD

$$\mathbf{A} = \begin{bmatrix} \mathbf{U} \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma} \end{bmatrix} \begin{bmatrix} \mathbf{V}^{\mathbf{T}} \end{bmatrix}$$

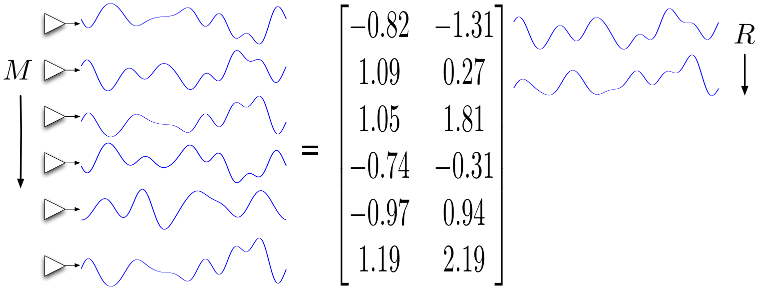
then an efficient sampling structure is to "whiten" with $\mathbf{U}^{\mathbf{T}}$, then sample

Known correlation structure $\to$ whiten then sample

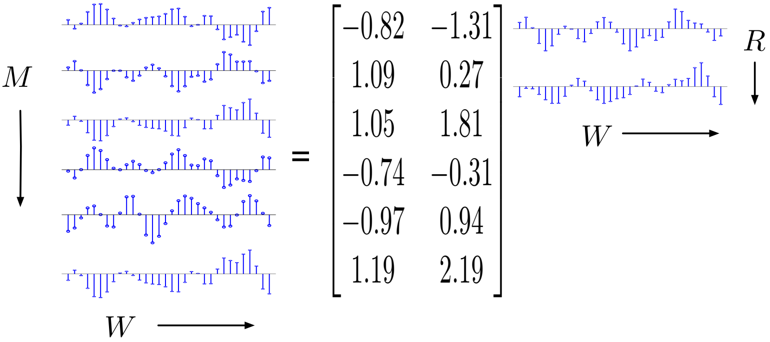


- Requires R ADCs and a total of RW samples
- Recover samples of original using $\mathbf{X} = \mathbf{U}\mathbf{Y}$

Sampling correlated signals



Sampling correlated signals



- Bandlimited $\Rightarrow$ this is just a *low-rank recovery problem*
- Sampling each channel separately takes MW total samples, we want strategies that take $\sim RW$ total samples

Low-rank matrix recovery

- Given p *linear samples* of a matrix,

$$\mathbf{y} = \mathcal{A}(\mathbf{X_0}), \quad \mathbf{y} \in \mathbb{R}^p, \quad \mathbf{X_0} \in \mathbb{R}^{M \times W}$$

  we solve

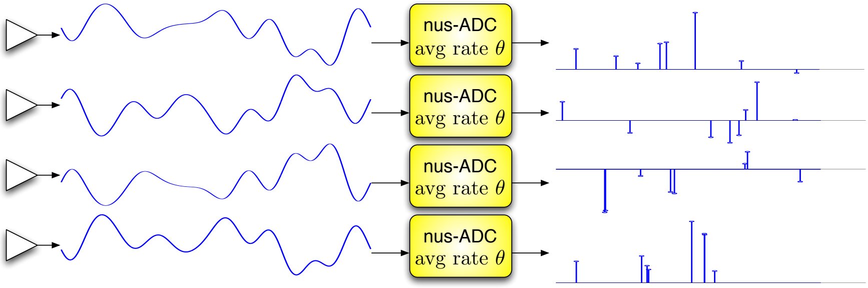
$$\min_{\mathbf{X}} \; \|\mathbf{X}\|_* \quad \text{subject to} \quad \mathcal{A}(\mathbf{X}) = \mathbf{y}$$

  where $\|\mathbf{X}\|_*$ is the nuclear norm: the sum of the singular values of $\mathbf{X}$.

