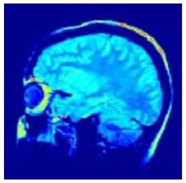
Measurements and Bits:

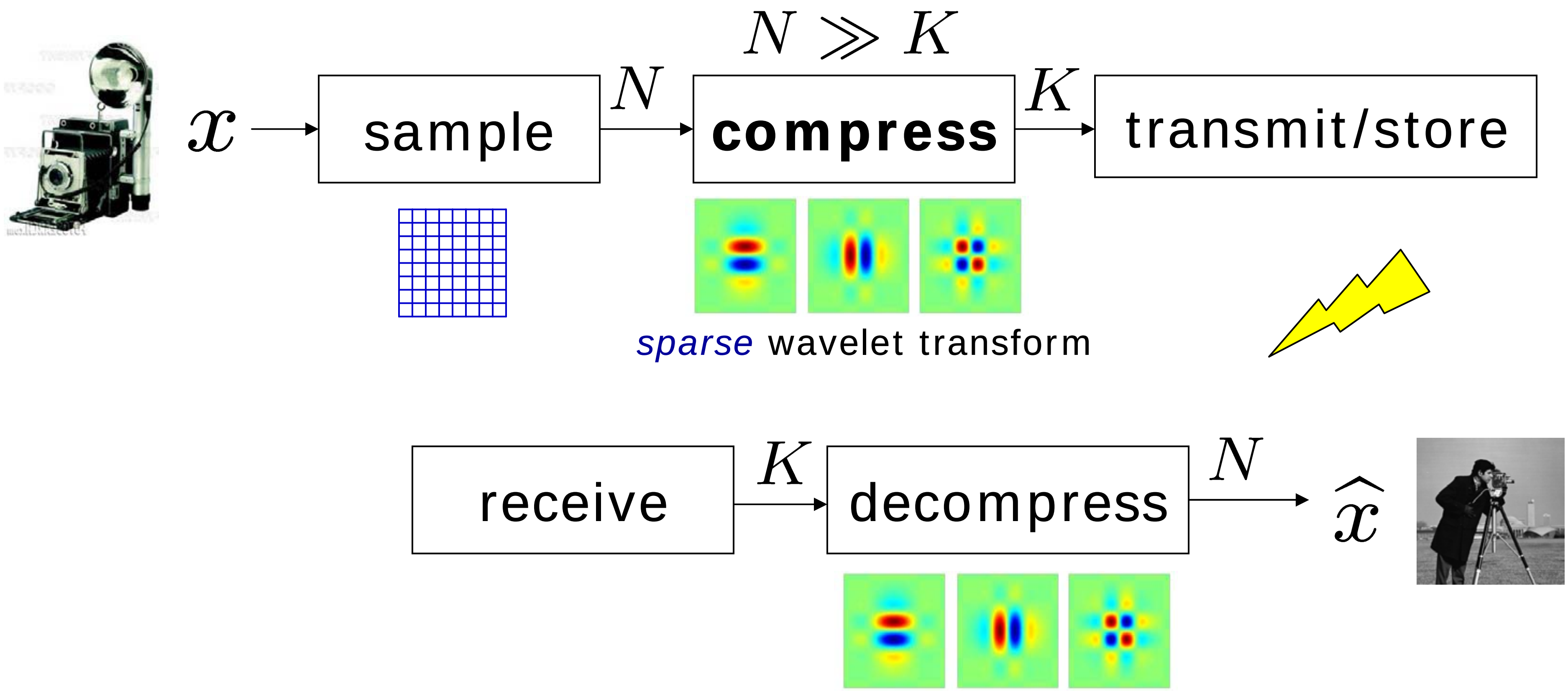
Compressed Sensing
meets
Information Theory



Dror Baron
ECE Department
Rice University
dsp.rice.edu/cs

Sensing by *Sampling*

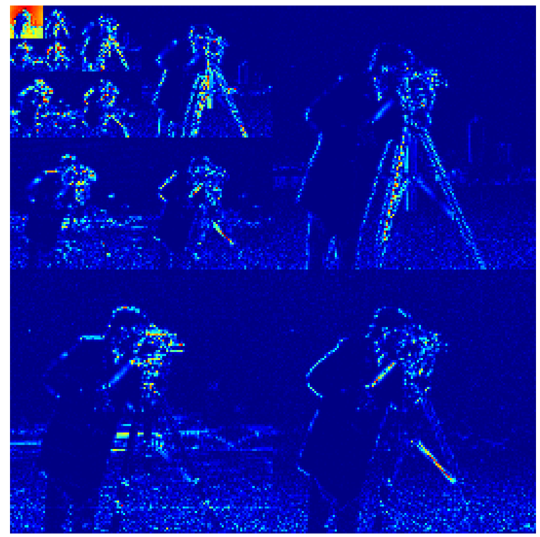
- ***Sample*** data at Nyquist rate
- ***Compress*** data using model (e.g., sparsity)
  - encode coefficient locations and values
- *Lots of work* to throw away >80% of the coefficients
- Most computation at *sensor* (asymmetrical)
- Brick wall to performance of modern acquisition systems



Sparsity / Compressibility

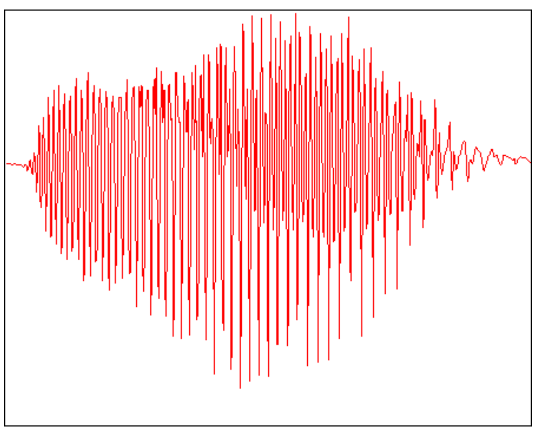
- Many signals are *sparse* or *compressible* in some representation/basis (Fourier, wavelets, ...)

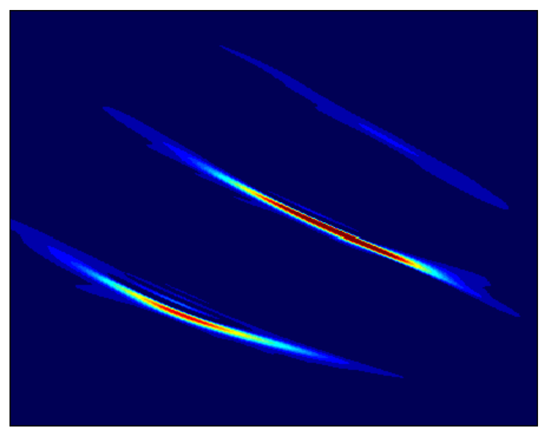
N
pixels



$K \ll N$
large
wavelet
coefficients

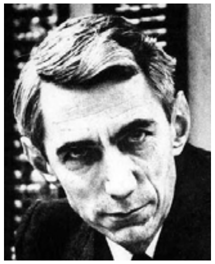
N
wideband
signal
samples





$K \ll N$
large
Gabor
coefficients

Compressed Sensing



- Shannon/Nyquist sampling theorem
  - worst case bound for *any* bandlimited signal
  - *too pessimistic* for some classes of signals
  - does not exploit signal *sparsity/compressibility*

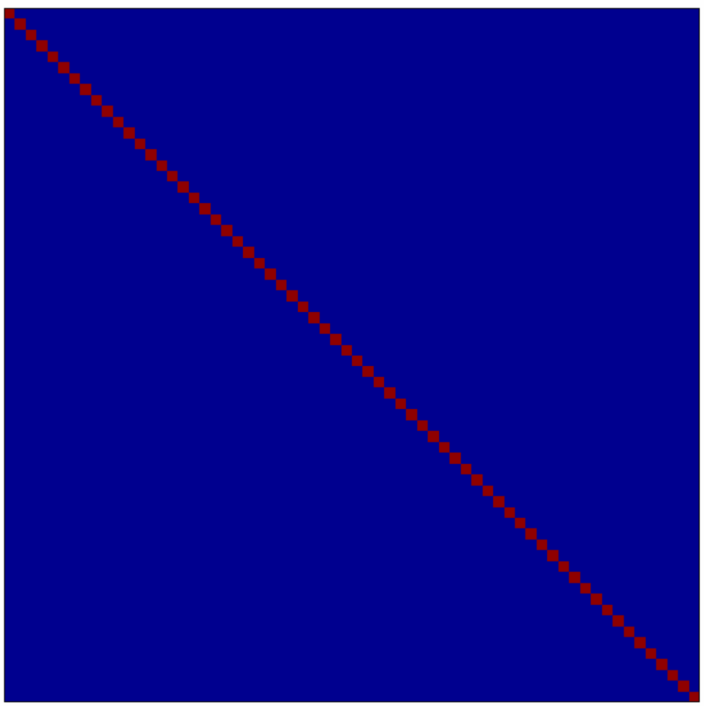
- *Seek direct sensing of compressible information*

- Compressed Sensing (CS)
  - sparse signals can be recovered from a small number of nonadaptive (fixed) linear measurements
  - [Candes et al.; Donoho; Kashin; Gluskin; Rice...]
  - based on new uncertainty principles beyond Heisenberg ("incoherency")

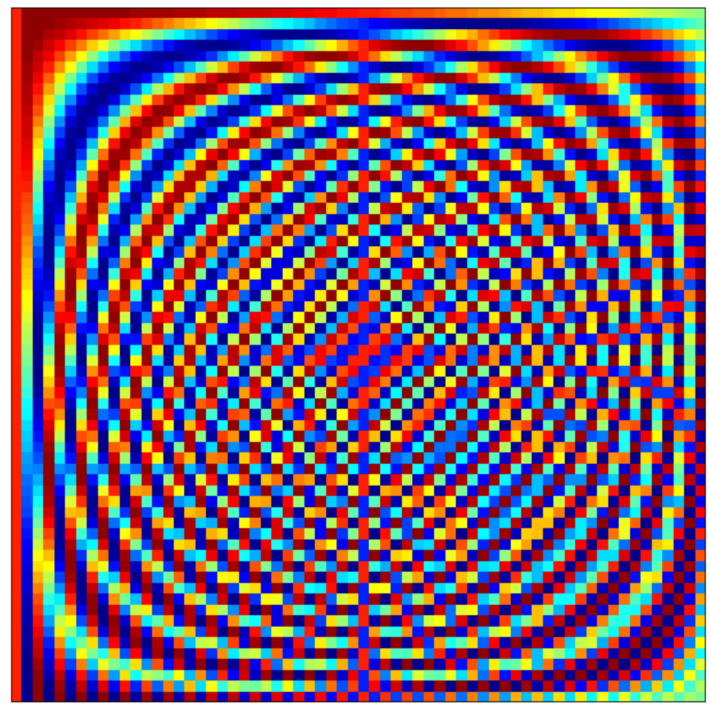
Incoherent Bases (matrices)

- Spikes and sines (Fourier)

$$\Psi = I$$



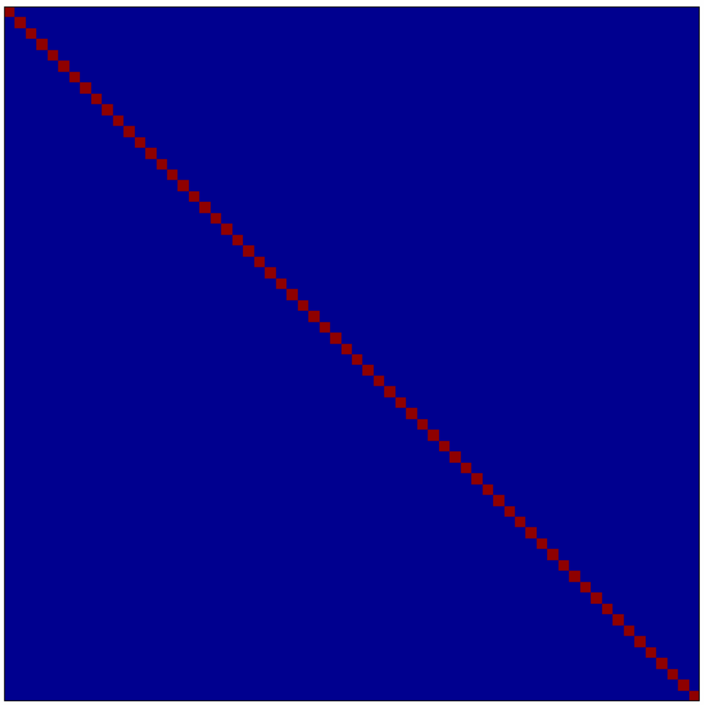
$$\Phi = \mathrm{idct}(I)$$



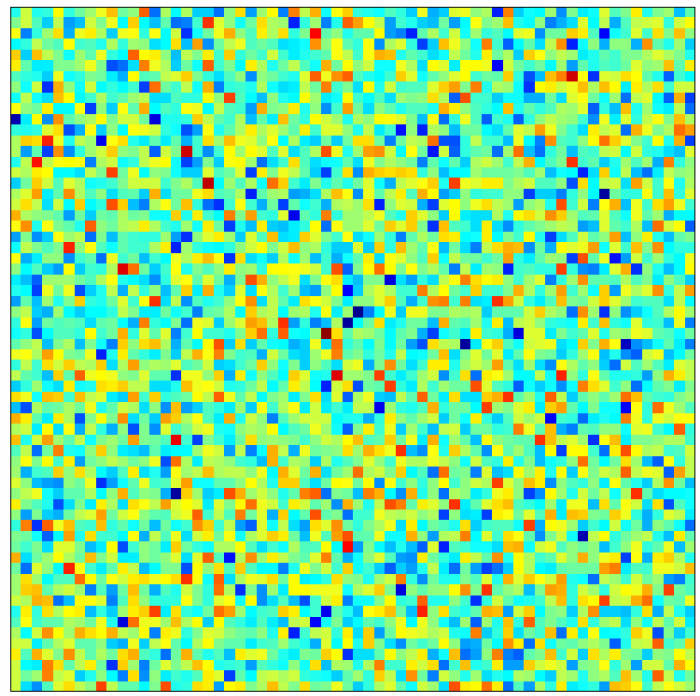
Incoherent Bases

- Spikes and "random noise"

