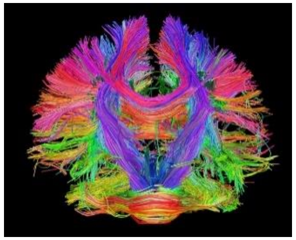



CS-570
Statistical Signal Processing

Lecture 10: Quantization & sampling

Spring Semester 2019

Grigorios Tsagkatakis

Matrix Completion (MC)

Relaxation

$$\min\{\ \|\mathbf{M}\|_* \ :\ \ \mathcal{A}(\mathbf{X}) = \mathcal{A}(\mathbf{M})\}$$

Performance

$$\left\|M - M^*\right\|_F^2 \le 4\sqrt{\frac{(2+p)\min(n_1, n_2)}{p}}\delta + 2\delta,$$

$$\text{where } p = \text{fraction of known entries} = \frac{m}{n_1 n_2} = \frac{|\Omega|}{n_1 n_2}$$

Noisy case

$$\min\{\ \|\mathbf{M}\|_* \ :\ \ \|\mathcal{A}(\mathbf{X}) - \mathcal{A}(\mathbf{M})\|_F^2 \le \epsilon\}$$

CS and MC

	Sparse recovery	*Rank minimization*
Unknown	Vector *x*	Matrix *A*
Observations	$y = Ax$	$y = L[A]$ (linear map)
Combinatorial objective	$\#\{\mathbf{x}_i \neq 0\} \quad = \quad \|\mathbf{x}\|_0$	$\mathrm{rank}(A) \quad = \quad \#\{\sigma_i(A) \neq 0\}$ $= \quad \|\sigma(A)\|_0$
Convex relaxation	$\|\mathbf{x}\|_1 = \sum_i \lvert\mathbf{x}_i\rvert$	$\|A\|_* = \sum_i \sigma_i(A)$
Algorithmic tools	Linear programming	Semidefinite programming

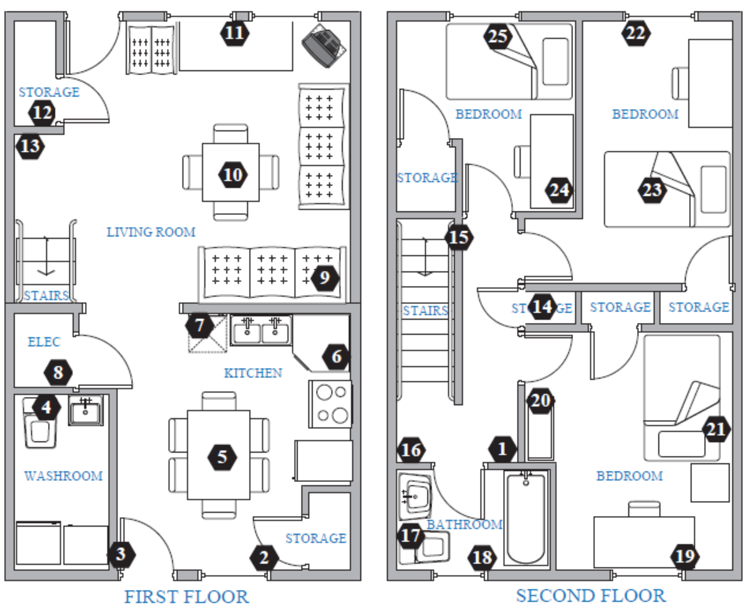
Yi Ma et al, "Matrix Extensions to Sparse Recovery", CVPR2009

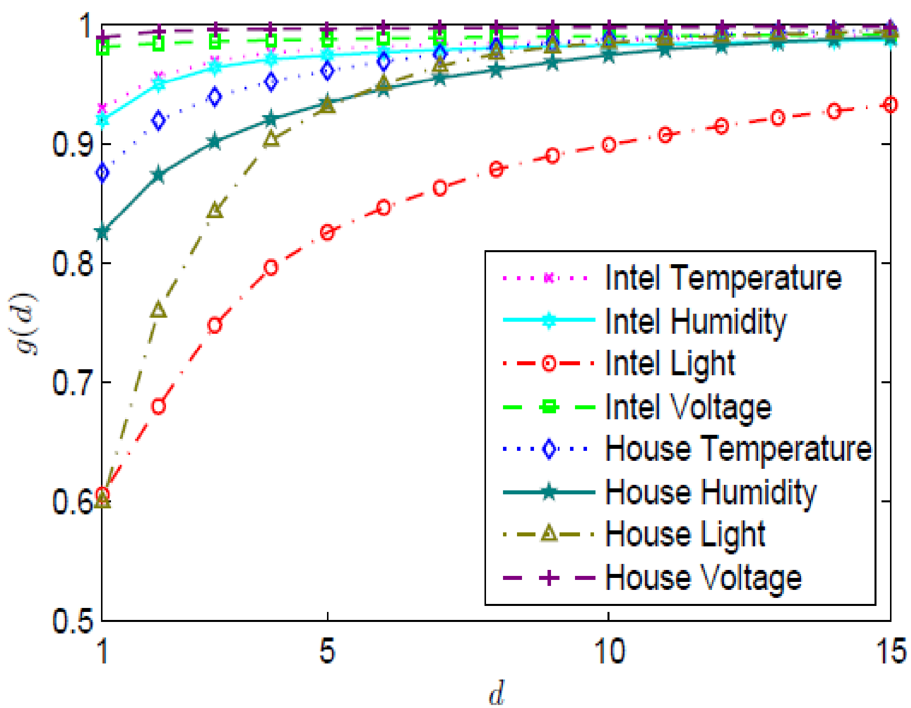
Applications of MC

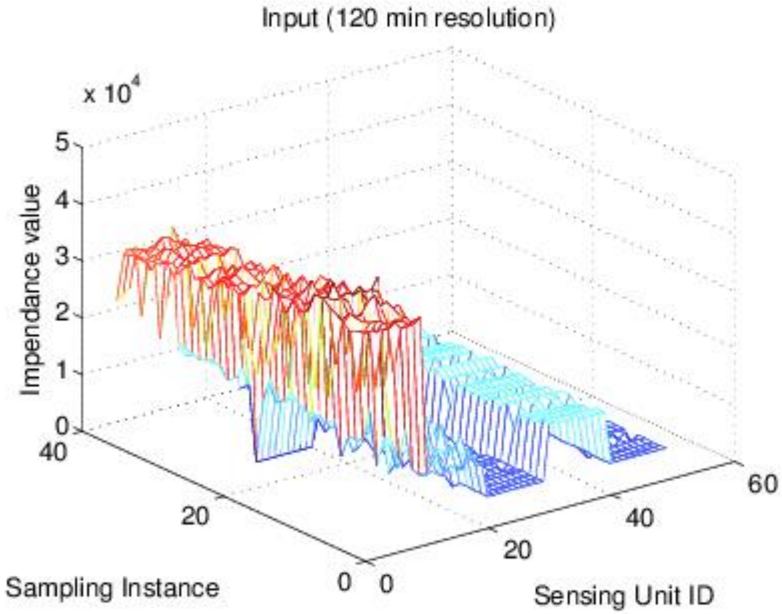
- Recommendation systems
  - Matrix (user, preference/quality/intention)
- Sensor localization
  - Matrix (location, physical quantity)
- Data recovery in Wireless Sensor Networks
  - Matrix (sensor, time)

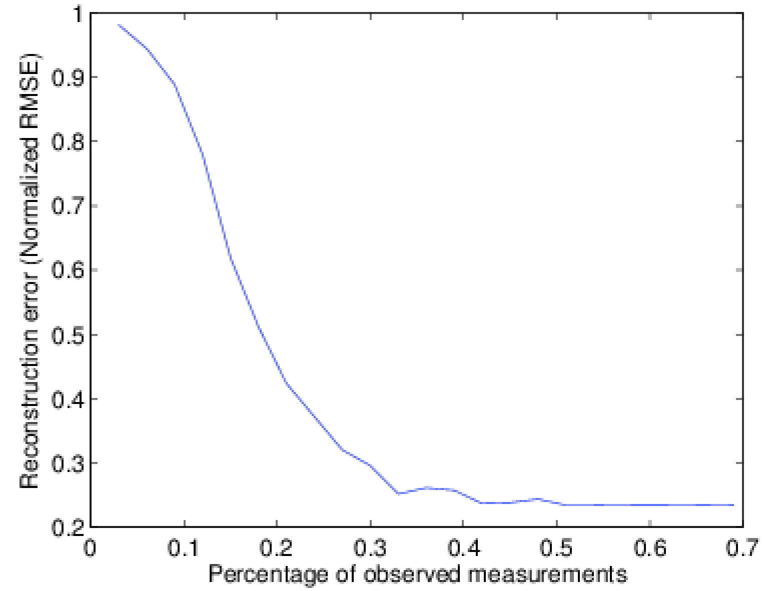
Data Gathering

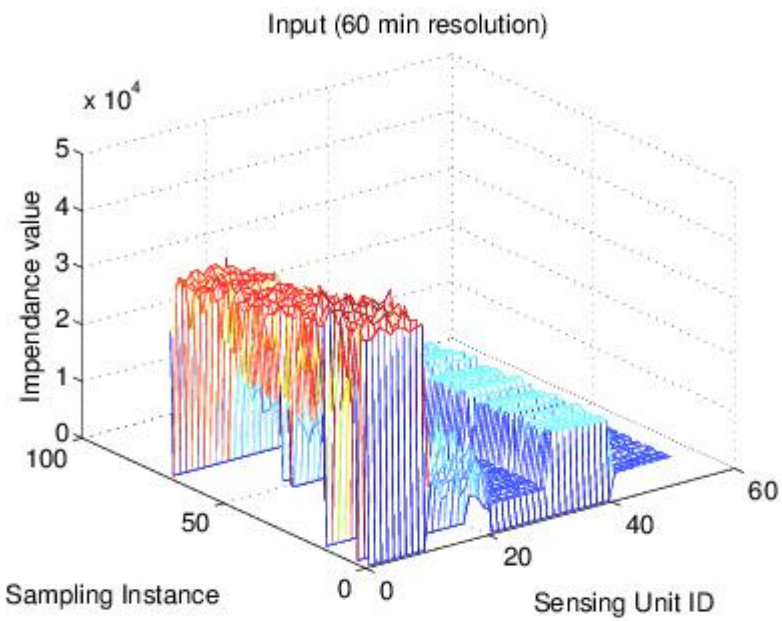
- STCDG: An Efficient Data Gathering Algorithm Based on Matrix Completion for Wireless Sensor Networks