- An "optimal" sampler $\mathcal{A}$ would (stably) recover $\mathbf{X_0}$ from $\mathbf{y}$ when

$$\begin{aligned} \#\text{samples} &\gtrsim \; R \cdot \max(M, W) \\ &\gtrsim RW \quad \text{(in our case)} \end{aligned}$$

Architecture 1: One non-uniform ADC per channel



- M individual nonuniform-ADCs with average rate θ
- Same as choosing $M\theta$ random samples from $M \times W$ matrix

Matrix completion

- Results of Candes, Recht, Tao, Keshavan, Montenari, Oh, Plan, ... $\Rightarrow$
  Given a small number of entries in a low-rank matrix,
  we can “fill in” the missing entries

Matrix completion

- Results of Candes, Recht, Tao, Keshavan, Montenari, Oh, Plan, ... $\Rightarrow$
  Given a small number of entries in a low-rank matrix,
  we can "fill in" the missing entries

- Recht '09: Suppose $M \times W$ matrix $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathbf{T}}$ is rank R with

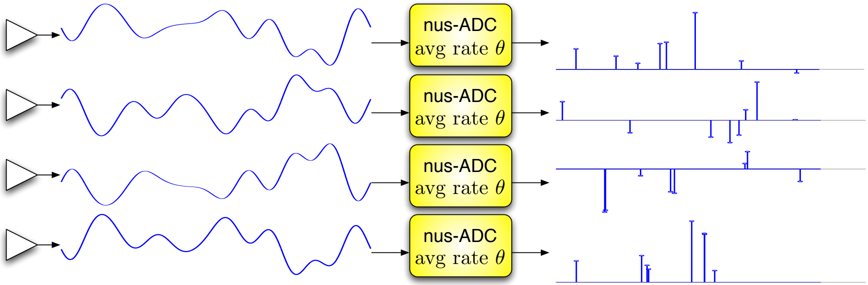
$$\mu = \max\left(\frac{M}{R}\max_i \|\mathbf{U}^{\mathbf{T}}\mathbf{e_i}\|_2^2,\ \frac{W}{R}\max_i \|\mathbf{V}^{\mathbf{T}}\mathbf{e_i}\|_2^2,\ \frac{MW}{R}\|\mathbf{U}\mathbf{V}^{\mathbf{T}}\|_\infty^2\right)$$

  then we can recover $\mathbf{X}$ whp from randomly chosen samples when

$$\#\text{samples} \;\geq\; \text{Const}\cdot\mu\cdot RW\log^2(W)$$

  using nuclear norm minimization

Architecture 1: One non-uniform ADC per channel



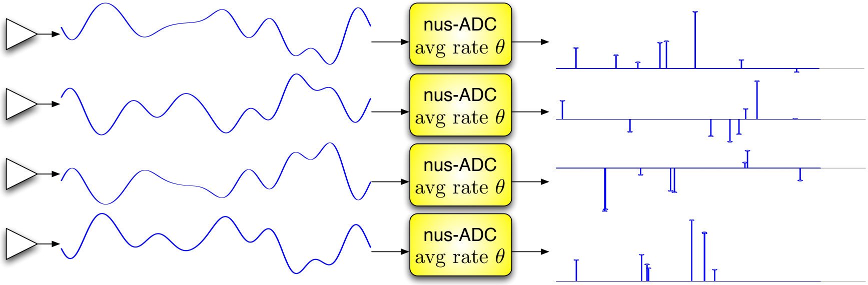
- Direct application of these results: we can recover "incoherent" ensembles when

$$\text{total samples} = M\theta \geq \text{Const.} \cdot RW \cdot \log^2(W)$$

so *we can take $\theta \sim \frac{R}{M}W$ instead of W*.