$$\Psi = I$$

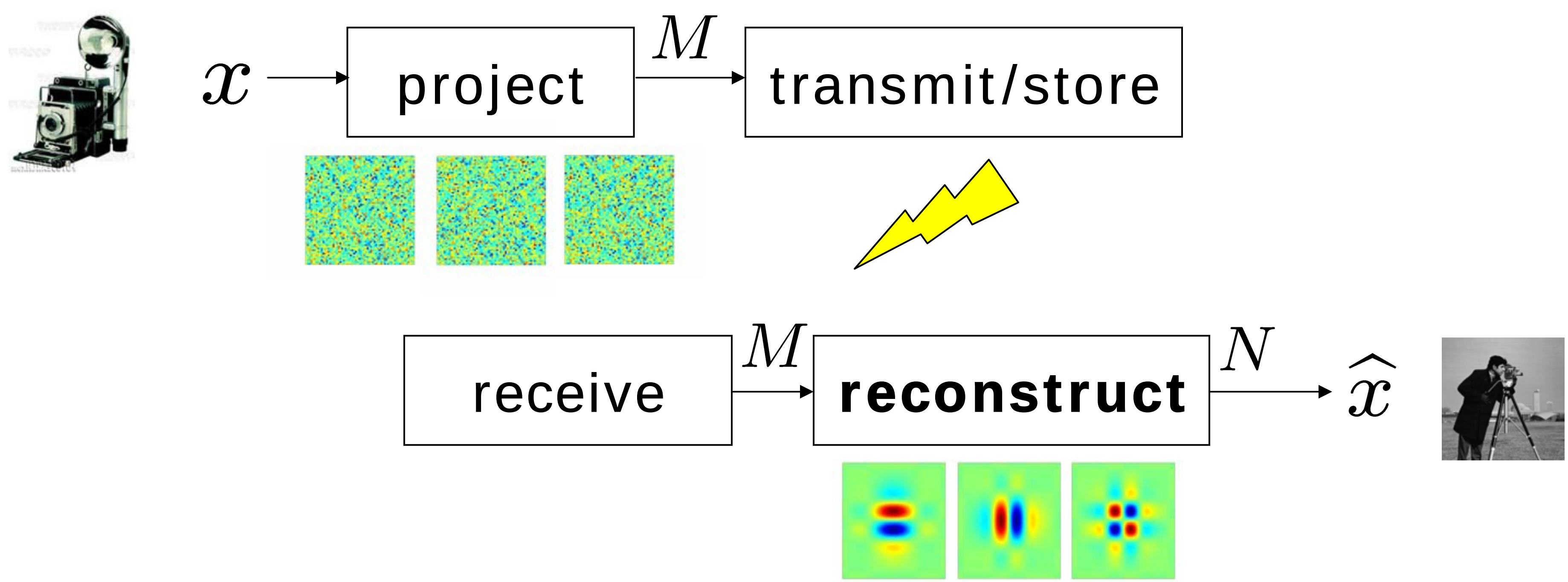


$$\Phi = \text{randn}(N, N)$$



Compressed Sensing via ***Random Projections***

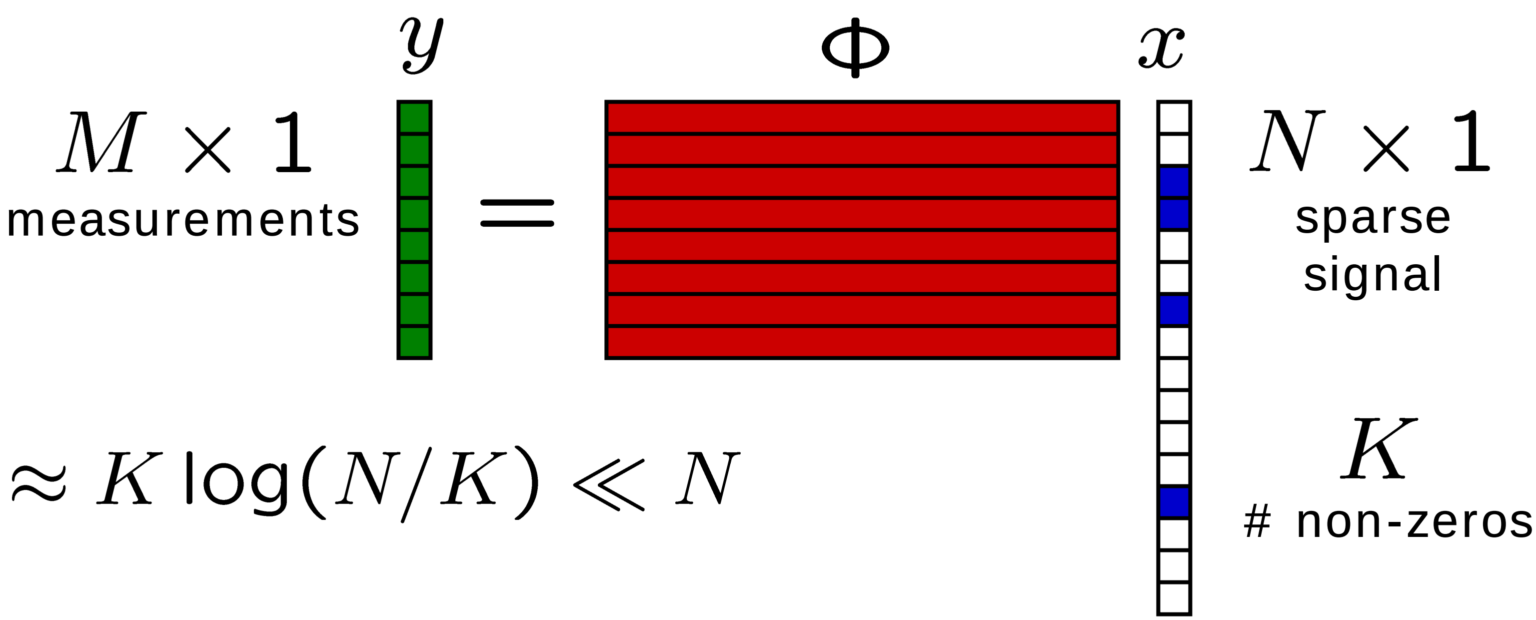
- Measure linear projections onto *incoherent* basis where data is *not sparse/compressible*
  - random projections are *universally incoherent*
  - fewer measurements $M \approx K \log(N/K) \ll N$
  - no location information
- Reconstruct via optimization
- Highly asymmetrical (most computation at *receiver*)



CS Encoding

- Replace **samples** by more general **encoder** based on a few linear projections (inner products)
- Matrix vector multiplication – potentially *analog*

$$y = \Phi x$$

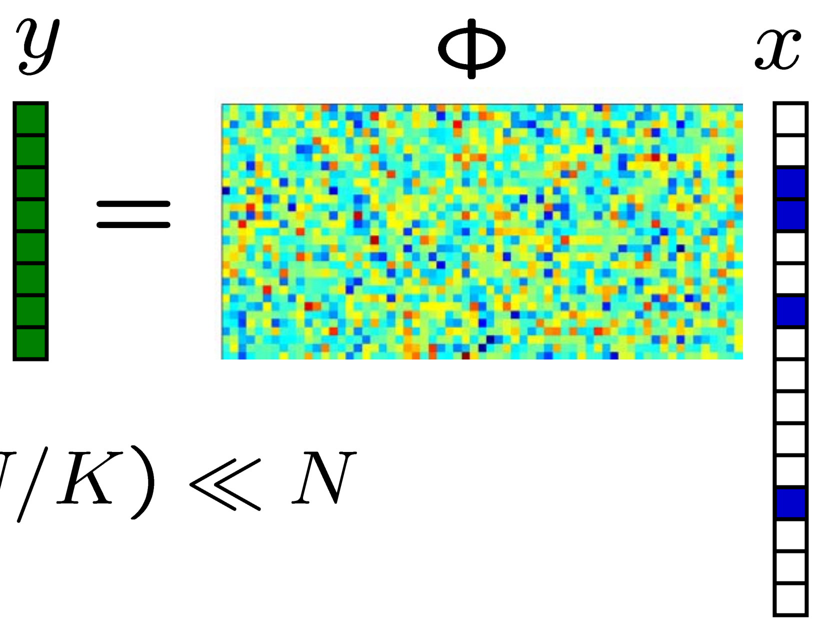


$$M \approx K \log(N/K) \ll N$$

Universality via Random Projections

- **Random** projections
- *Universally incoherent* with any compressible/sparse signal class

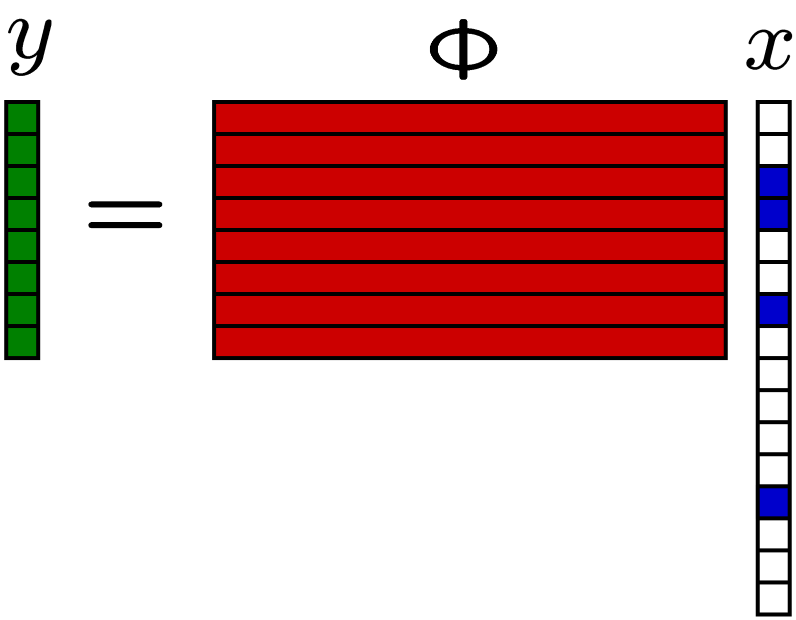
$$y = \Phi x$$



$$M \approx K \log(N/K) \ll N$$

Reconstruction Before-CS – ℓ_2

- **Goal:** Given measurements y find signal x
- Fewer rows than columns in measurement matrix Φ
- *Ill-posed*: infinitely many solutions $\widehat{x}$
- Classical solution: *least squares*



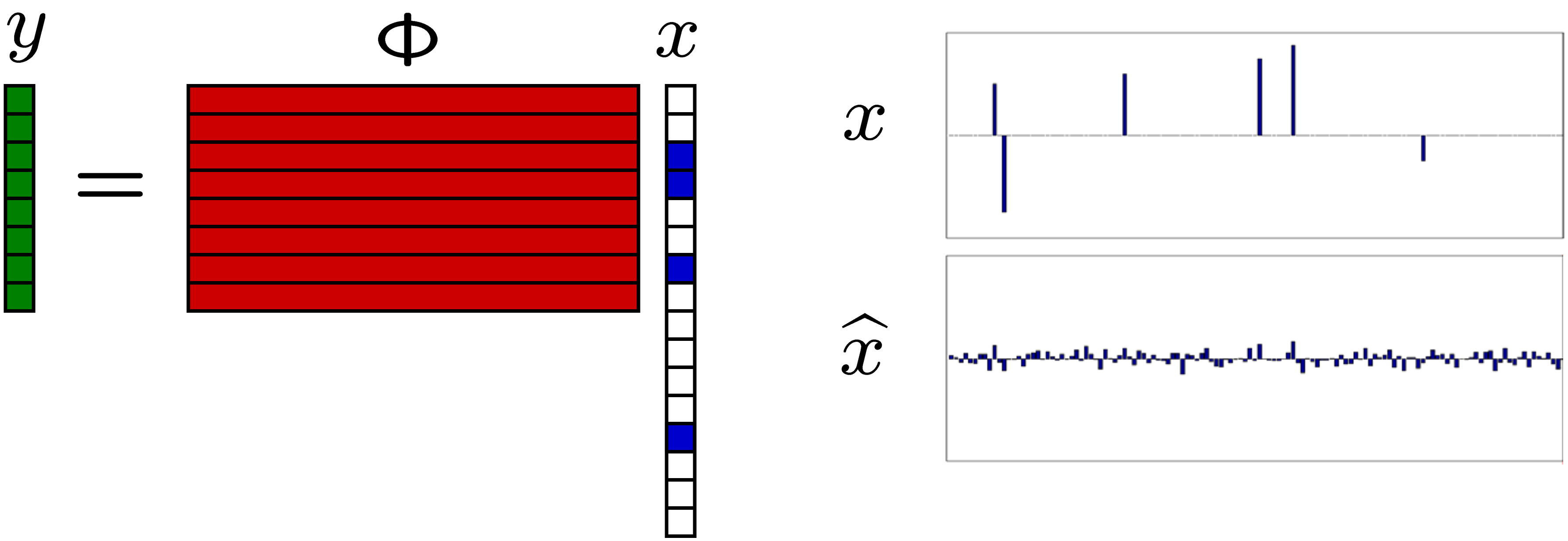
$$\widehat{x} = \arg\min_{y=\Phi x} \|x\|_2$$

$$\sum_i |x_i|^2$$

$$\widehat{x} = (\Phi^T \Phi)^{-1} \Phi^T y$$

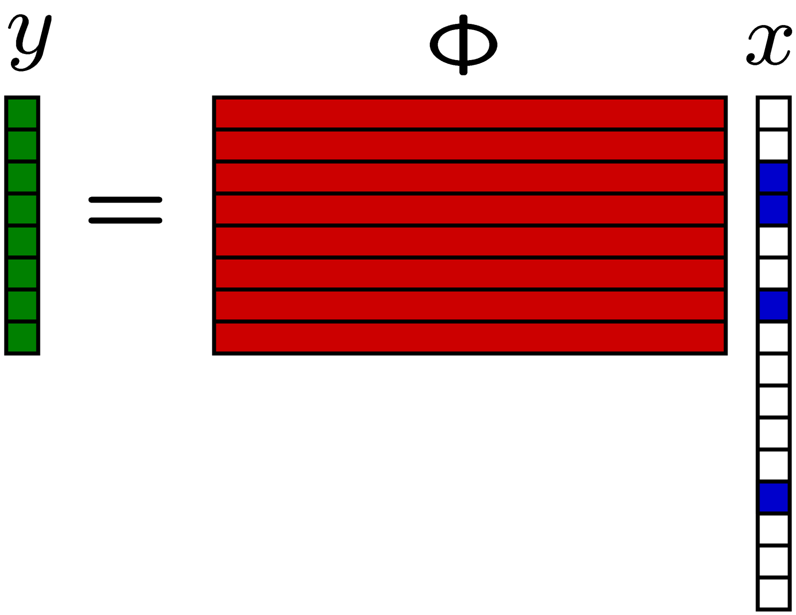
Reconstruction Before-CS – ℓ_2

- Goal: Given measurements y find signal x
- Fewer rows than columns in measurement matrix Φ
- *Ill-posed*: infinitely many solutions $\widehat{x}$
- Classical solution: *least squares*
- Problem: *small L_2 doesn't imply sparsity*



Ideal Solution – ℓ_0

- *Ideal* solution: exploit sparsity of x
- Of the infinitely many solutions $\widehat{x}$ seek *sparsest* one

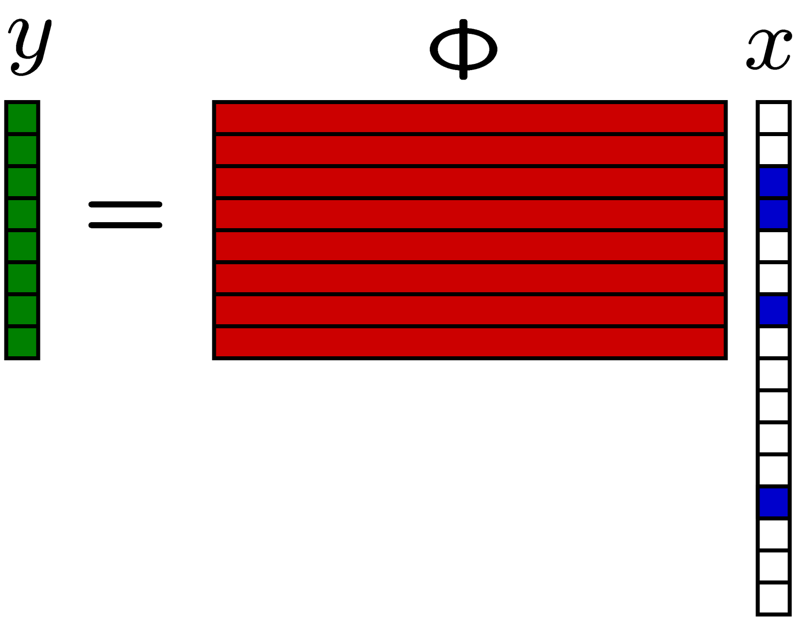


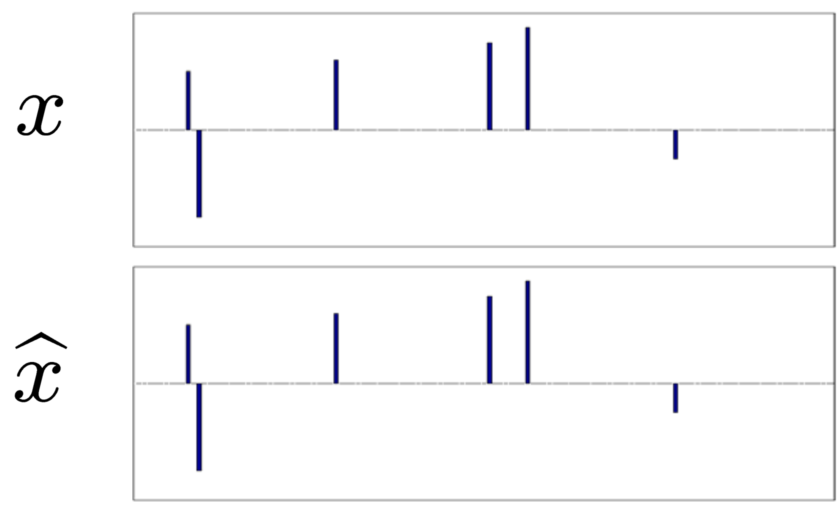
$$\widehat{x} = \arg\min_{y=\Phi x} \|x\|_0$$