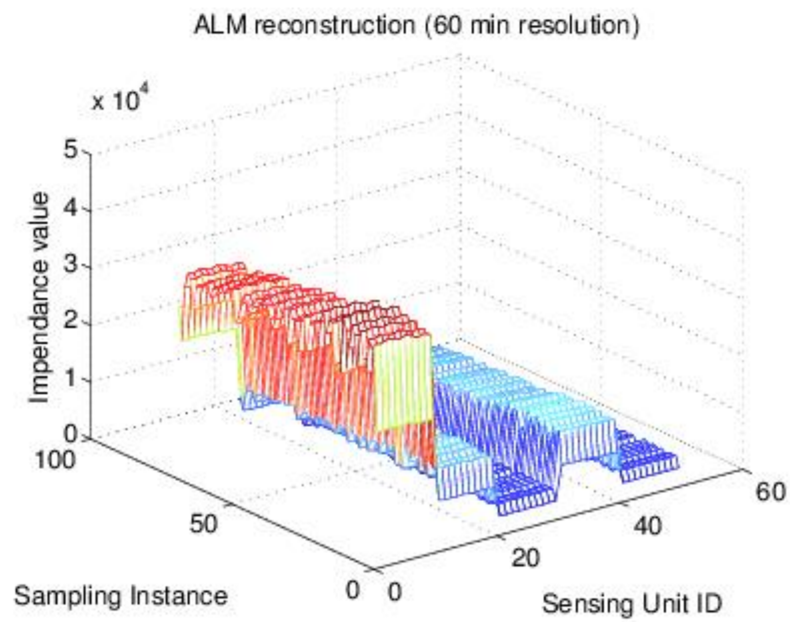




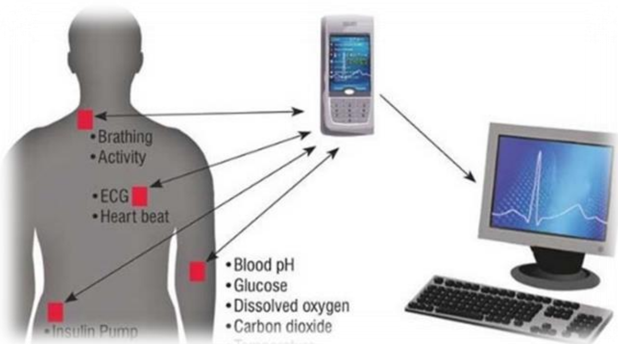

Input (120 min resolution)
x 10^4
Impendance value
Sampling Instance
Sensing Unit ID


Reconstruction error (Normalized RMSE)
Percentage of observed measurements


Input (60 min resolution)
x 10^4
Impendance value
Sampling Instance
Sensing Unit ID


ALM reconstruction (60 min resolution)
x 10^4
Impendance value
Sampling Instance
Sensing Unit ID

WSNs for Human Activity Recognition



Elderly assistance

Chronic conditions management

HAR

Health monitoring

Acting upon alerting events

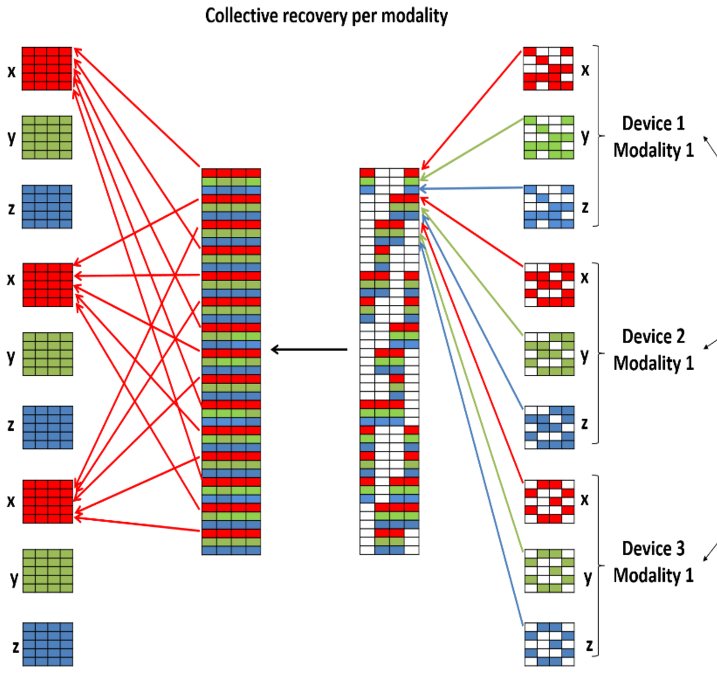
Fitness coaching

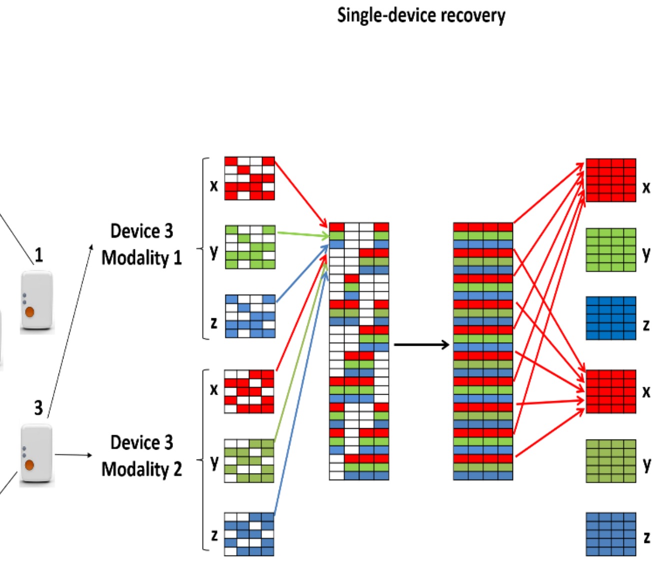
FORTH
Institute of Computer Science

Single-device vs collective recovery: matrices

Scenario 2
Collective per modality

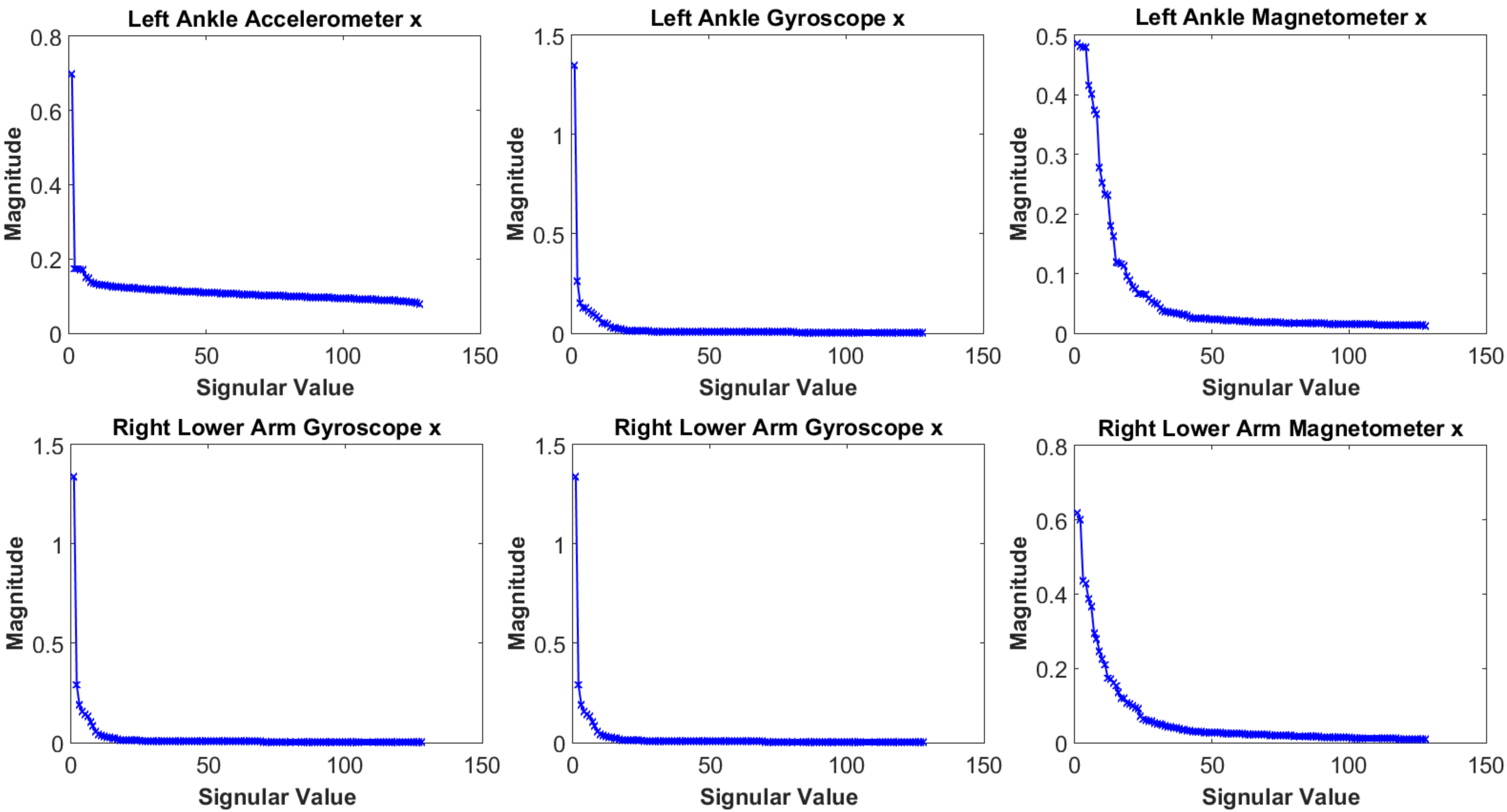
Scenario 1
Single-device





Scenario 3: Overall collective recovery structured similarly

Body sensor network



RTT estimation

Decentralized Matrix Factorization by Stochastic Gradient Descent (DMFSGD),

Estimation of end-to-end network distances

- Network nodes exchange messages with each other
- Each node collects and processes local measurements