- Incoherent $\Rightarrow$
signal energy is spread out evenly across time and channels

Architecture 1: One non-uniform ADC per channel

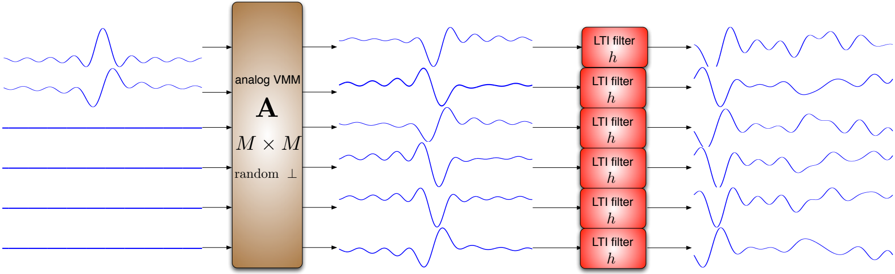


Drawbacks:

- Incoherence assumptions (not universal)
- Requires M ADCs (time-multiplexing would be delicate...)

Spreading the signals out

$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\mathbf{T}$ $\qquad\qquad$ $\tilde{\mathbf{X}} = \tilde{\mathbf{U}}\boldsymbol{\Sigma}\tilde{\mathbf{V}}^\mathbf{T}$



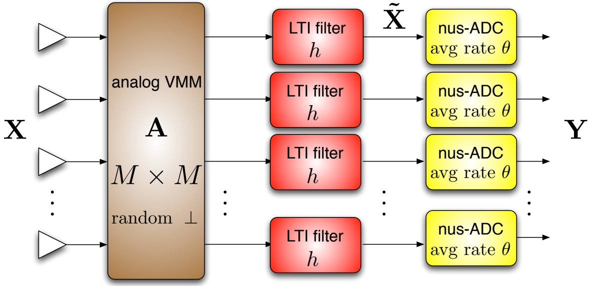
- Take $\mathbf{A}$ $M \times M$ and orthogonal,
  $\mathbf{H} = \mathrm{circ}(h[n])$ orthogonal:

$$\mathbf{H} = \mathbf{F}^\mathbf{H}\boldsymbol{\Lambda}\mathbf{F}, \quad \boldsymbol{\Lambda} = \mathrm{diag}(\{\lambda_i\}), \quad |\lambda_i| = 1$$

then

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\mathbf{T} \quad \Rightarrow \quad \tilde{\mathbf{X}} = \tilde{\mathbf{U}}\boldsymbol{\Sigma}\tilde{\mathbf{V}}^\mathbf{H}, \quad \tilde{\mathbf{U}} = \mathbf{A}\mathbf{U}, \; \tilde{\mathbf{V}} = \mathbf{H}\mathbf{V}$$

Architecture 2: Pre-mix + prefilter + non-uniform ADCs



- We can recover the ensemble $\tilde{\mathbf{X}}$ when

$$\text{total samples} \gtrsim RW \log^4(W)$$

- From $\tilde{\mathbf{X}}$, we recover $\mathbf{X}$ using

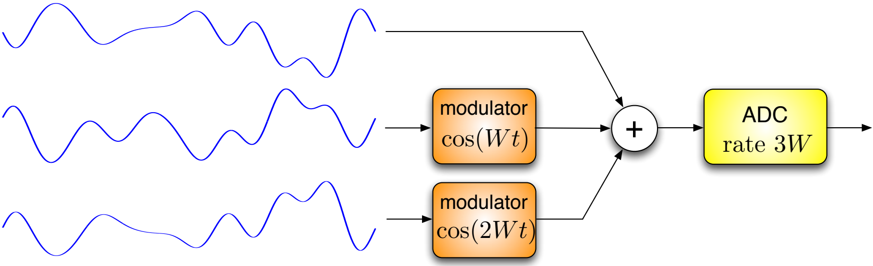
$$\mathbf{X} = \mathbf{A}^{\mathbf{T}}\tilde{\mathbf{X}}\mathbf{H}$$

- Universal, but still using an ADC for every channel...