number of nonzero entries

Ideal Solution – ℓ_0

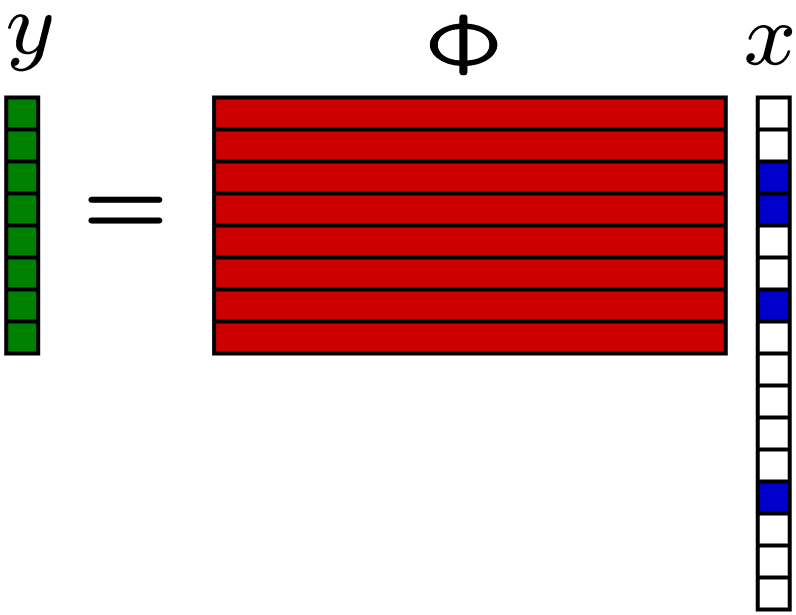
- *Ideal* solution: exploit sparsity of x
- Of the infinitely many solutions $\widehat{x}$ seek *sparsest* one
- If M · K then w/ high probability this can't be done
- If M ¸ K+1 then ***perfect reconstruction*** w/ high probability [Bresler et al.; Wakin et al.]
- But not robust and *combinatorial* complexity





The CS Revelation – ℓ_1

- Of the infinitely many solutions $\widehat{x}$ seek the one with smallest ℓ_1 norm



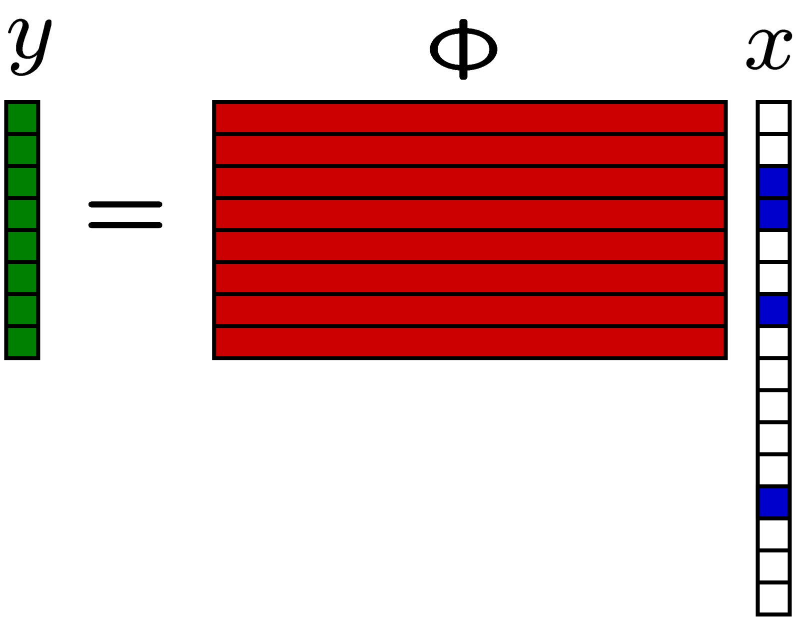
$$\widehat{x} = \arg\min_{y=\Phi x} \|x\|_1$$

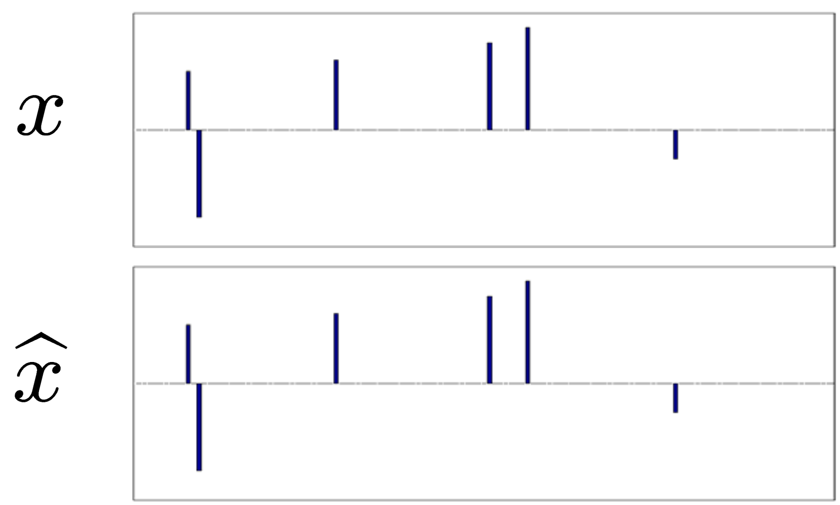
↑

$\sum_i |x_i|$

The CS Revelation – ℓ_1

- Of the infinitely many solutions $\widehat{x}$ seek the one with smallest ℓ_1 norm
- If $M \approx K \log(N/K)$ then *perfect reconstruction* w/ high probability [Candes et al.; Donoho]
- Robust to measurement noise
- *Linear programming*





CS Hallmarks

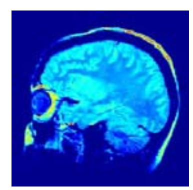
- CS changes the rules of data acquisition game
  - exploits a priori signal *sparsity* information (signal is compressible)

- **Hardware: *Universality***
  - same random projections / hardware for *any* compressible signal class
  - simplifies hardware and algorithm design

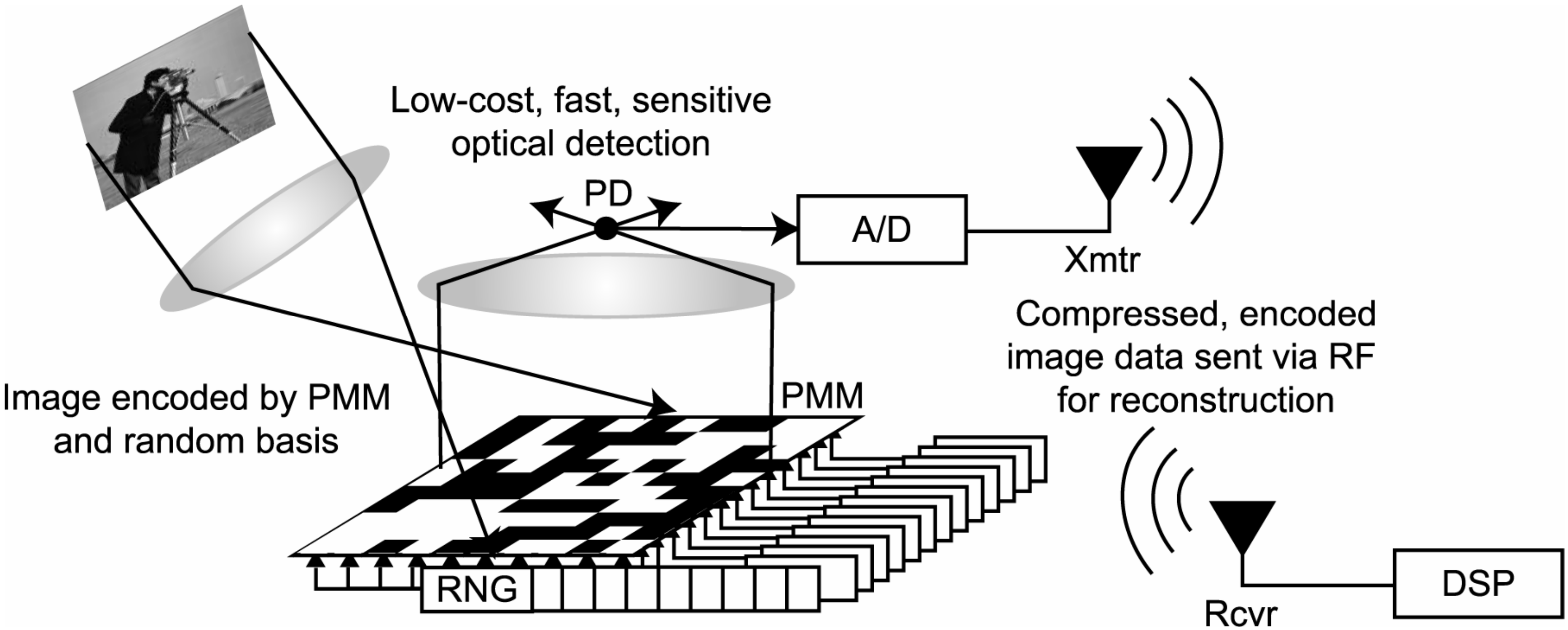
- **Processing: *Information scalability***
  - random projections ~ sufficient statistics
  - same random projections for range of tasks
    - reconstruction > estimation > recognition > detection
  - far fewer measurements required to detect/recognize

- Next generation data acquisition
  - new imaging devices and A/D converters [DARPA]
  - new reconstruction algorithms
  - new distributed source coding algorithms [Baron et al.]

Random Projections in Analog



Optical Computation of Random Projections



- CS encoder integrates sensing, compression, processing
- Example: new *cameras* and *imaging* algorithms

First Image Acquisition (M=0.38N)

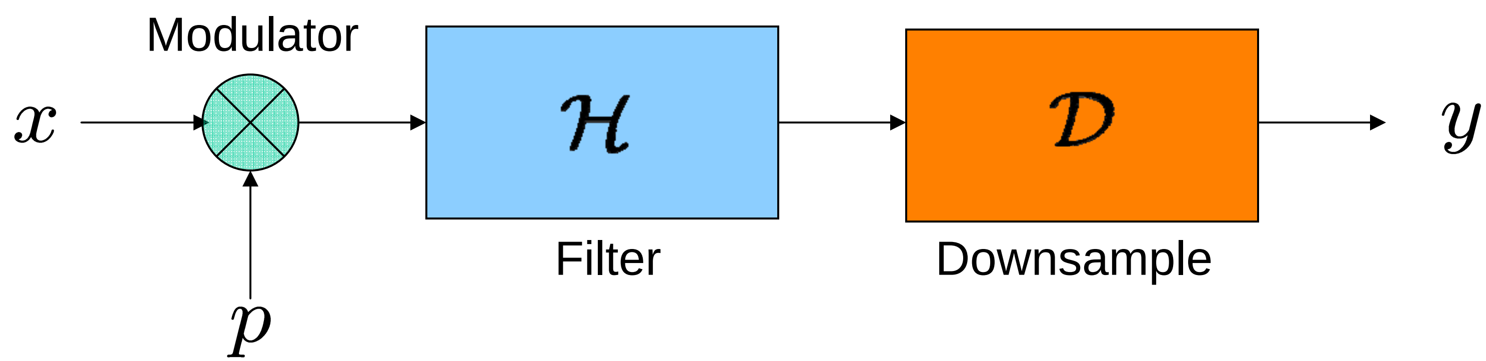
ideal 64x64 image
(4096 pixels)

400
wavelets

image on
DMD array

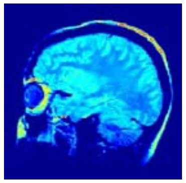
1600
random meas.

A/D Conversion Below Nyquist Rate



- *Challenge:*
  - wideband signals (radar, communications, ...)
  - currently impossible to sample at Nyquist rate
- Proposed CS-based solution:
  - sample at “information rate”
  - simple hardware components
  - good reconstruction performance

Connections Between Compressed Sensing and Information Theory



Measurement Reduction via CS

- CS reconstruction via ℓ_1
  - If $M \approx K \log(N/K)$ then *perfect reconstruction* w/ high probability [Candes et al.; Donoho]
  - *Linear programming*

- *Compressible* signals (signal components decay)
  - also requires $M = O(K \log(N/K))$
  - polynomial complexity (BPDN) [Candes et al.]
  - cannot reduce order of M [Kashin,Gluskin]

Fundamental Goal: Minimize M

- Compressed sensing aims to minimize resource consumption due to measurements

- Donoho:

  “Why go to so much effort to acquire all the data when most of what we get will be thrown away?”

Fundamental Goal: Minimize M

- Compressed sensing aims to minimize resource consumption due to measurements

- Donoho:

  "Why go to so much effort to acquire all the data when most of what we get will be thrown away?"

- Recall *sparse signals*
  - only $M = K + 1$ measurements for ℓ_0 reconstruction
  - not robust and *combinatorial* complexity

Rich Design Space

- *What performance metric to use?*
  - Determine support set of nonzero entries [Wainwright]
    - this is ℓ_0 distortion metric
    - but why let tiny nonzero entries spoil the fun?
  - ℓ_1 metric? ℓ_2?

Rich Design Space

- What performance metric to use?
  - Determine support set of nonzero entries [Wainwright]
    - this is ℓ_0 distortion metric
    - but why let tiny nonzero entries spoil the fun?
  - ℓ_1 metric? ℓ_2?

- *What complexity class of reconstruction algorithms?*
  - any algorithms?
  - polynomial complexity?
  - near-linear or better?

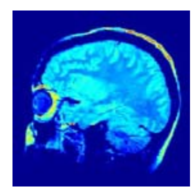
Rich Design Space

- What performance metric to use?
  - Determine support set of nonzero entries [Wainwright]
    - this is ℓ_0 distortion metric
    - but why let tiny nonzero entries wreck spoil the fun?
  - ℓ_1 metric? ℓ_2?

- What complexity class of reconstruction algorithms?
  - any algorithms?
  - polynomial complexity?
  - near-linear or better?

- *How to account for imprecisions?*
  - noise in measurements?
  - compressible signal model?

Lower Bound on Number of Measurements



Measurement Noise

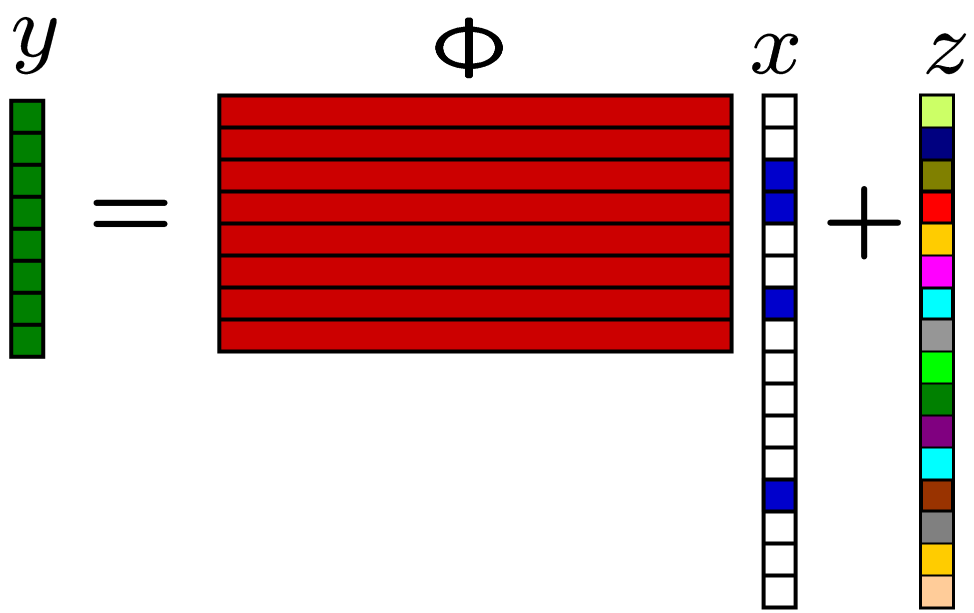
- Measurement process is analog
- Analog systems add noise, non-linearities, etc.