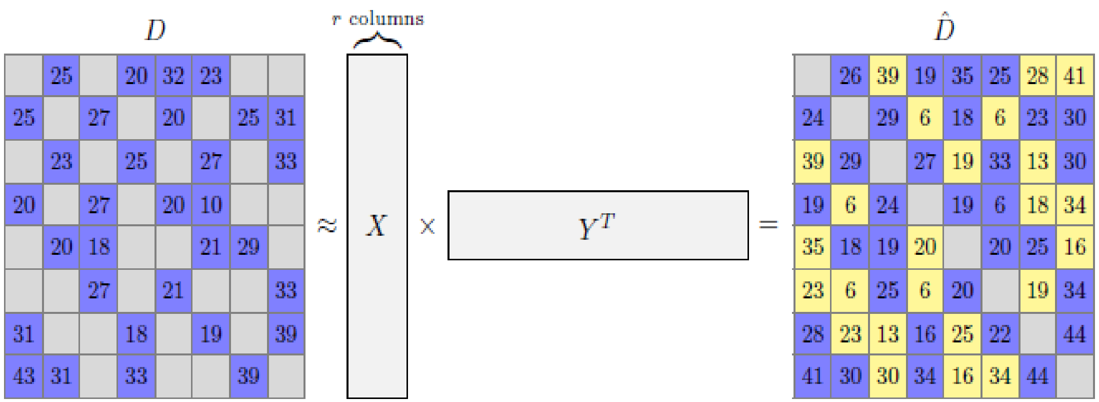


Fig. 2. Network distance prediction by matrix factorization. Note that the diagonal entries of D and $\hat{D}$ are empty.

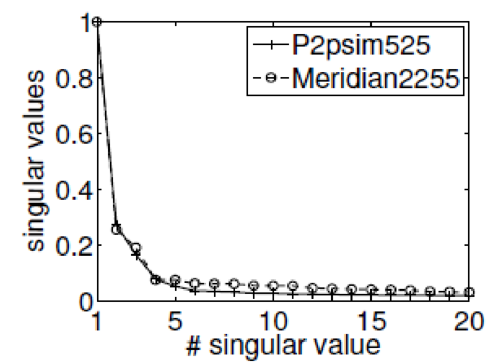
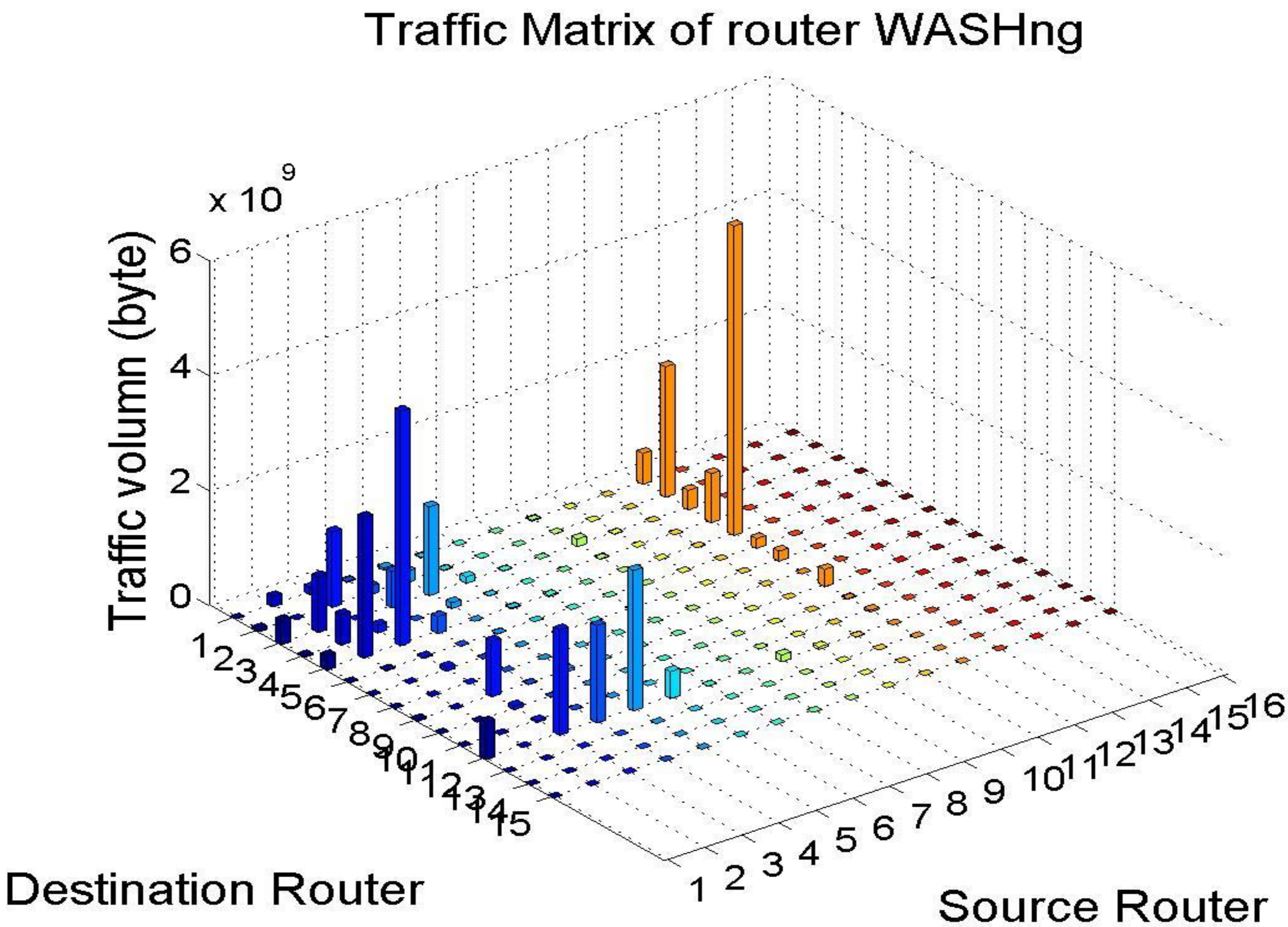
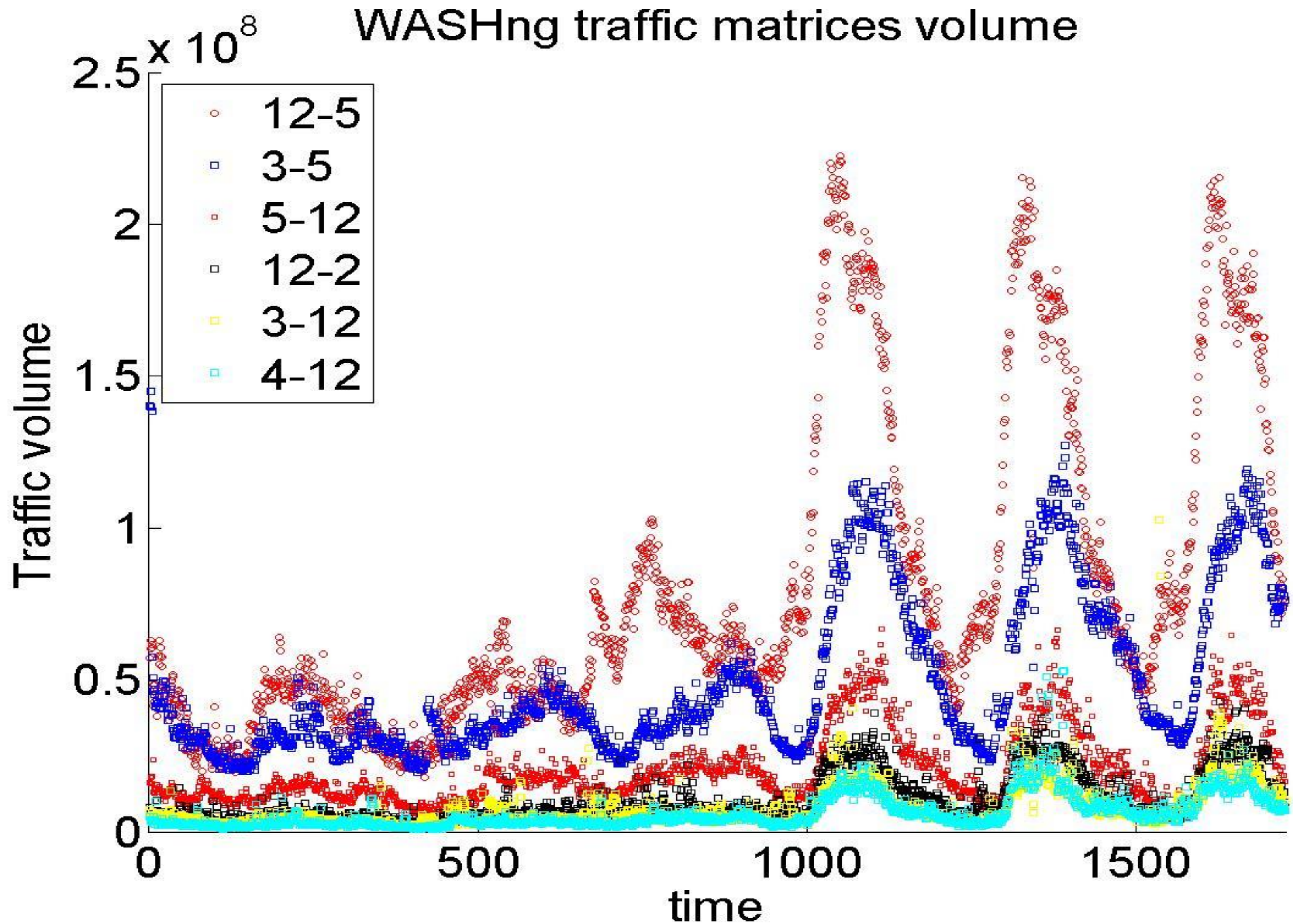
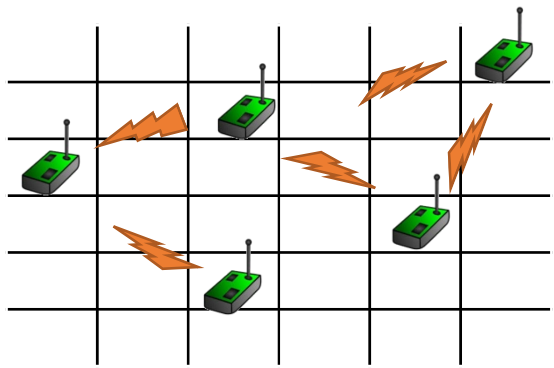