Multiplexing onto one channel

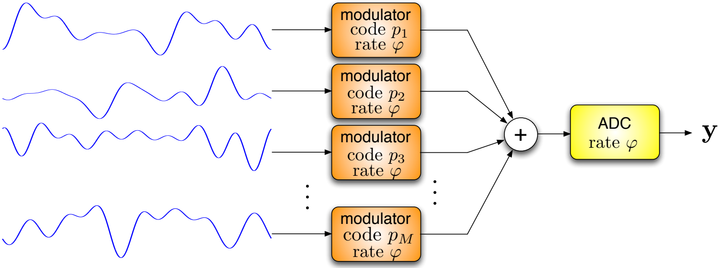
- We can always combine M channels into 1 by *multiplexing* in either time or frequency

  Frequency multiplexer:



- Replace M ADCs running at rate W with 1 ADC at rate MW

Architecture 3: modulated multiplexing

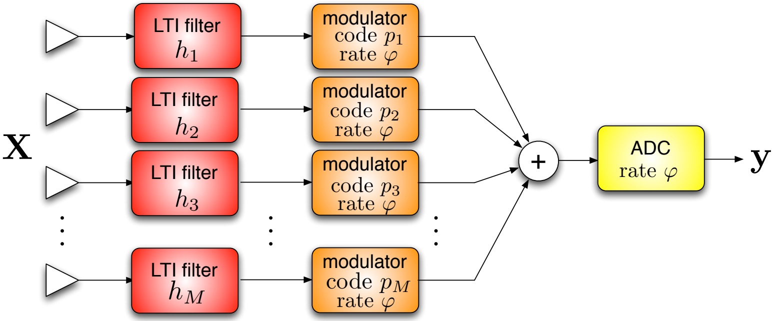


- If the signals are spread out uniformly in time, then the ADC and modulators can run at rate

$$\varphi \gtrsim RW \log^{3/2}(MW)$$

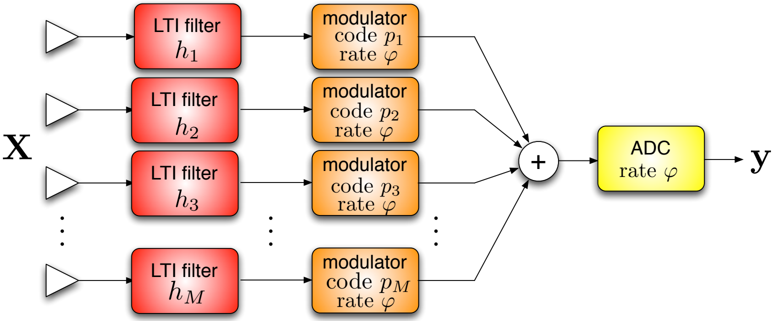
- This requires a (milder) “incoherence across time” assumption

Architecture 4: prefilter + modulated multiplexing



- Prefiltering spreads signals out over time (low-bandwidth filters)
- Modulate and sum diversifies and then combines the channels
- We use one standard ADC operating at rate φ (modulation rate is the same as the ADC sample rate)
- How big does φ need to be to recover $\mathbf{X}$?

Architecture 4: prefilter + modulated multiplexing



Matrix formulation

$$\mathbf{y} = \begin{bmatrix} \mathbf{H_1} & \mathbf{H_2} & \mathbf{H_3} & \cdots & \mathbf{H_M} \end{bmatrix} \mathbf{P}\, \mathrm{vec}(\mathbf{X})$$

($\varphi \times \varphi M$ random matrix)(samples of $\mathbf{X}$ at rate φ)

We have structured random linear measurements of a rank R matrix...

Compressive multiplexing theory

- Recht et al '07:
  Recovery is possible given the matrix-RIP:

$$(1-\delta)\|\mathbf{X}\|_F^2 \;\leq\; \|\mathcal{A}(\mathbf{X})\|_2^2 \;\leq\; (1+\delta)\|\mathbf{X}\|_F^2, \quad \forall \mathbf{X} : \mathrm{rank}(\mathbf{X}) \leq 2R,$$

Compressive multiplexing theory

- Recht et al '07 Candes and Plan '09
  The mRIP can follow from a certain concentration bound.