- *Assume Gaussian noise for ease of analysis*

Setup

- Signal x is iid $x_i \sim p_X(x)$
- Additive white Gaussian noise $z_i \sim \mathcal{N}(0,1)$
- Noisy measurement process

$$y = y_0 + z = \Phi x + z$$

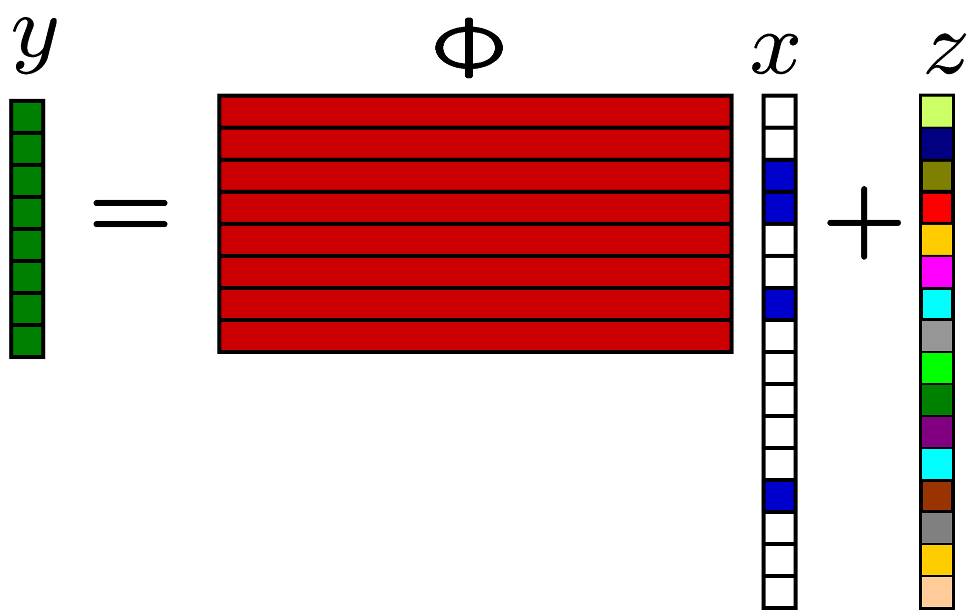


Setup

- Signal x is iid $x_i \sim p_X(x)$
- Additive white Gaussian noise $z_i \sim \mathcal{N}(0,1)$
- Noisy measurement process

$$y = y_0 + z = \Phi x + z$$

- *Random projection of tiny coefficients (compressible signals) similar to measurement noise*



Measurement and Reconstruction Quality

- Measurement signal to noise ratio

$$\mathrm{SNR} = \frac{E[\|y_0\|_2^2]}{E[\|z\|_2^2]} = \frac{E[\|y_0\|_2^2]}{M}$$

- Reconstruct using decoder mapping $D_x : \;\; y \rightarrow \widehat{x}$

- Reconstruction distortion metric

$$D = \frac{E[\|\widehat{x} - x\|_2^2]}{E[\|x\|_2^2]}$$

- Goal: minimize CS *measurement rate*

$$\delta = \lim_{N \rightarrow \infty} \inf_{\{D_x:\ \mathrm{SNR},\ \text{achieves}\ D\}} \frac{M}{N}$$

Measurement Channel

- Model process $y_0 \longrightarrow y$ as *measurement channel*

- *Capacity* of measurement channel

$$C = \frac{1}{2}\log_2(1 + \mathsf{SNR})$$

- ***Measurements are bits!***

Lower Bound [Sarvotham et al.]

- **Theorem:** For a sparse signal with rate-distortion function $R(D)$, lower bound on measurement rate δ subject to measurement quality SNR and reconstruction distortion D satisfies

$$\delta \geq \frac{2R(D)}{\log_2(1+\mathsf{SNR})}$$

- Direct relationship to *rate-distortion* content

- *Applies to any linear signal acquisition system*

Lower Bound [Sarvotham et al.]

- **Theorem:** For a sparse signal with rate-distortion function $R(D)$, lower bound on measurement rate δ subject to measurement quality SNR and reconstruction distortion D satisfies

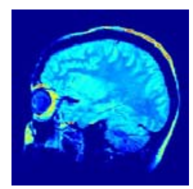
$$\delta \geq \frac{2R(D)}{\log_2(1+\text{SNR})}$$

- **Proof sketch:**
  - each measurement provides $C = \frac{1}{2}\log_2(1+\text{SNR})$ bits
  - information content of source $\approx NR(D)$ bits
  - source-channel separation for continuous amplitude sources
  - minimal number of measurements $M \approx \frac{NR(D)}{\frac{1}{2}\log_2(1+\text{SNR})}$
  - obtain measurement rate δ via normalization by N

Example

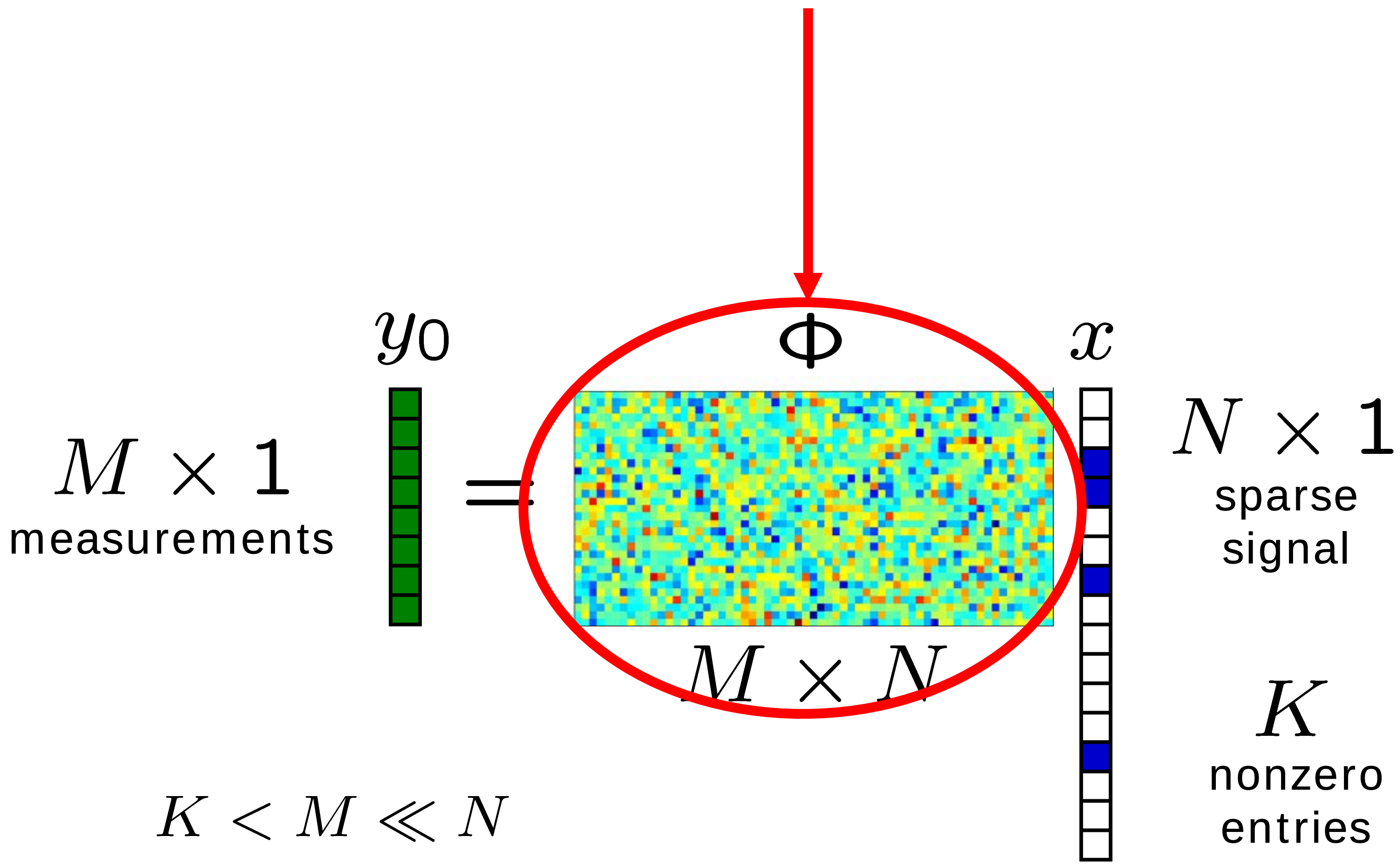
- Spike process - K spikes of uniform amplitude
- Rate-distortion function $NR(D) \approx K \log(N/K)$
- Lower bound $\delta \gtrsim \frac{2K \log_2(N/K)}{N \log_2(1+\mathrm{SNR})}$
- Numbers:
  - signal of length 10^7
  - 10^3 spikes
  - SNR=10 dB $\Rightarrow$ $M \gtrsim 7,682$
  - SNR=-20 dB $\Rightarrow$ $M \gtrsim 1.85 \cdot 10^6$
- *If interesting portion of signal has relatively small energy then need significantly more measurements!*
- Upper bound (achievable) in progress...

CS Reconstruction
Meets
Channel Coding



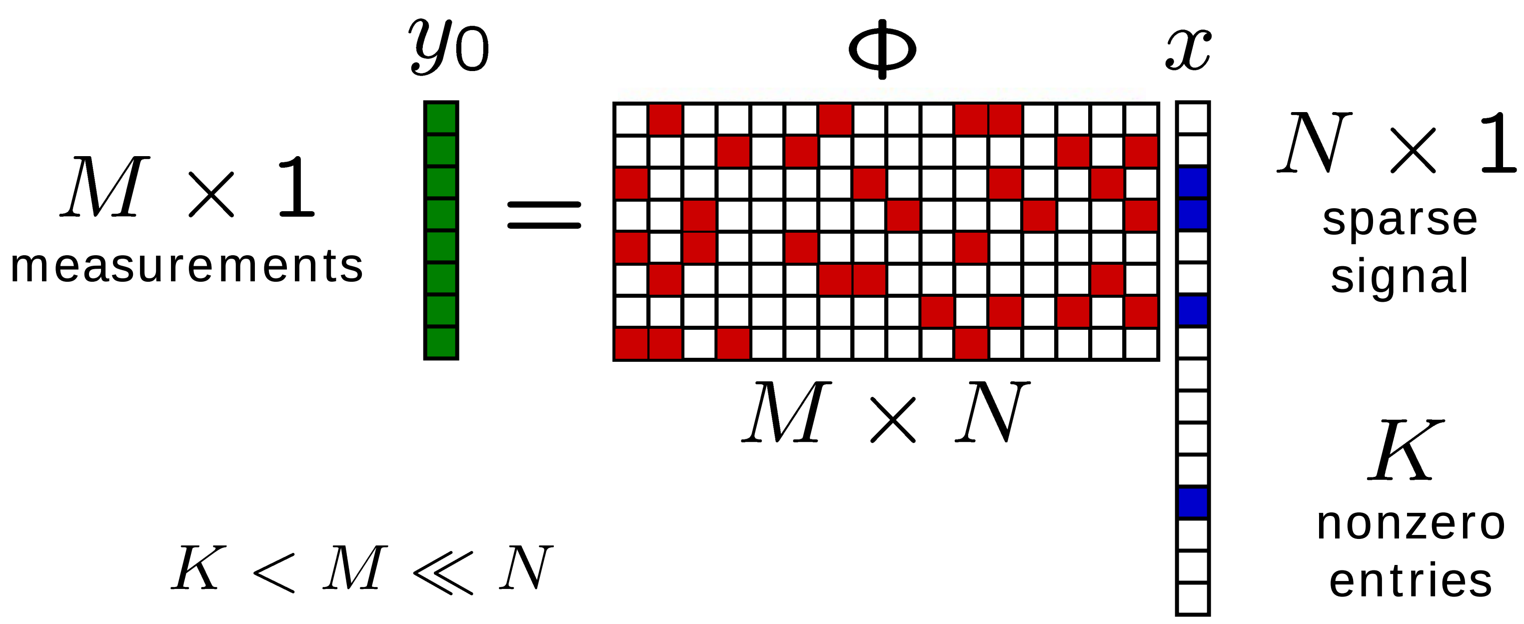
Why is Reconstruction Expensive?

Culprit: dense, unstructured Φ



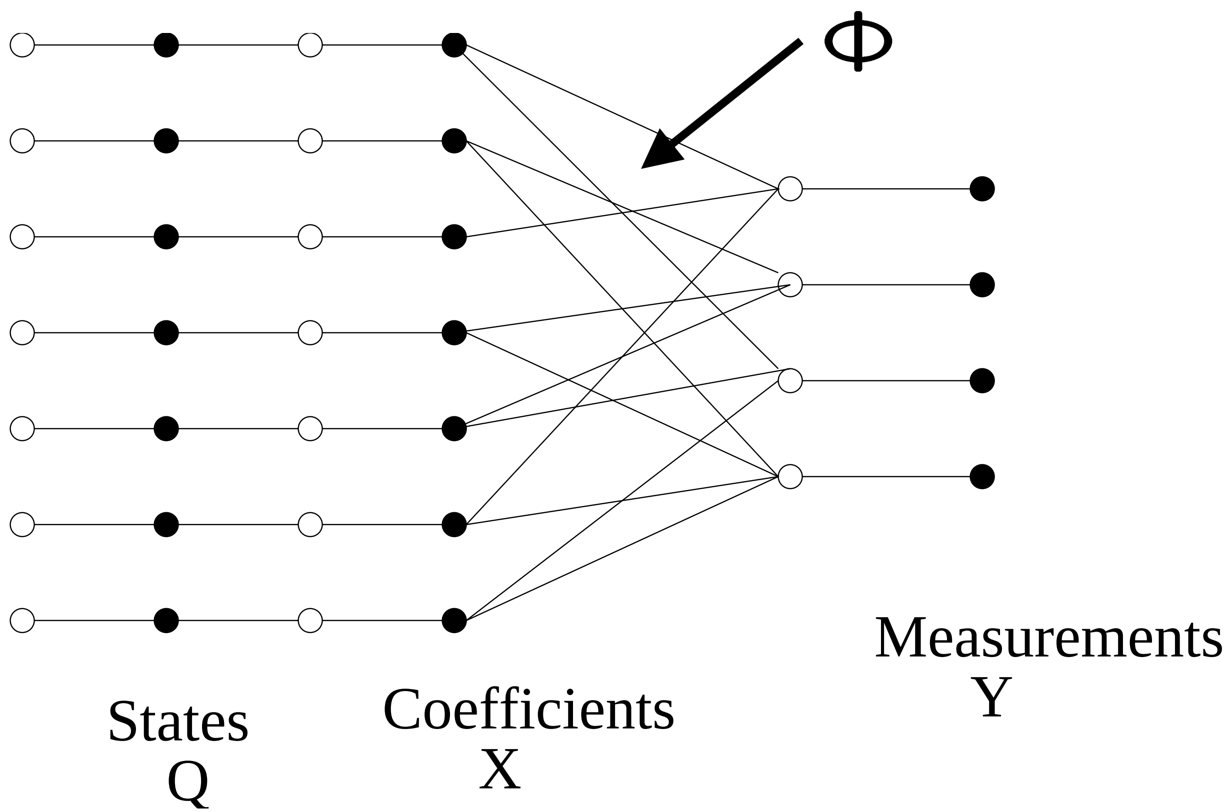
Fast CS Reconstruction

- LDPC measurement matrix (sparse)
- Only 0/1 in Φ
- Each row of Φ contains L randomly placed *1*'s
- Fast matrix multiplication
  - ✓ fast encoding
  - ✓ fast reconstruction