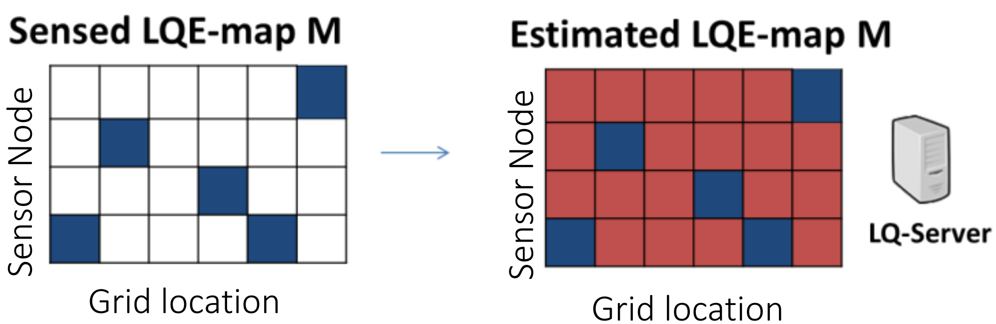
Fig. 3. The singular values of a RTT matrix of 2255×2255, extracted from the Meridian dataset [30] and called "Meridian2255", and of a RTT matrix of 525×525, extracted from the P2psim dataset [30] and called "P2psim525". The singular values are normalized so that the largest singular values of both matrices are equal to 1.


Traffic Matrix of router WASHng
x 10^9
Traffic volumn (byte)
0
2
4
6
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
Destination Router
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Source Router

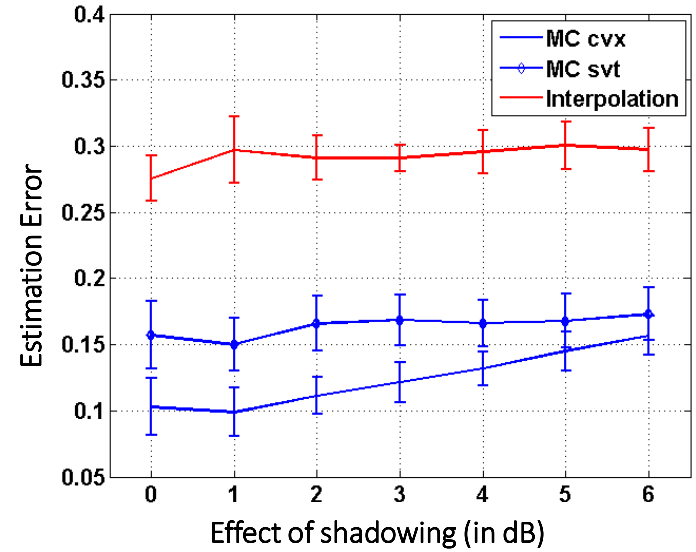

WASHng traffic matrices volume
x 10^8
2.5
2
1.5
1
0.5
0
12-5
3-5
5-12
12-2
3-12
4-12
Traffic volume
0
500
1000
1500
time

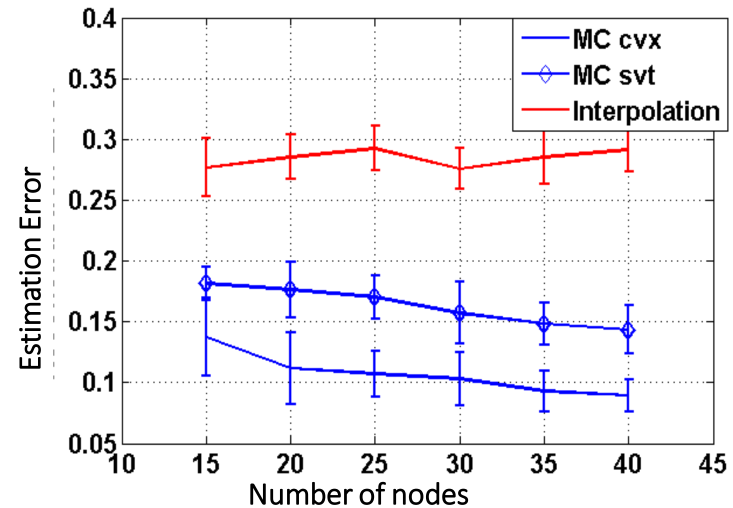
LQM Estimation



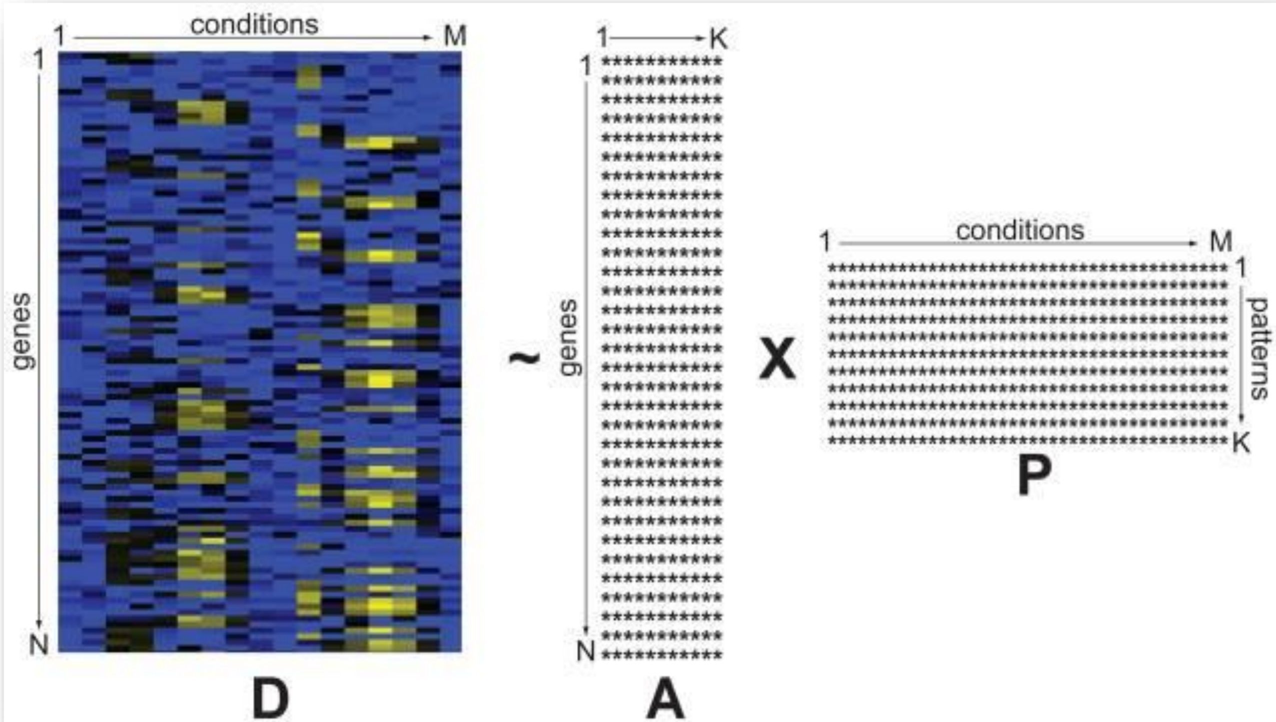


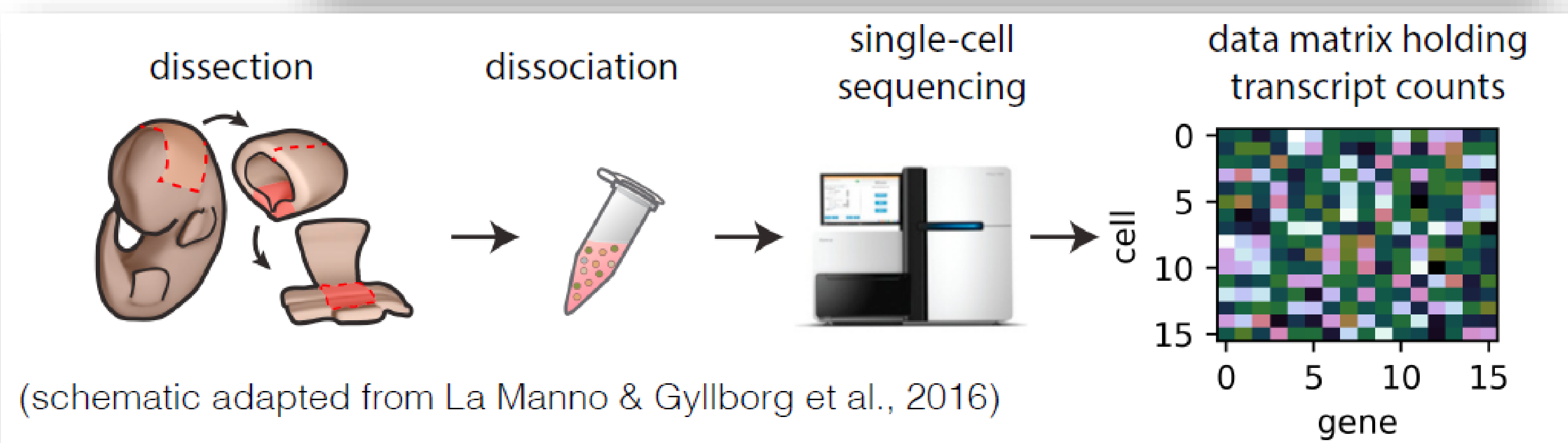
- Dataset
  - Testbed @ FORTH (144m^2, 1x1m grid)
  - RSSI values (channel quality)
  - 13 IEEE802.11b/g channels