  If for any fixed $M \times W$ matrix $\mathbf{X}$ and some $0 < t < 1$ we have

  $$\mathrm{P}\left\{\left|\ \|\mathcal{A}(\mathbf{X})\|_2^2 - \|\mathbf{X}\|_F^2\right| > t\|\mathbf{X}\|_F^2\right\} \ \le\ Ce^{-\frac{p}{\mu}},$$

  then $\delta < .307$ for

  $$p \gtrsim \mu R W$$

Compressive multiplexing theory

- Krahmer and Ward '10:
  Modulating columns of a sparse-RIP matrix yields concentration.

  Suppose Φ satisfies

  $$(1-\delta)\|\mathbf{x}\|_2^2 \ \leq \ \|\mathbf{\Phi x}\|_2^2 \ \leq \ (1+\delta)\|\mathbf{x}\|_2^2 \quad \forall \ K\text{-sparse } \mathbf{x}.$$

  Set $\mathbf{\Phi}' = \mathbf{\Phi P}$. Then there is a $t < 1$ s.t. for *any* fixed $\mathbf{x}$

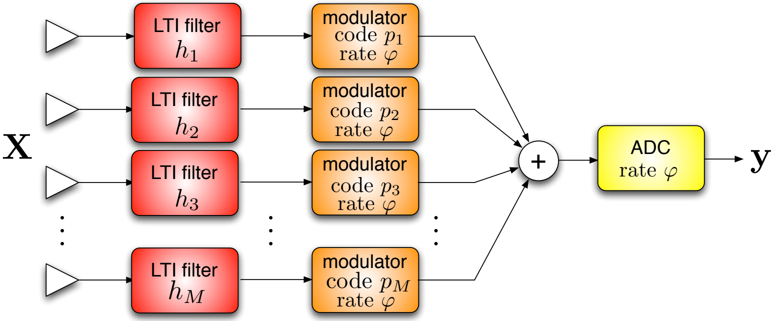
  $$\mathrm{P}\left\{\left| \|\mathbf{\Phi}'\mathbf{x}\|_2^2 - \|\mathbf{x}\|_2^2\right| > t\|\mathbf{x}\|_2^2\right\} \ \leq \ Ce^{-K/c}$$

Compressive multiplexing theory

- R '09:
  Concatenated random Toeplitz matrices obey a sparse-RIP.

  Take
  $$\mathbf{\Phi} = \begin{bmatrix}\mathbf{H_1} & \mathbf{H_2} & \cdots & \mathbf{H_M}\end{bmatrix}$$
  then with high probability
  $$(1-\delta)\|\mathbf{x}\|_2^2 \;\leq\; \|\mathbf{\Phi x}\|_2^2 \;\leq\; (1+\delta)\|\mathbf{x}\|_2^2 \quad \forall\ K\text{-sparse } \mathbf{x}$$
  when
  $$\varphi \;\gtrsim\; K \cdot \log^4(\varphi M)$$

Architecture 4: prefilter + modulated multiplexing



- We can stably recover a rank-R ensemble $\mathbf{X}$ when the modulators and ADC operate at rate

$$\varphi \gtrsim RW \log^4(MW)$$

- This architecture is *universal* in that it works for any low-rank correlation structure

Numerical experiments

A few data points form a rough idea of how this works in practice:

	$M \times W$	R	sample threshold	factor above RW
A2	300×1000	5	18400	3.68
	300×1000	7	20800	2.97
A3	400×1000	2	7600	3.8
	400×1000	4	10680	2.67
	400×1000	7	14000	2
A4	100×1000	4	6600	1.65
	100×1000	7	10000	1.43
	300×1000	5	13000	2.6
	300×1000	7	14500	2.07

A2 is non-uniform sampling (matrix samples)

A3 is modulated multiplexing

A4 is prefilter+modulated multiplexing

Summary

- We saw several *compressive sampling architectures* for acquiring ensembles of *correlated signals* where the total number of samples we take scales like

$$(\text{bandwidth}) \times (\text{rank})$$

(to within log factors)