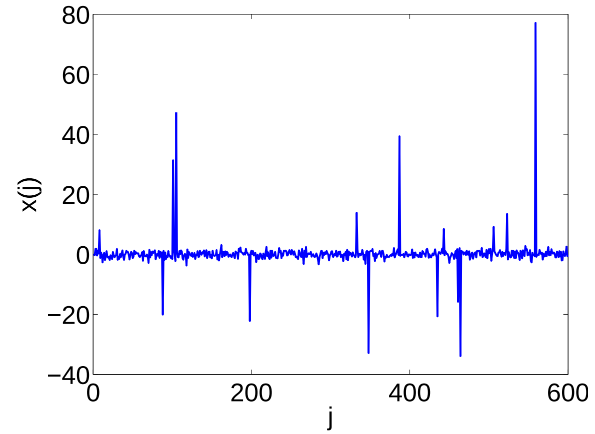


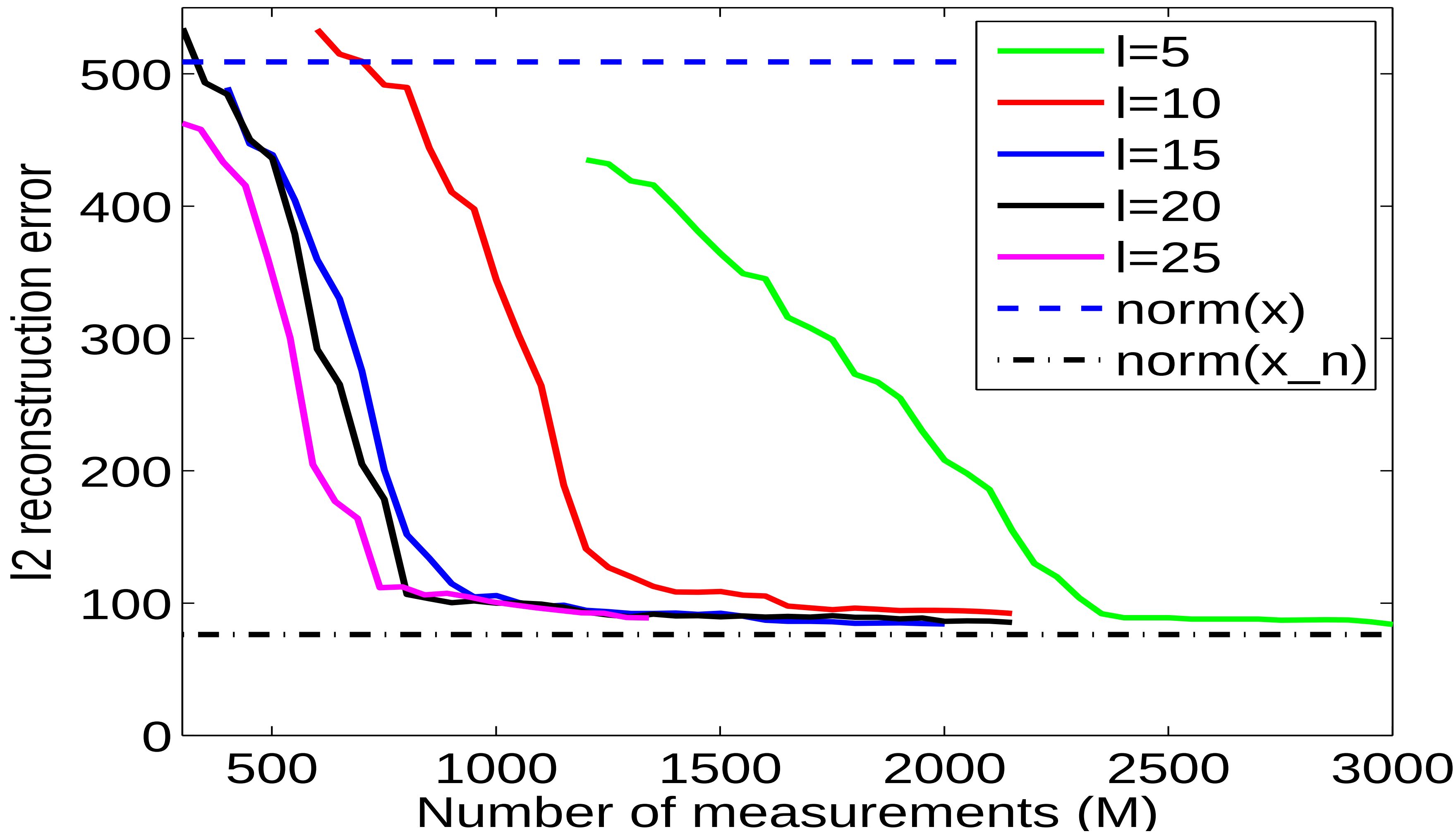
Ongoing Work: CS Using BP

- Considering noisy CS signals
- Application of Belief Propagation
  - BP over real number field
  - sparsity is modeled as prior in *graph*



Promising Results





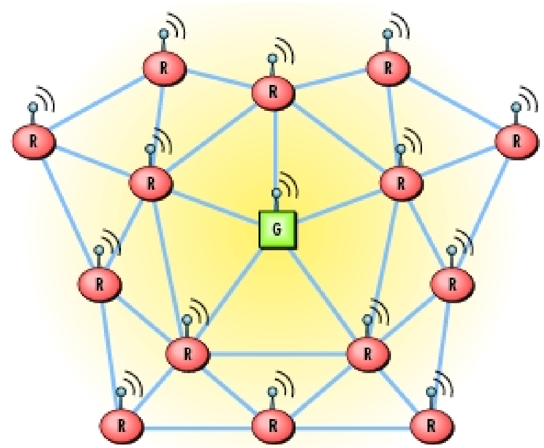
Theoretical Advantages of CS-BP

- Low complexity $O(N \log(N))$
- Provable reconstruction with noisy measurements using $M = O(K \log(N/K))$
- *Success of LDPC+BP in channel coding carried over to CS!*

Distributed

Compressed Sensing (DCS)

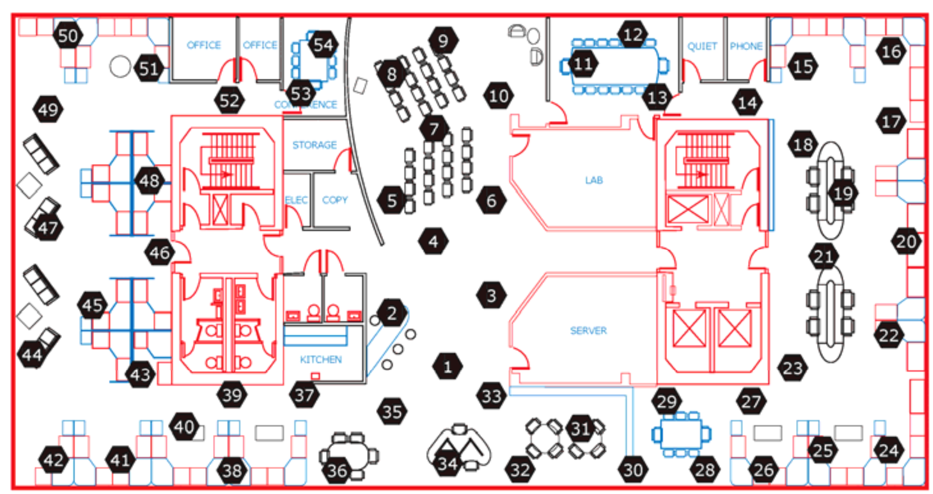
CS for distributed signal ensembles



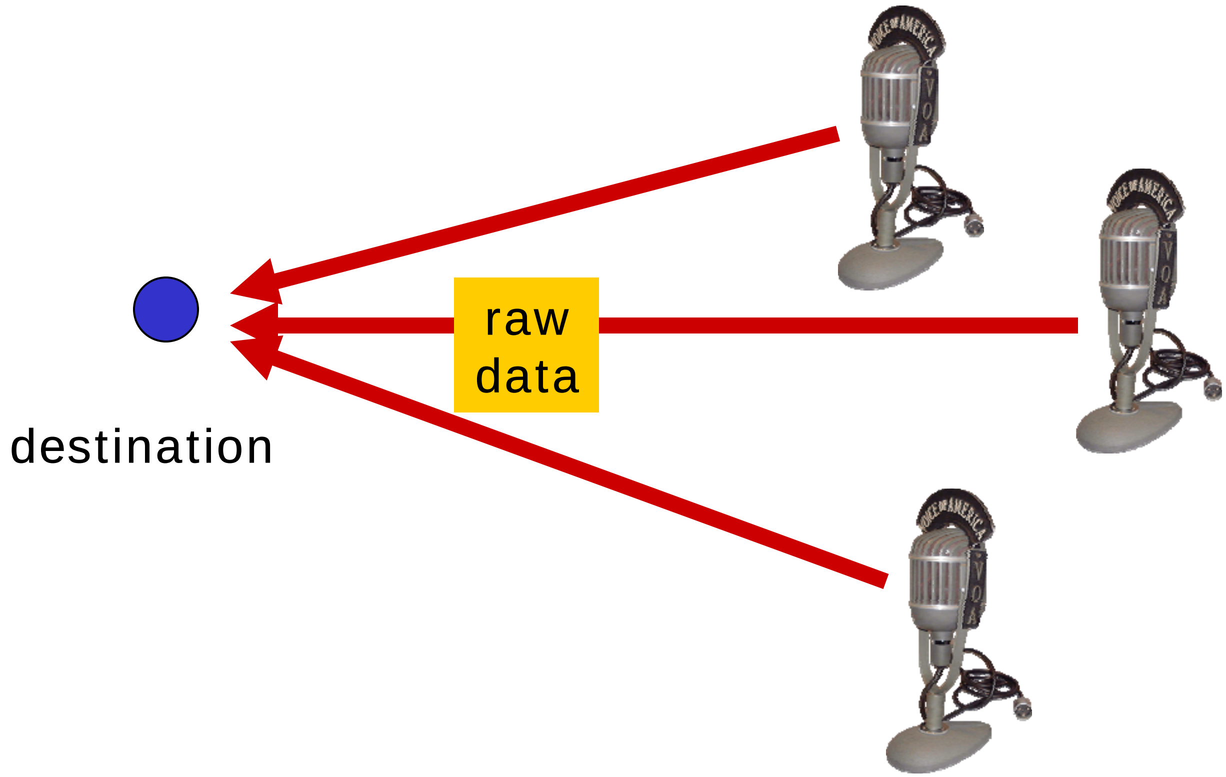
Why Distributed?

- Networks of many *sensor nodes*
  - sensor, microprocessor for computation, wireless communication, networking, battery
  - can be spread over large geographical area

- Must be *energy efficient*
  - *minimize communication* at expense of computation
  - motivates *distributed compression*

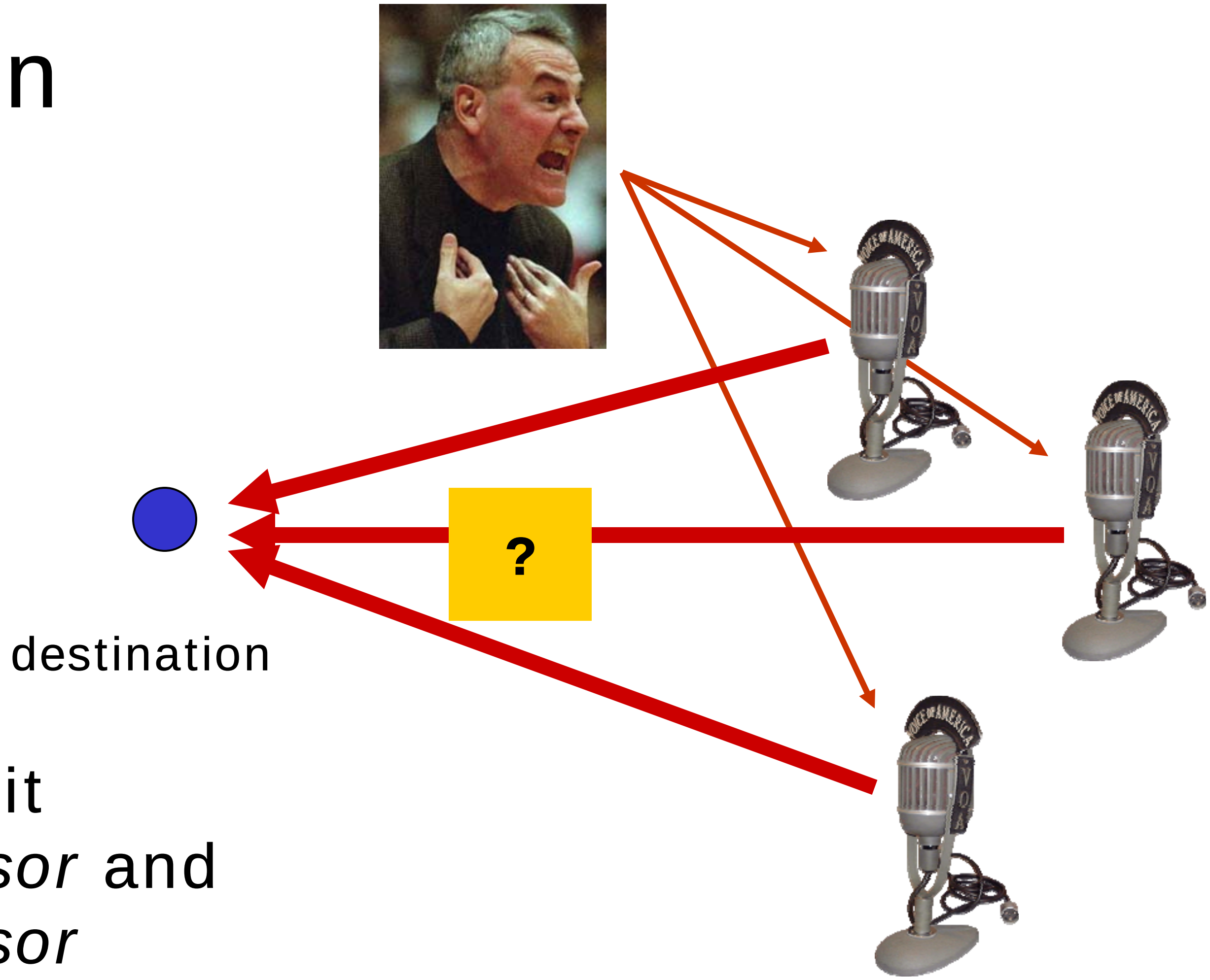


Distributed Sensing



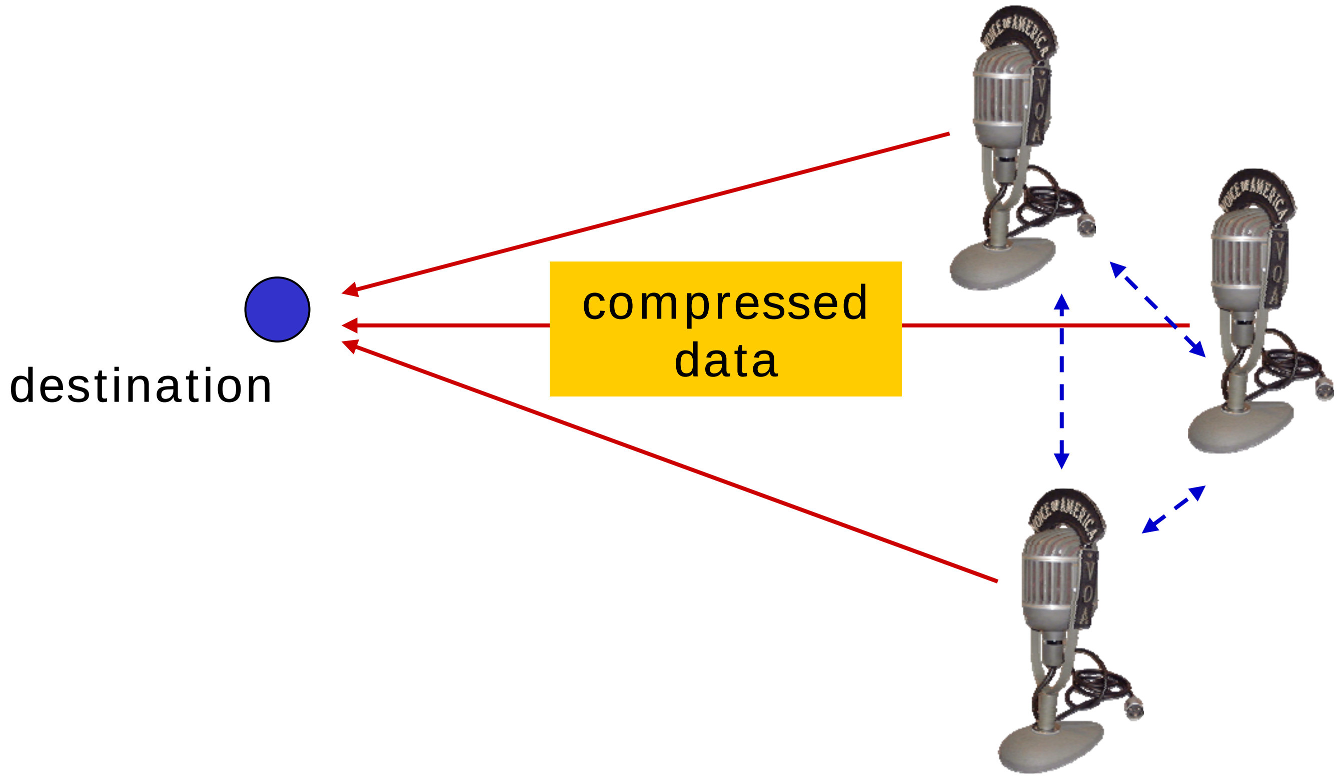
- Transmitting raw data typically inefficient