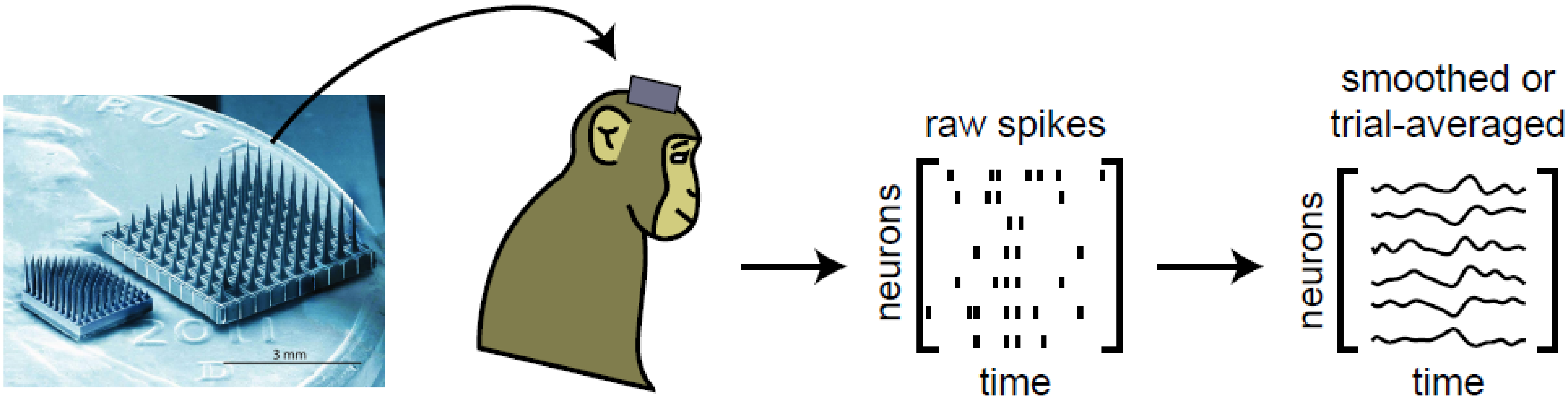
Microarray data

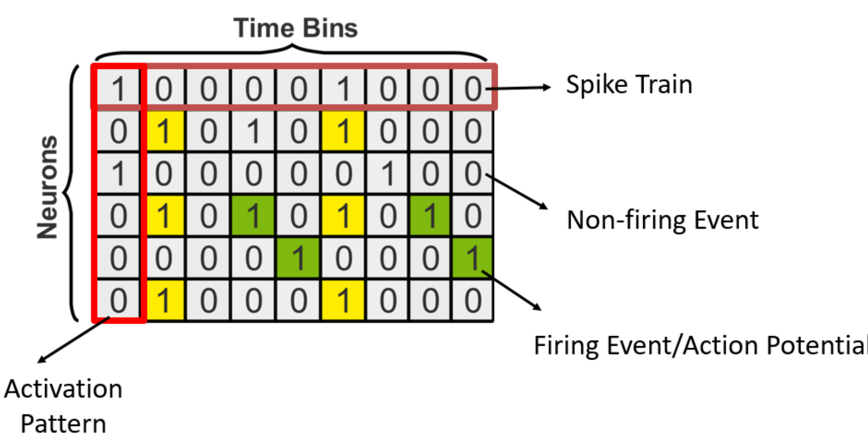


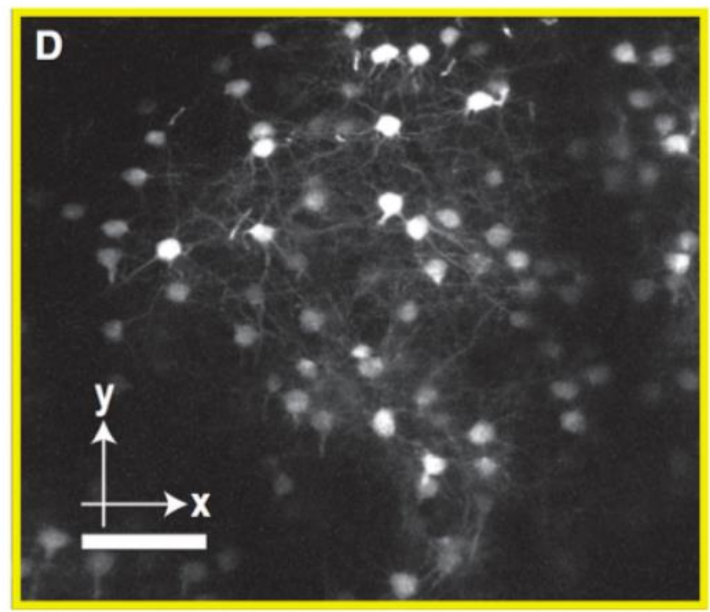


(schematic adapted from La Manno & Gyllborg et al., 2016)

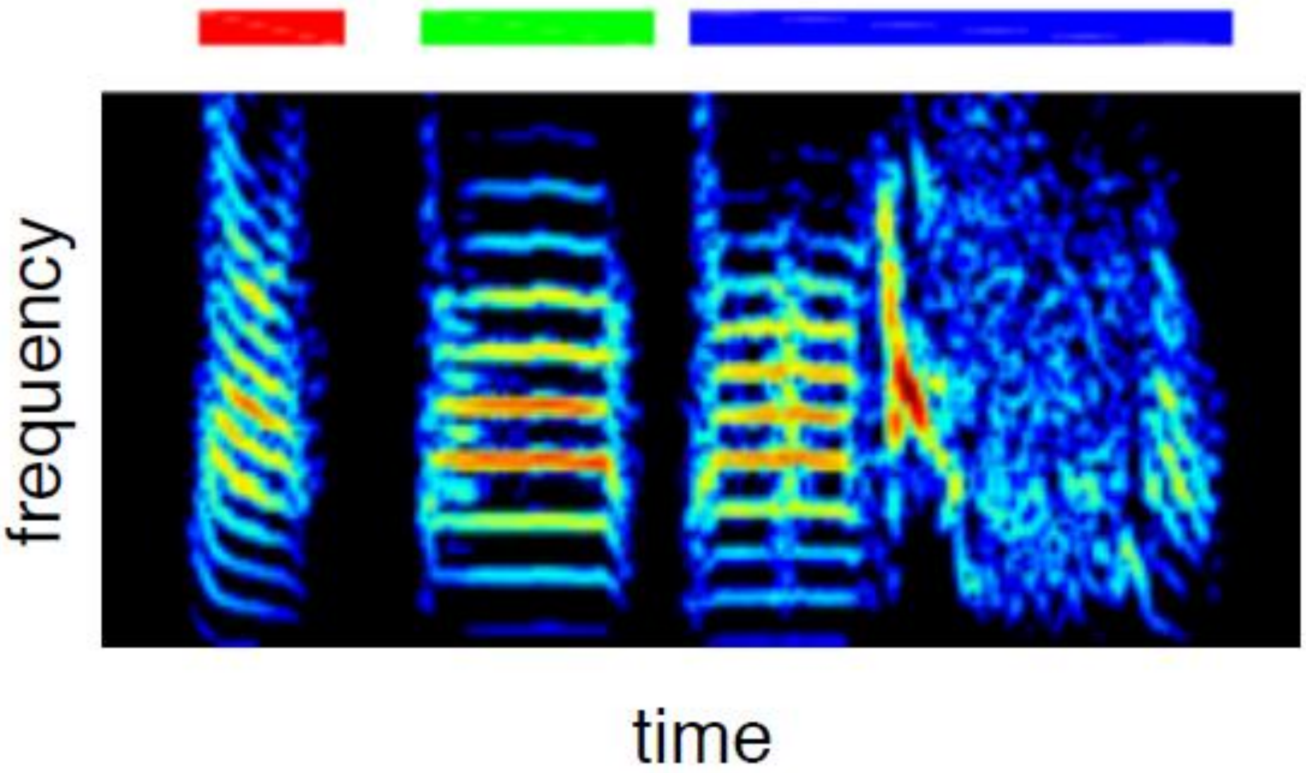
Neural Activity







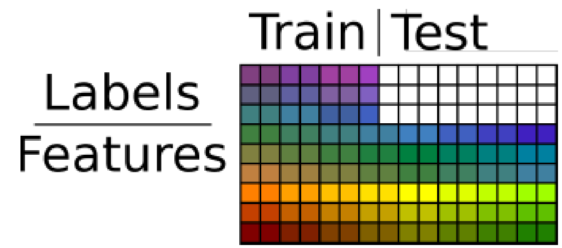
Audio



Transductive Classification

- Unknown labels and corrupted or missing training features and corresponding labels leading to incomplete matrix

$$\mathbf{Z} = \begin{bmatrix} \mathbf{Y}_{tr} & \mathbf{Y}_{tst} \\ \mathbf{X}_{tr} & \mathbf{X}_{tst} \end{bmatrix} \Rightarrow \begin{bmatrix} \mathbf{Y}_{tr} & \mathbf{?} \\ \mathbf{X}_{tr} & \mathbf{X}_{tst} \end{bmatrix}$$



- Formulate classification as rank minimization

$$\begin{aligned} \min \quad & \mu \|\mathbf{Z}\|_* + l_X(\mathbf{Z_X}) + \lambda\, l_Y(\mathbf{Z_Y}) \\ \text{s.t.} \quad & \mathbf{Z_X} \geq \mathbf{0} \ \text{ and } \ \mathbf{Z_1} = \mathbf{1}^{\mathrm{T}} \end{aligned}$$

where:

$$l_X(\mathbf{Z_X}) = \sum_{ij \in \mathbf{Z_X}} (z_{ij} - z_{0ij})^2$$

$$l_Y(\mathbf{Z_Y}) = \sum_{ij \in \mathbf{Y}_{train}} \frac{1}{\gamma} log(1 + exp(-\gamma z_{ij} z_{0ij}))$$

Multi-label image classification

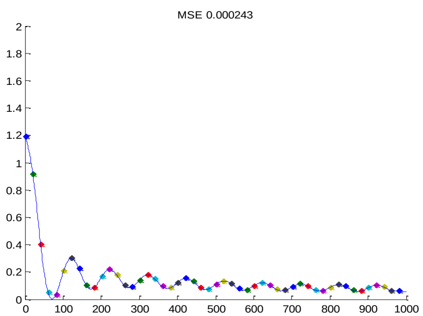
$$\mathbf{Z}^{obs} = \mathbf{Z} + \mathbf{E} = \left[\begin{array}{ccccc|ccc|cccc} 0 & 0 & 1 & \dots & 0 & ? & \dots & ? & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & \dots & 0 & ? & \dots & ? & 0 & 1 & 0 & 0 \\ 1 & ? & 1 & \dots & 0 & ? & \dots & ? & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & \dots & 0 & ? & \dots & ? & 0 & 0 & 0 & 1 \\ \mathbf{h}_1^{tr} & \mathbf{h}_2^{tr} & \mathbf{h}_3^{tr} & \dots & \mathbf{h}_N^{tr} & \mathbf{h}_1^{tst} & \dots & \mathbf{h}_M^{tst} & ? & ? & ? & ? \end{array}\right]$$

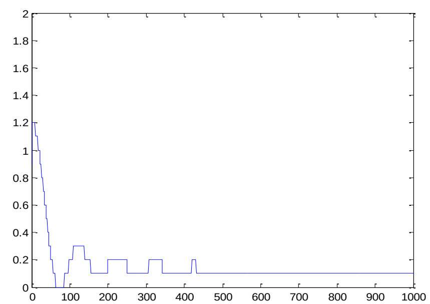
(a) Training (b) Testing (c) Class Representatives

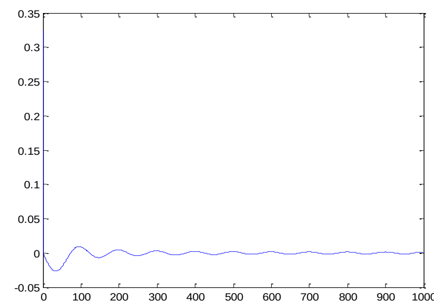
Analog-to-digital transformation

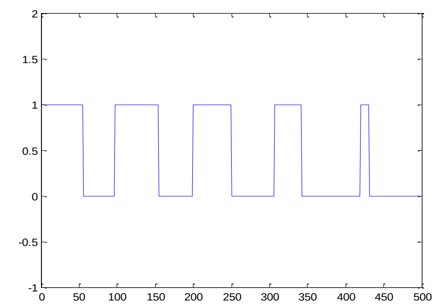
Signal Sensing

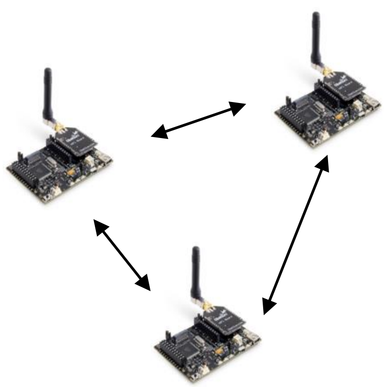
Sample	Quantize	Process	Transmit









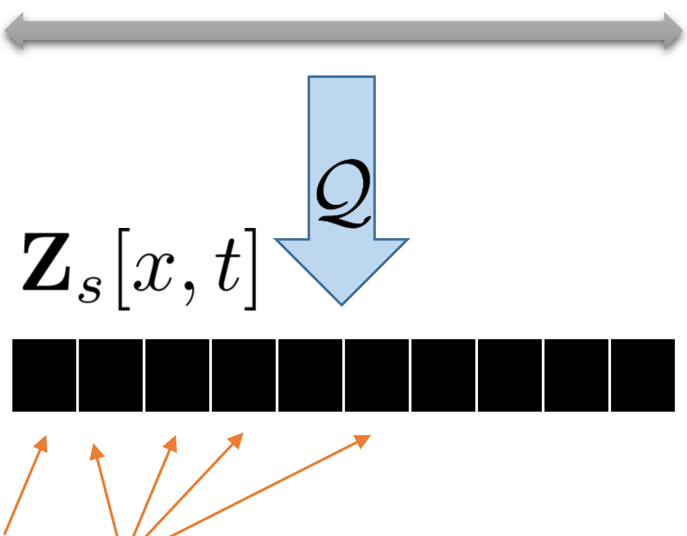


Signal Quantization

Map values to symbols

- Bits per measurements

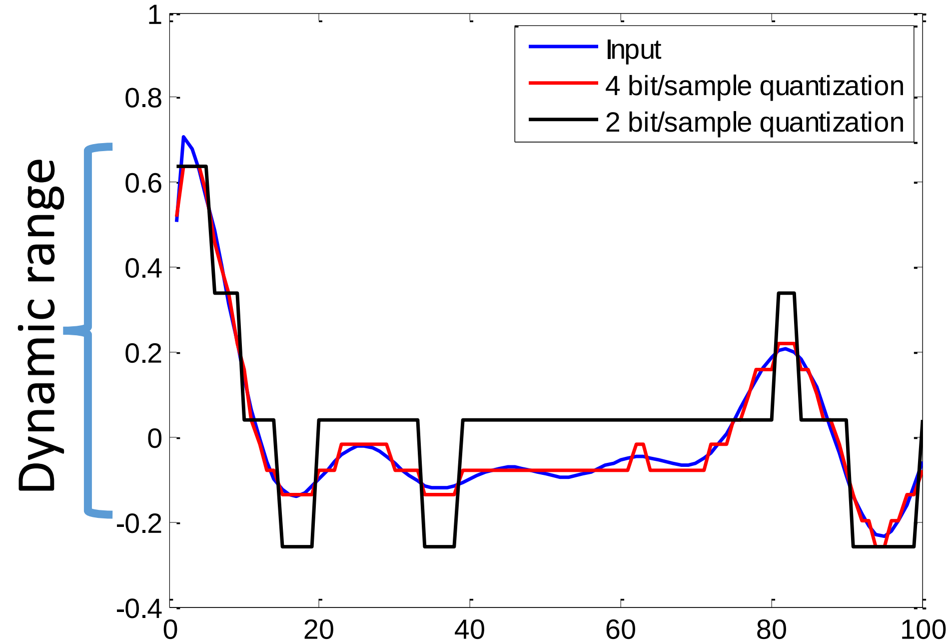
$$\mathbf{Y}_s[x,t]$$



$$\mathcal{Q}$$

$$\mathbf{Z}_s[x,t]$$

$$|\mathcal{C}| = \mathcal{B} = 2^R$$

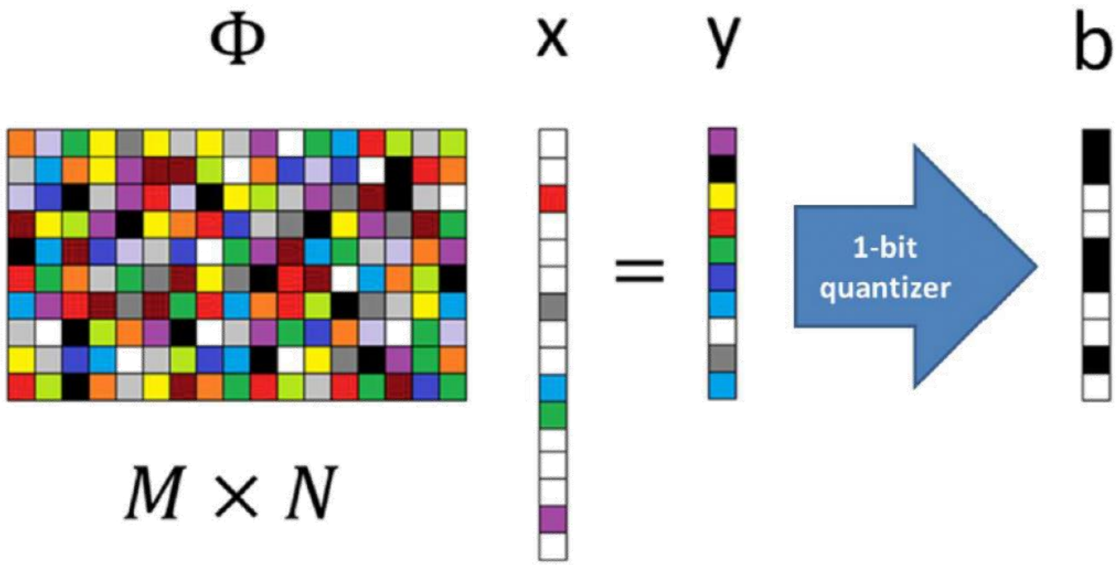


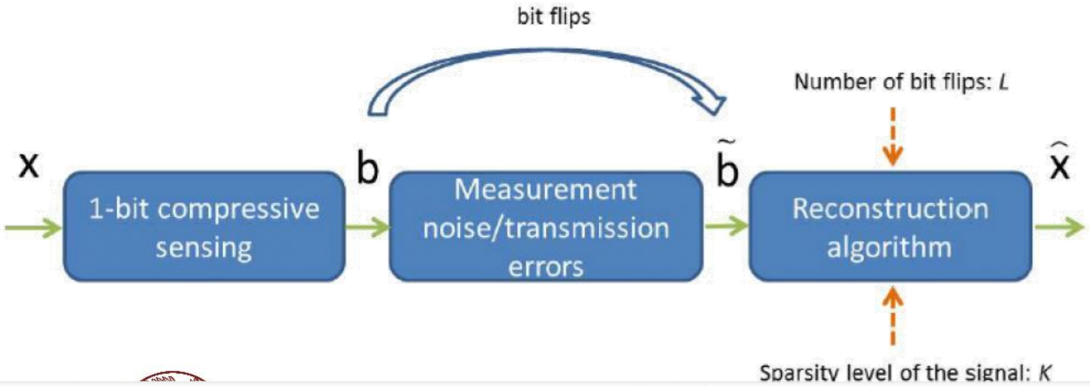
$$\mathcal{T} = \{t_1, ..t_{\mathcal{B}} | t_i < t_j, \forall i, j\}$$

$$Q(x) = \Delta \cdot \left\lfloor \frac{x}{\Delta} + \frac{1}{2} \right\rfloor = \Delta \cdot \text{floor}\left(\frac{x}{\Delta} + \frac{1}{2}\right)$$