Correlation



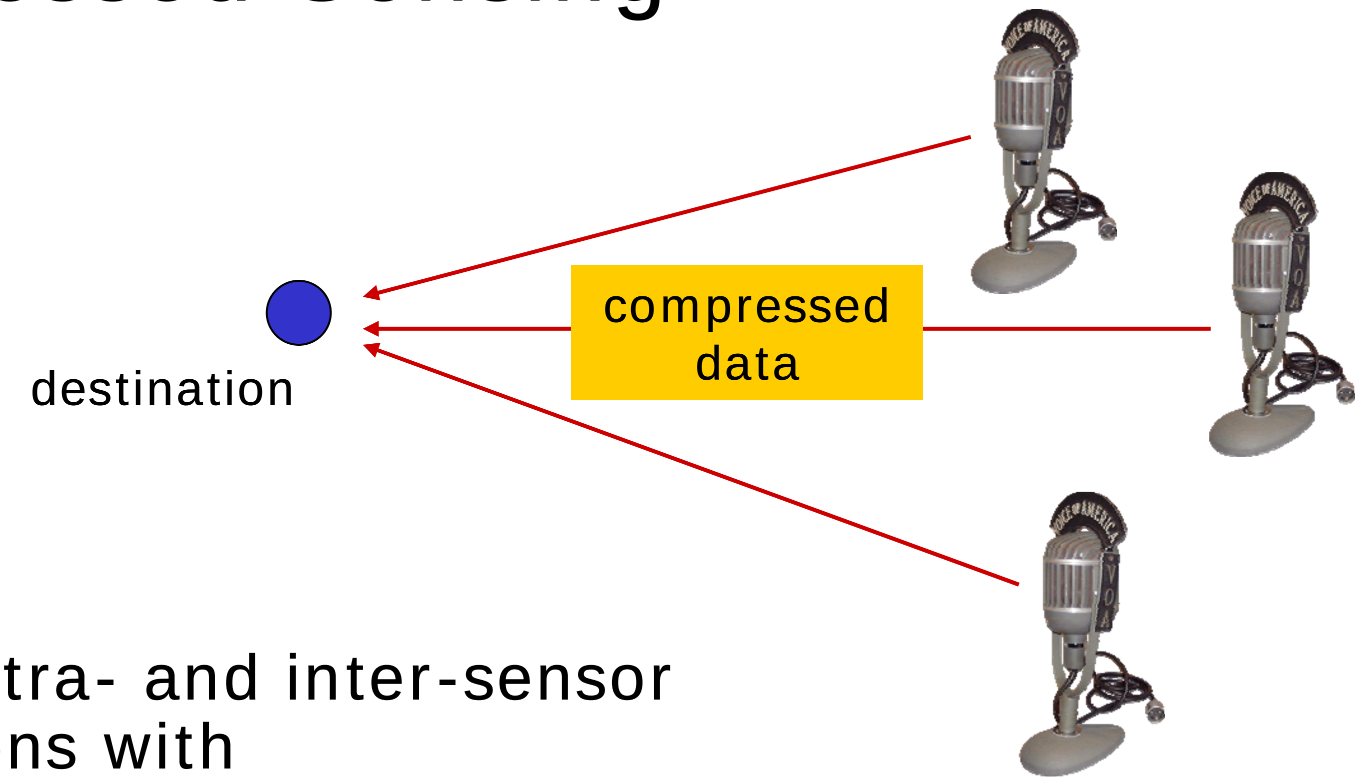
- Can we exploit *intra-sensor* and *inter-sensor* correlation to *jointly compress?*
- *Ongoing challenge* in information theory (distributed source coding)

Collaborative Sensing



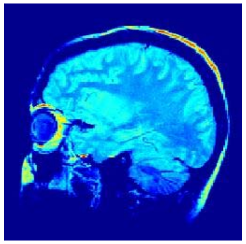
- Collaboration introduces
  - inter-sensor *communication overhead*
  - complexity at sensors

Distributed Compressed Sensing



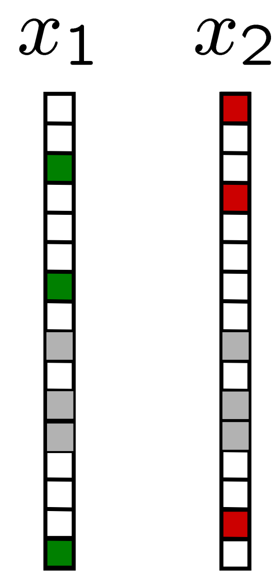
- Exploit intra- and inter-sensor correlations with
  - zero inter-sensor communication overhead
  - low complexity at sensors
- *Distributed source coding* via CS

Model 1:
Common + Innovations



Common + Innovations Model

- *Motivation:* measuring signals in smooth field
  - "average" temperature value common at multiple locations
  - "innovations" driven by wind, rain, clouds, etc.



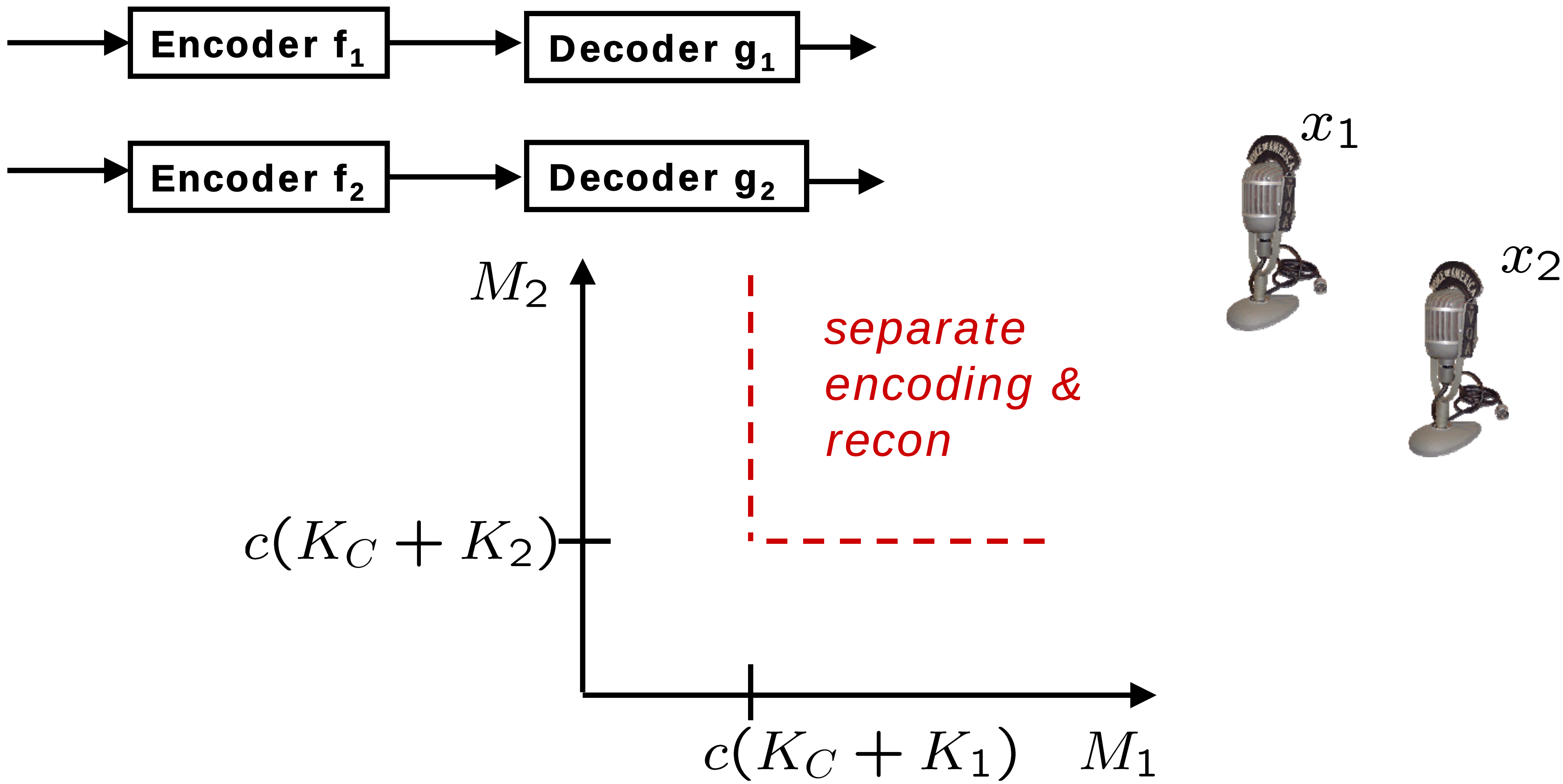
- *Joint sparsity model*:
  - length-N sequences x_1 and x_2

$$\begin{aligned} x_1 &= z_C + z_1 \\ x_2 &= z_C + z_2 \end{aligned}$$

  - z_C is length-N *common* component
  - z_1, z_2 length-N *innovations* components
  - z_C, z_1, z_2 have sparsity K_C, K_1, K_2

- Measurements
$$\begin{aligned} y_1 &= \Phi_1 x_1 \\ y_2 &= \Phi_2 x_2 \end{aligned}$$

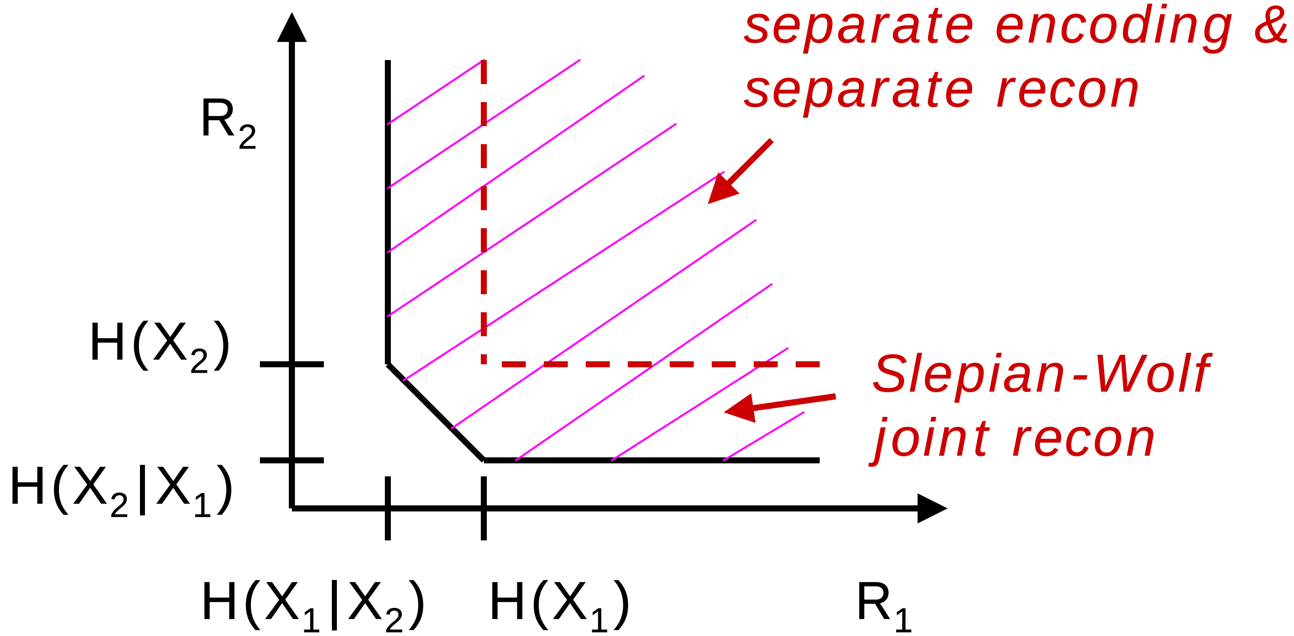
Measurement Rate Region with *Separate* Reconstruction



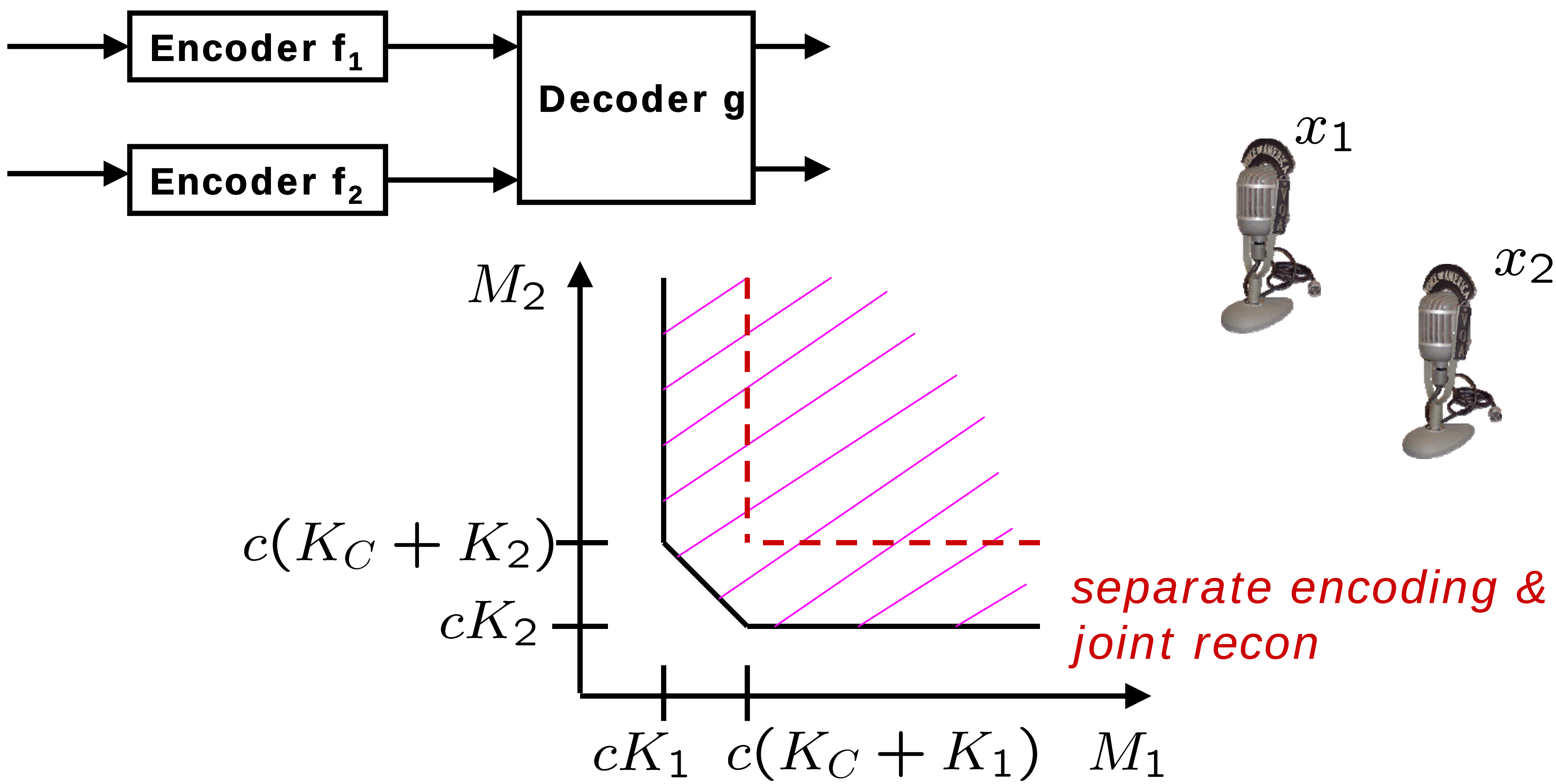
Slepian-Wolf Theorem
(Distributed lossless coding)

- **Theorem:** [Slepian and Wolf 1973]

$R_1 > H(X_1|X_2)$ (conditional entropy)
$R_2 > H(X_2|X_1)$ (conditional entropy)
$R_1+R_2 > H(X_1,X_2)$ (joint entropy)

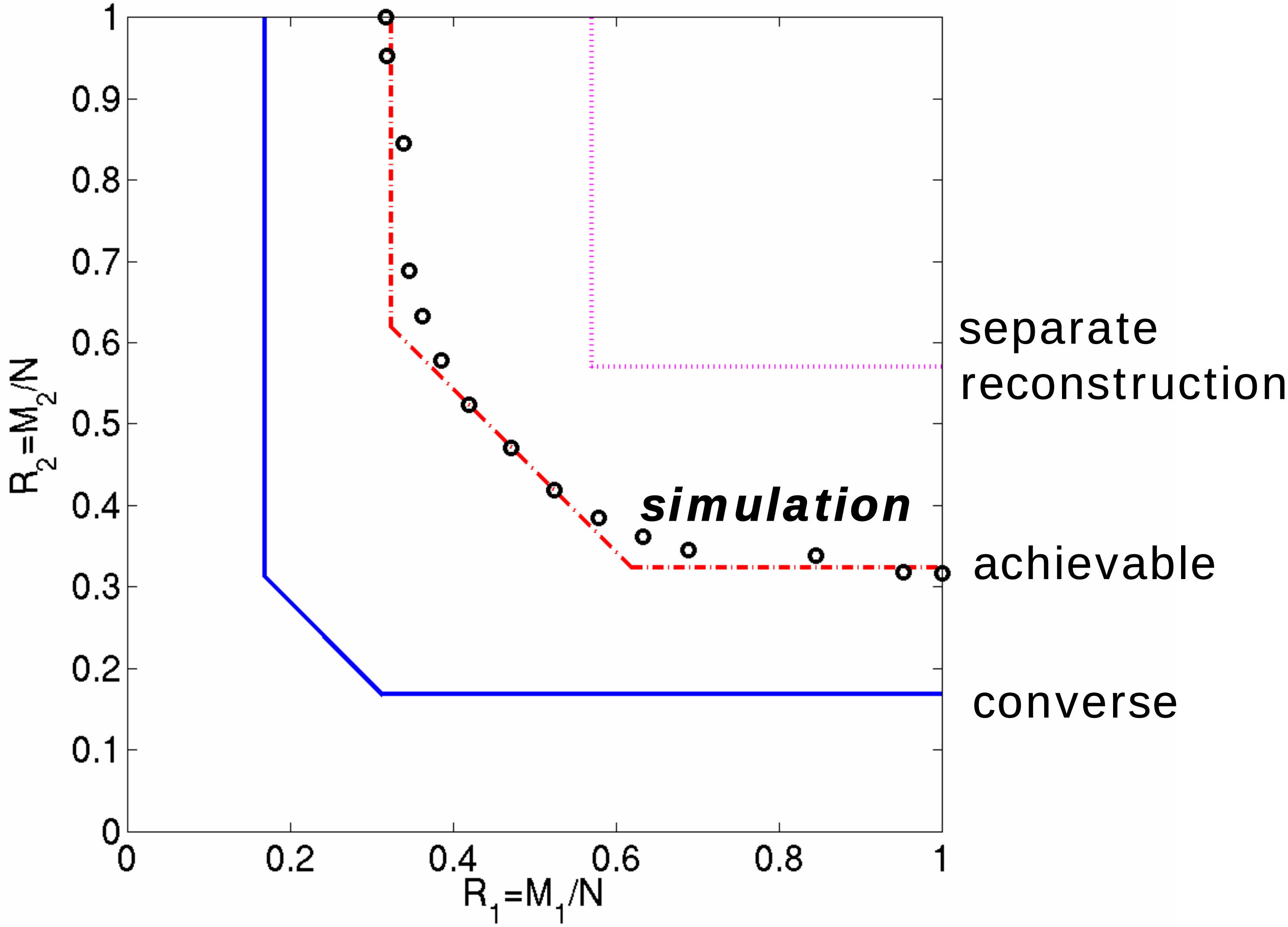


Measurement Rate Region with *Joint* Reconstruction

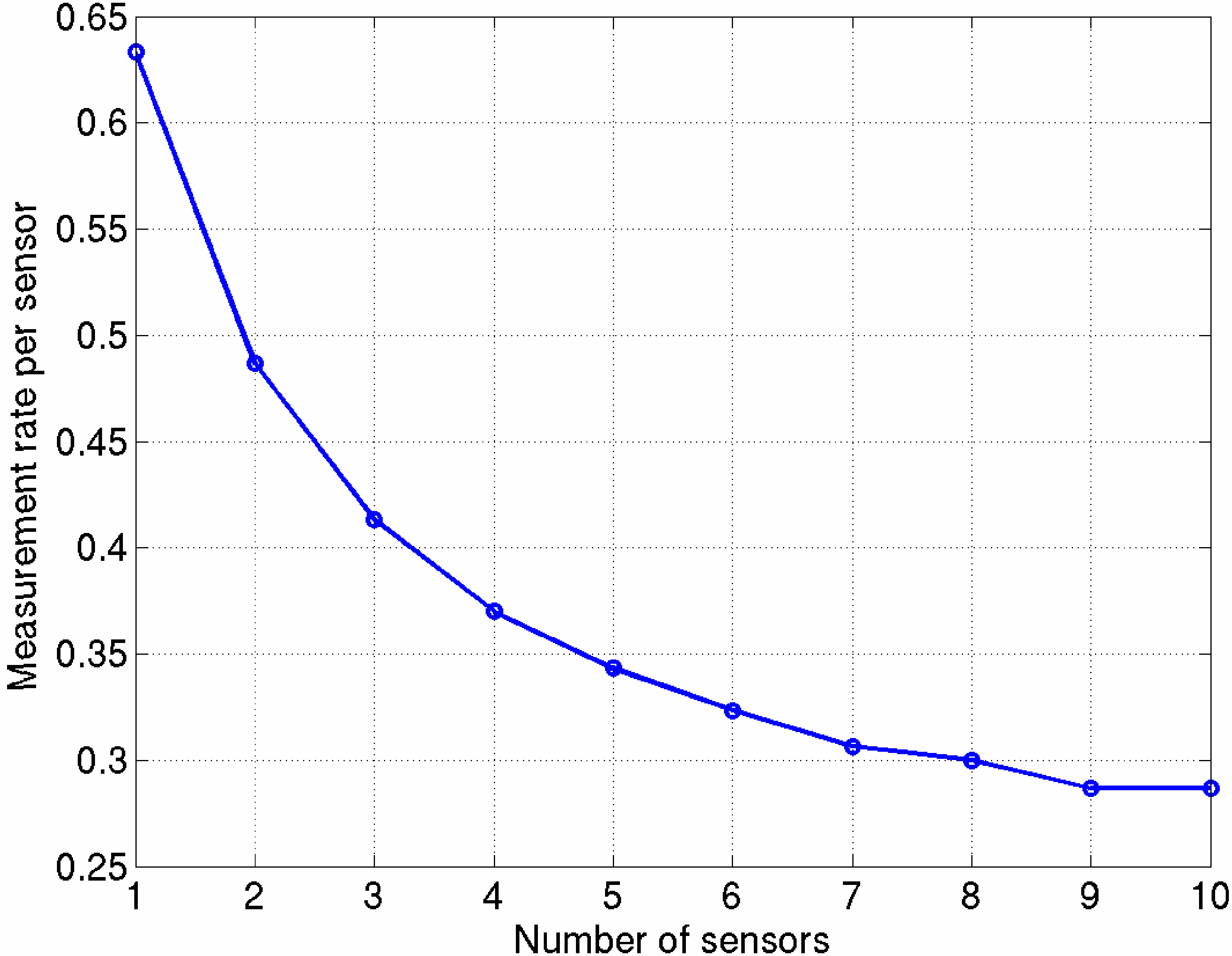


- Inspired by Slepian-Wolf coding

Measurement Rate Region [Baron et al.]



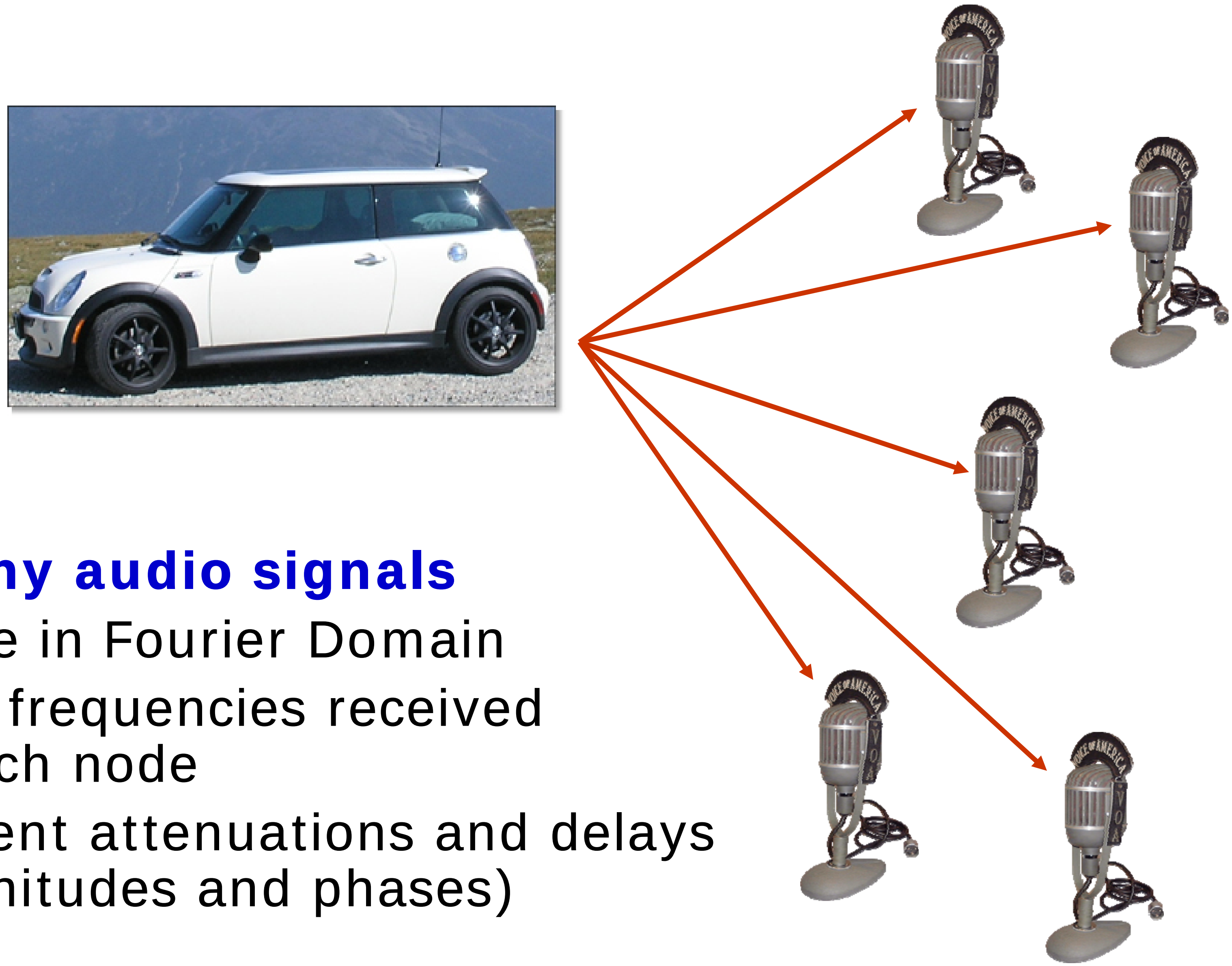
Multiple Sensors



Model 2:

Common Sparse Supports

Common Sparse Supports Model

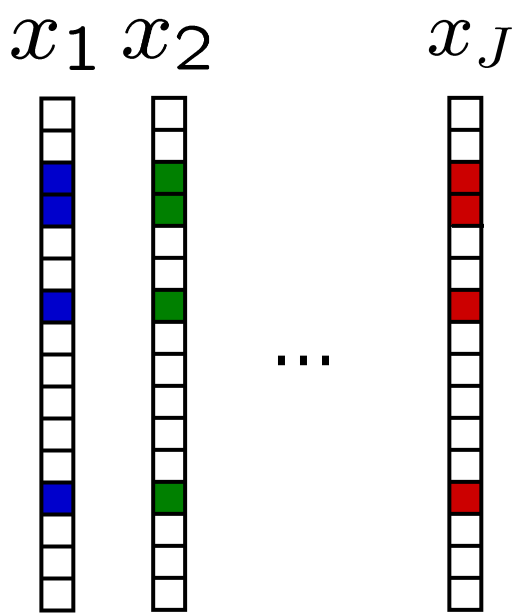


Ex: Many audio signals

- sparse in Fourier Domain
- same frequencies received by each node
- different attenuations and delays (magnitudes and phases)