FORTH
Institute of Computer Science

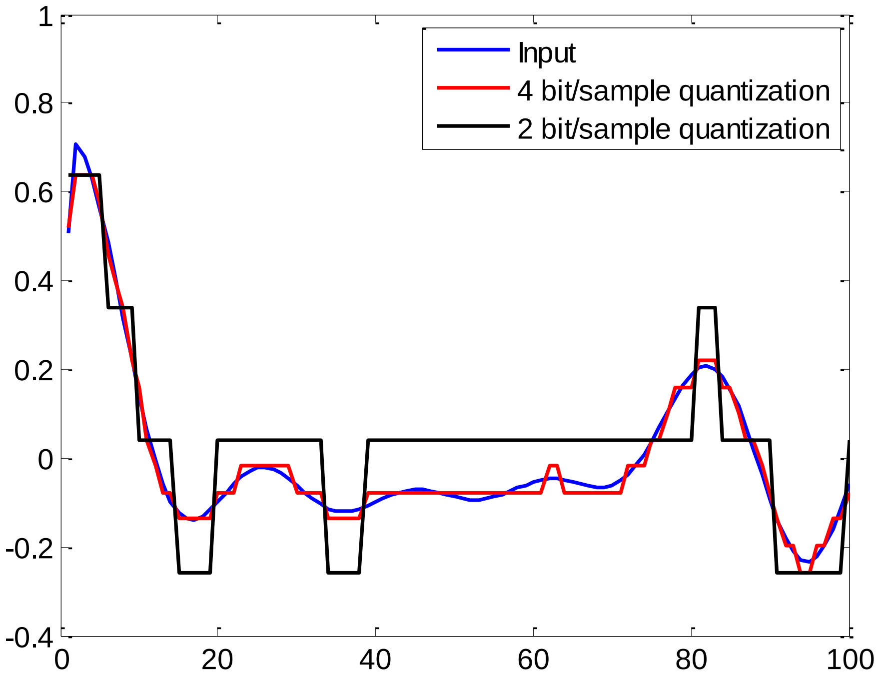
1-Bit Compressive Sensing





End-to-End CS imaging

- Difference aspects of sensing process
  - Dictionary
  - Noise sources
  - **Quantization**
    - ADC
    - Storage
    - Transmission
  - Entropy Coding



(Scalar) Signal Quantization

Map from real to indexed $\mathcal{Q} : \mathbb{R} \rightarrow 2^R$

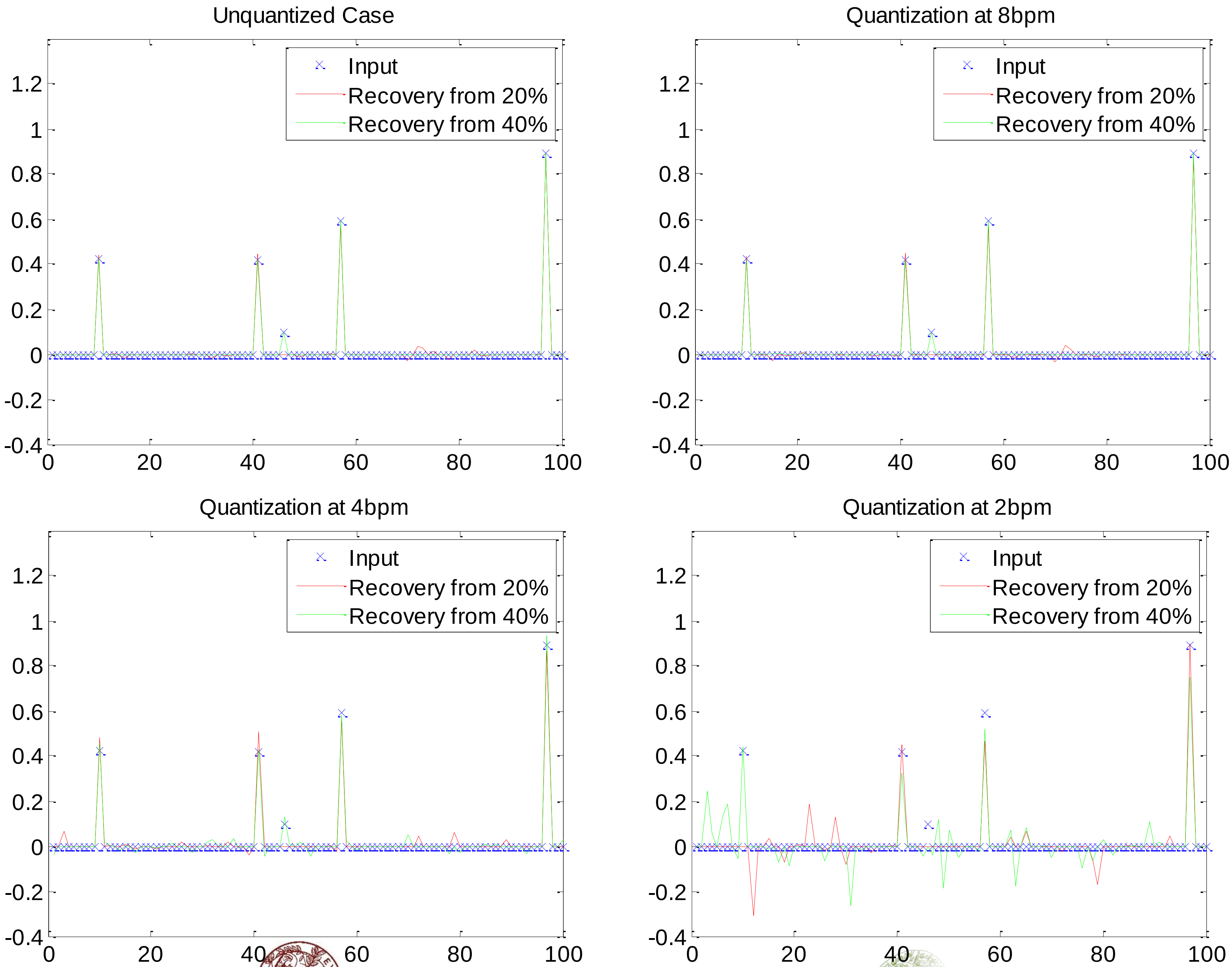
- Codebook $|\mathcal{C}| = \mathcal{B} = 2^R$
- Thresholds $\mathcal{T} = \{t_1, ..t_{\mathcal{B}} | t_i < t_j, \forall i, j\}$
- Dynamic range $M = |\max(x) - \min(x)|$
- Saturation $\hat{f} = \min(f, M)$

Uniform scalar

Non-uniform scalar: Lloyd-Max conditions

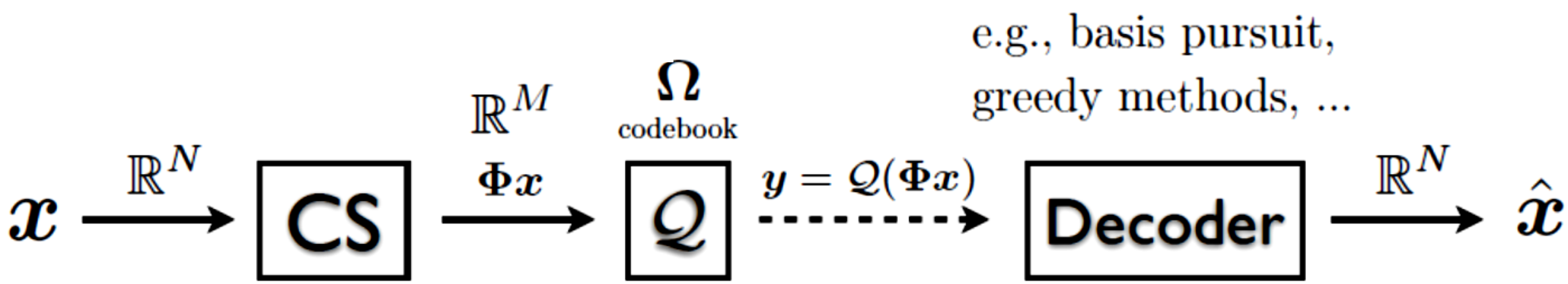
$$|t_j - t_i| = \Delta \qquad n = \log_2 M$$

$$t_j = \frac{1}{2}(\mathbb{E}\{X|X \in C_j\} + \mathbb{E}\{X|X \in C_{j+1}\})$$


Unquantized Case
Quantization at 8bpm
Quantization at 4bpm
Quantization at 2bpm
Input
Recovery from 20%
Recovery from 40%

Quantizing Compressed Sensing

- In a perfect noiseless system



Finite codebook $\Rightarrow \hat{\boldsymbol{x}} \neq \boldsymbol{x}$

i.e., impossibility to encode continuous domain
in a finite number of elements.

Objective: Minimize $\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|$

given a certain number of:
bits, measurements, or bits/meas.