Common Sparse Supports Model

- *Signals share sparse components but different coefficients*



- Intuition: Each measurement vector holds clues about coefficient support set

Required Number of Measurements

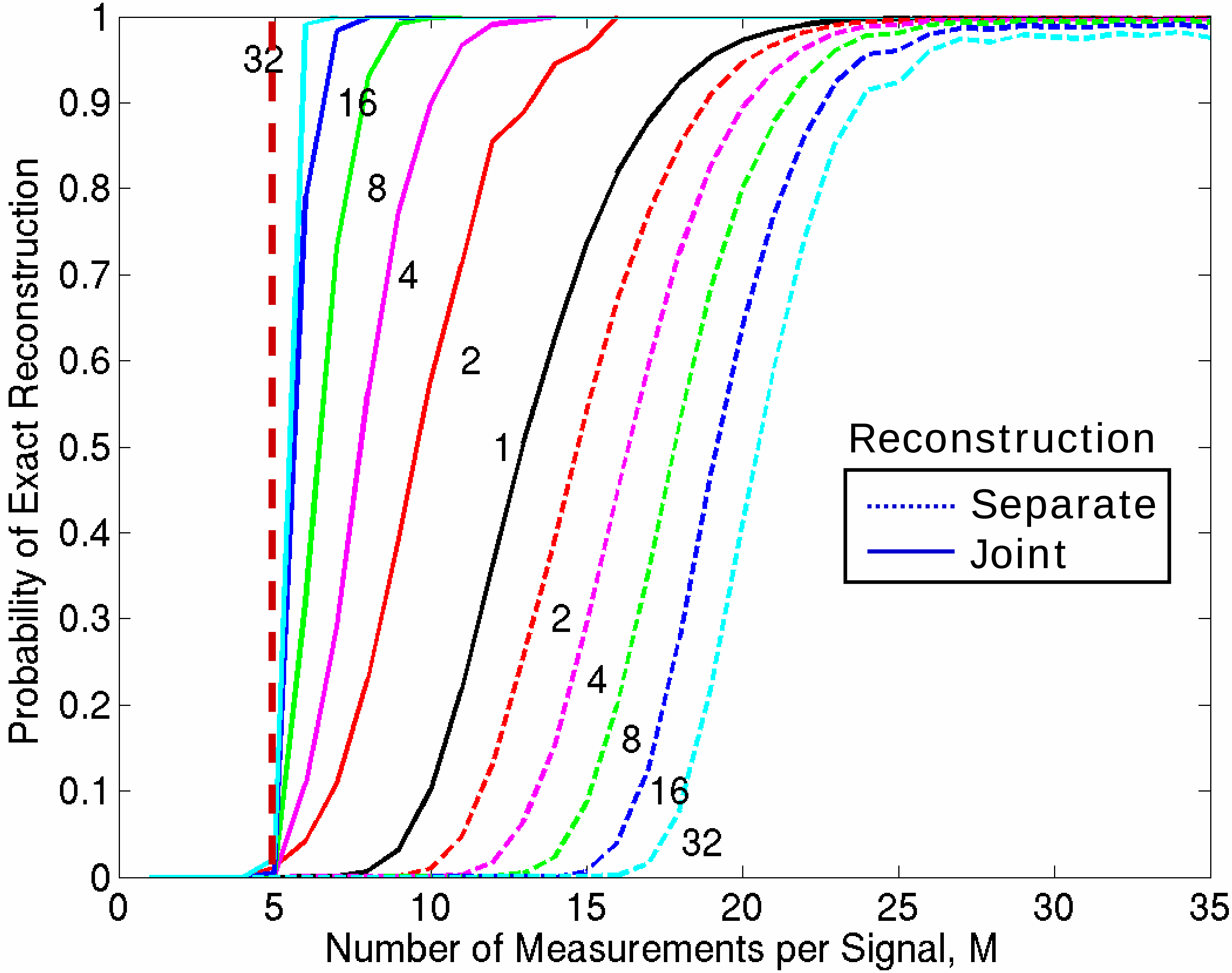
[Baron et al. 2005]

- **Theorem:** M=K measurements per sensor do not suffice to reconstruct signal ensemble
- **Theorem:** As number of sensors J increases, M=K+1 measurements suffice to reconstruct
- *Joint reconstruction with reasonable computational complexity*

***K* = 5**

N = 50

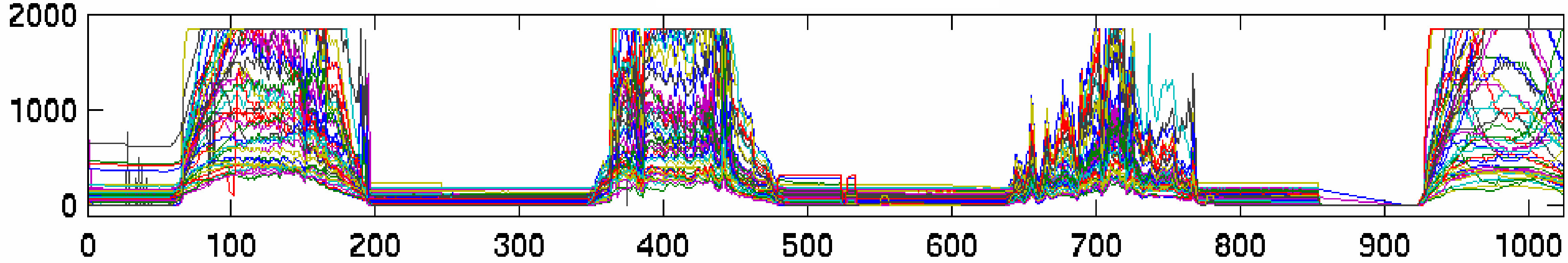
Results for Common Sparse Supports



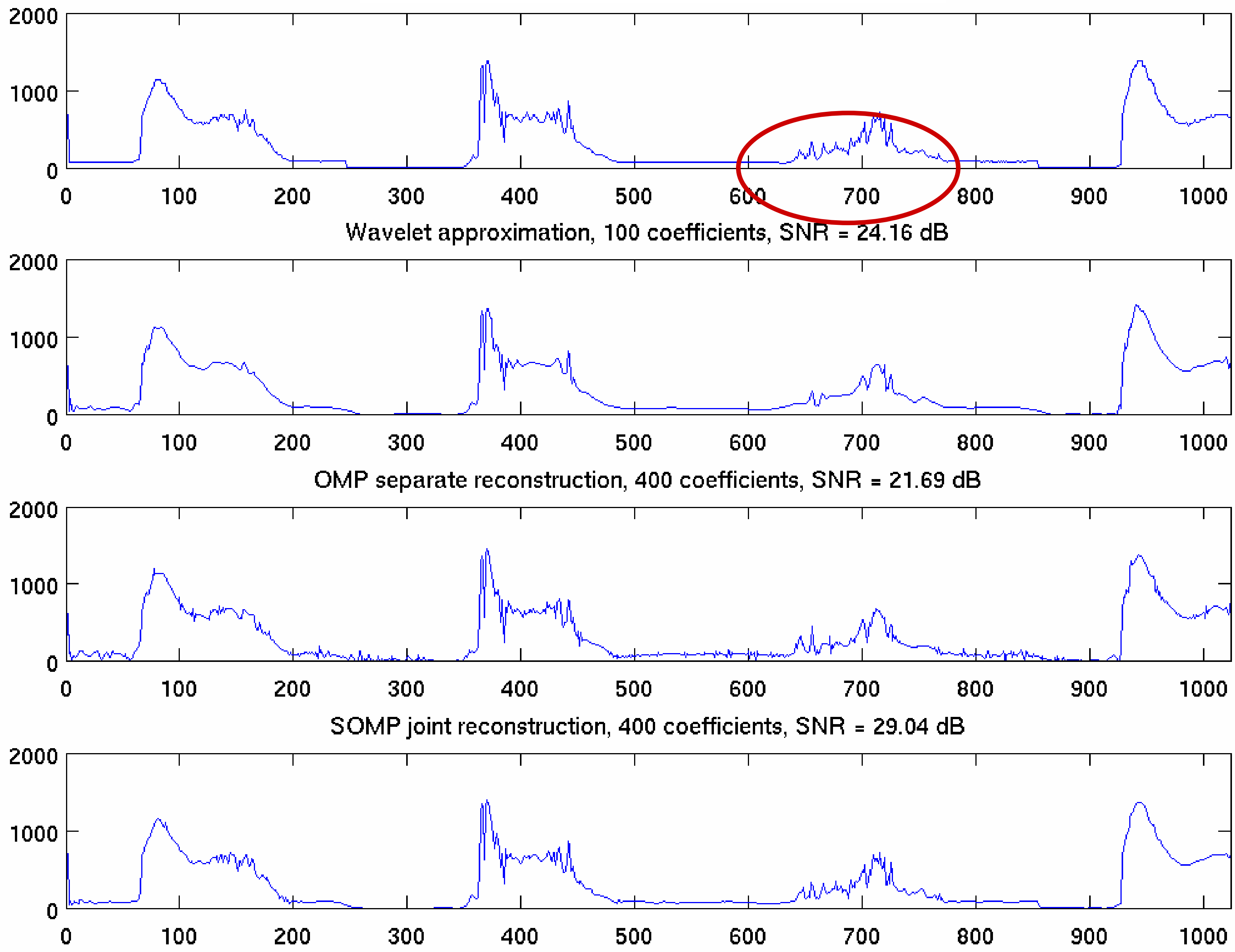
Real Data Example

- Light levels in Intel Berkeley Lab
- 49 sensors, 1024 samples each
- Compare:
  - wavelet approx — 100 terms per sensor
  - *separate CS* — 400 measurements per sensor
  - *joint CS* (SOMP) — 400 measurements per sensor

- Correlated signal ensemble



Light Intensity at Node 19



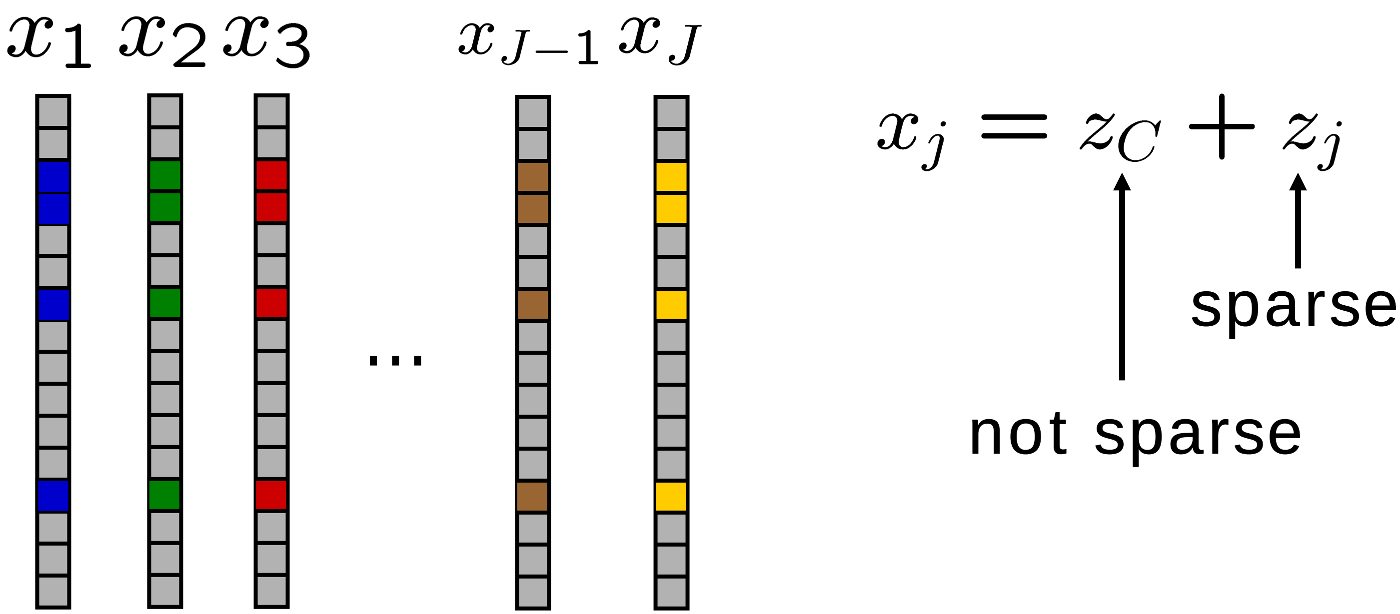
Model 3:
Non-Sparse Common Component

Non-Sparse Common Model

- *Motivation:* non-sparse video frame + sparse motion
- Length-N common component z_C is non-sparse

⇒ *Each signal is incompressible*

- Innovation sequences z_j may share supports

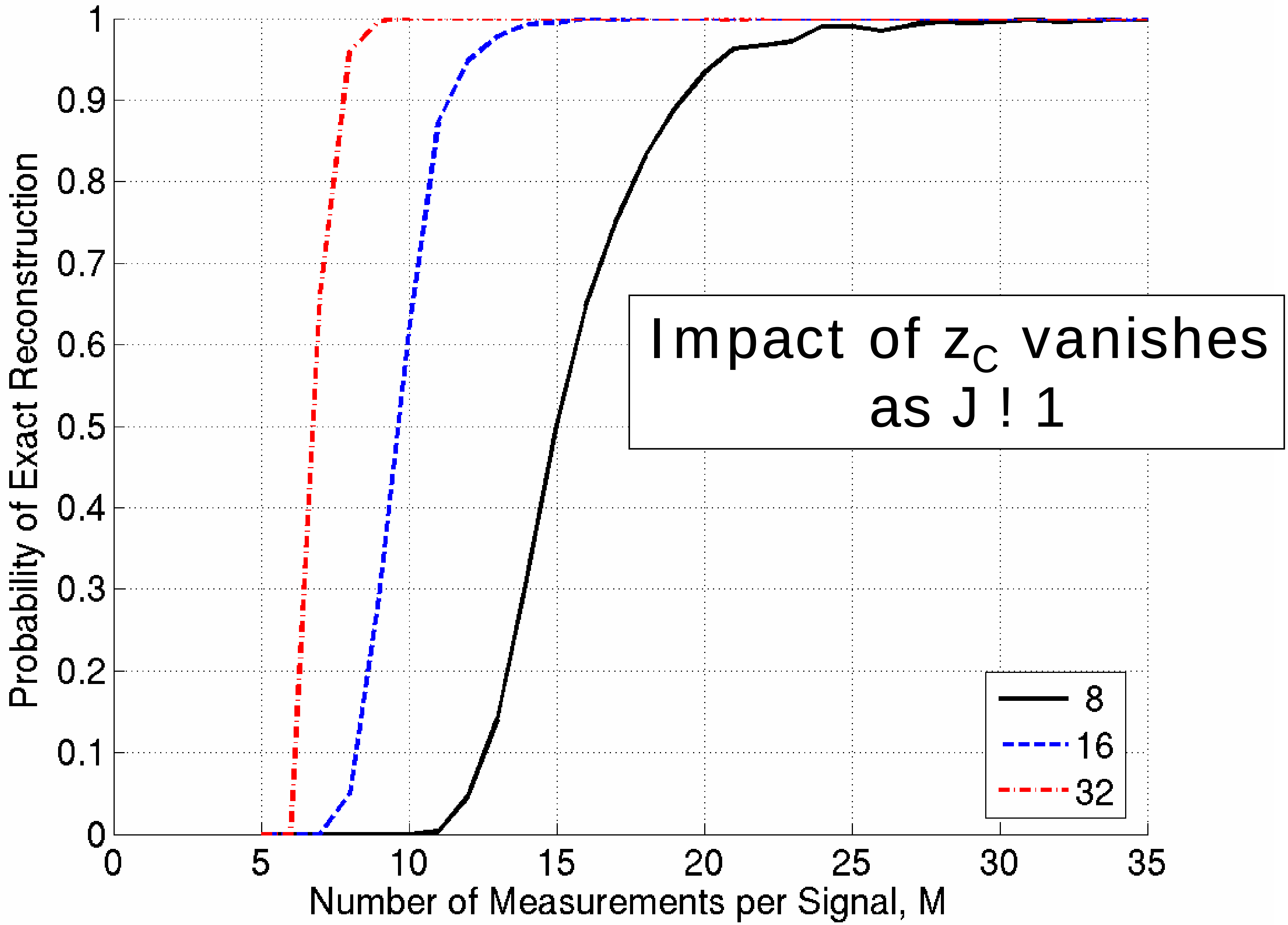


- *Intuition*: each measurement vector contains clues about common component z_C

$K=5$
$N=50$

Results for Non-Sparse Common

(same supports)



Summary

- **Compressed Sensing**
  - "random projections"
  - process sparse signals using far fewer measurements
  - universality and information scalability
- Determination of measurement rates in CS
  - measurements are bits
  - lower bound on measurement rate
    - direct relationship to rate-distortion content
- Promising results with LDPC measurement matrices
- Distributed CS
  - new models for joint sparsity
  - analogy with Slepian-Wolf coding from information theory
  - compression of sources w/ intra- and inter-sensor correlation
- ***Much potential and much more to be done***
- ***Compressed sensing meets information theory***

dsp.rice.edu/cs

THE END

"With High Probability"