Addressing Quantization in CS

- Operating regimes
  - Measurement Compression
  - Quantization Compression

➢ High resolution model (noise source)

$$\Delta << \|\mathbf{s}\|_2 \longrightarrow \Delta \approx \epsilon_Q : \min \|\mathbf{s}\|_1 \;\; \text{s.t.} \;\; \|\mathbf{y} - \mathbf{\Phi s}\|_2 \leq \epsilon_Q$$

➢ 1-bit CS $\mathbf{y} = \text{sign}(\mathbf{\Phi s})$

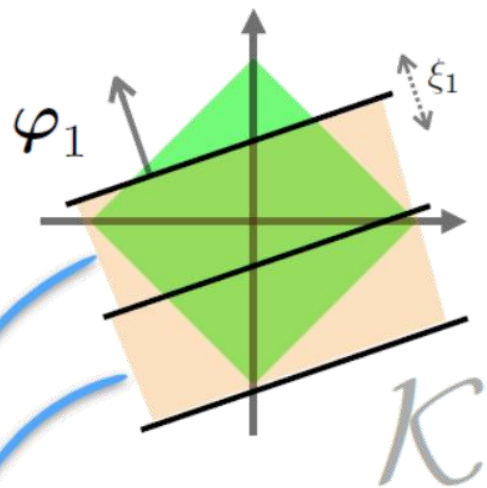
J. Laska, P. Boufounos, M. Davenport, and R. Baraniuk, “Democracy in action: Quantization, saturation, and compressive sensing," Applied and Computational Harmonic Analysis , 2011.

J.N. Laska, and R.G. Baraniuk. "Regime change: Bit-depth versus measurement-rate in compressive sensing." *Signal Processing, IEEE Transactions on Signal Processing* , 2012.

L. Jacques, J.N. Laska, P.T. Boufounos, and R.G. Baraniuk, “Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors," IEEE Transactions on Information Theory, 2013.

Properties of $\mathsf{A}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi x} + \boldsymbol{\xi})$

- 1. For consistent reconstruction method



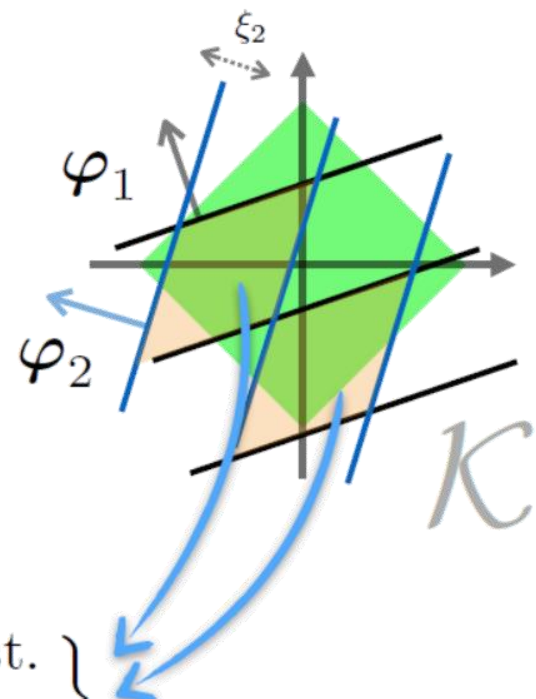
$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\varphi}_1^T \\ \vdots \\ \boldsymbol{\varphi}_M^T \end{pmatrix}$$

Signals $\boldsymbol{u}$ s.t.

$$\underbrace{\mathcal{Q}(\boldsymbol{\varphi}_1^\top \boldsymbol{u} + \xi_1)}_{\delta\lfloor(\boldsymbol{\varphi}_1^\top \boldsymbol{u} + \xi_1)/\delta\rfloor} = \text{cst.}$$

Properties of $\mathsf{A}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{\xi})$

- 1. For consistent reconstruction method



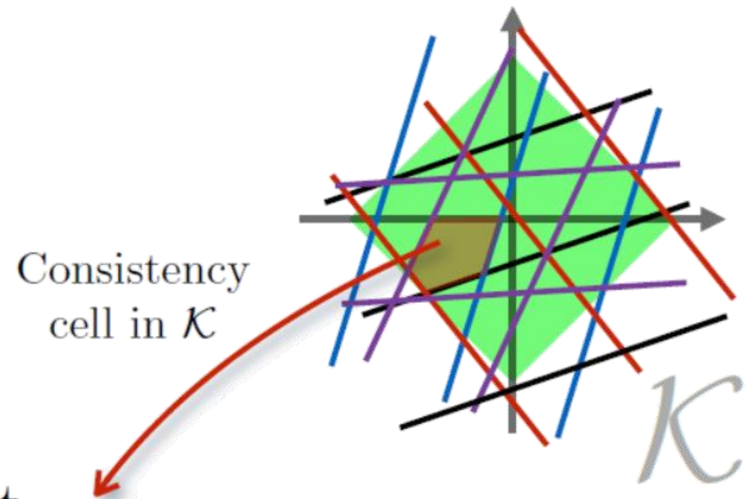
$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\varphi}_1^T \\ \vdots \\ \boldsymbol{\varphi}_M^T \end{pmatrix}$$

Signals $\boldsymbol{u}$ s.t.

$$\left.\begin{aligned} \mathcal{Q}(\boldsymbol{\varphi}_1^\top \boldsymbol{u} + \xi_1) &= \text{cst.} \\ \mathcal{Q}(\boldsymbol{\varphi}_2^\top \boldsymbol{u} + \xi_2) &= \text{cst.} \end{aligned}\right\}$$

Properties of $\mathsf{A}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi x} + \boldsymbol{\xi})$

- 1. For consistent reconstruction method



$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\varphi}_1^T \\ \vdots \\ \boldsymbol{\varphi}_M^T \end{pmatrix}$$

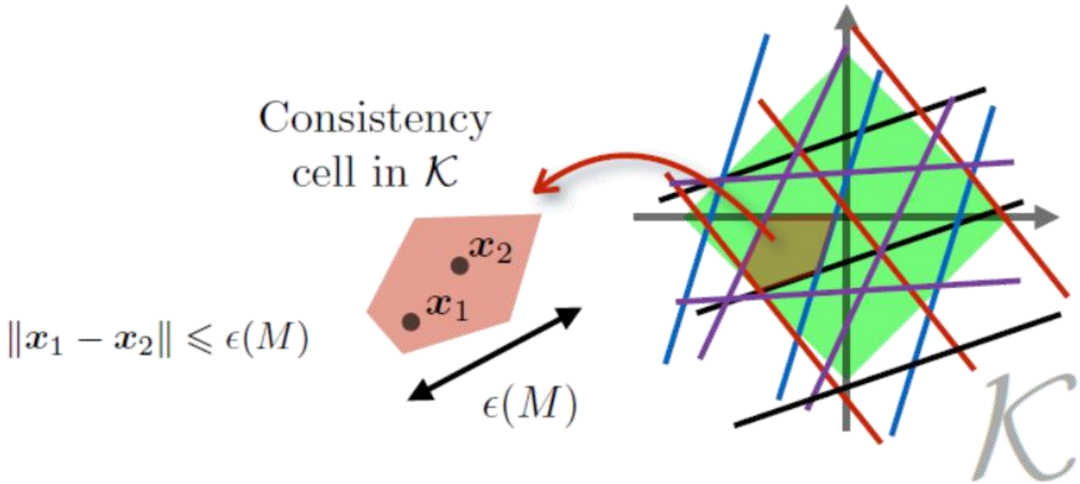
Signals $\boldsymbol{u}$ s.t.

$$\mathsf{A}(\boldsymbol{u}) := \mathcal{Q}(\boldsymbol{\Phi u} + \boldsymbol{\xi}) = \boldsymbol{y}$$

for some $\boldsymbol{y}$

Properties of $\mathsf{A}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi x} + \boldsymbol{\xi})$

- 1. For consistent reconstruction method

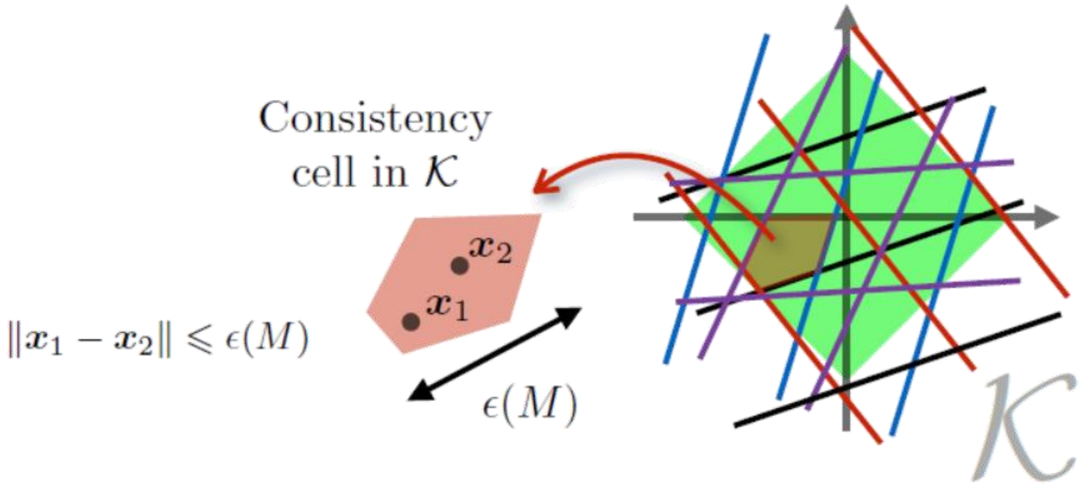


$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\varphi}_1^T \\ \vdots \\ \boldsymbol{\varphi}_M^T \end{pmatrix}$$

Size $\epsilon(M)$ should decay for large M!

Properties of $\mathsf{A}(\boldsymbol{x}) := \mathcal{Q}(\boldsymbol{\Phi}\boldsymbol{x} + \boldsymbol{\xi})$

- 1. For consistent reconstruction method



$$\boldsymbol{\Phi} = \begin{pmatrix} \boldsymbol{\varphi}_1^T \\ \vdots \\ \boldsymbol{\varphi}_M^T \end{pmatrix}$$

Size $\epsilon(M)$ should decay for large M!

For $\boldsymbol{\Phi}$ a random Gaussian matrix, $\epsilon(M) \leqslant C_{\mathcal{K}}\, M^{-q}$.

with $q = 1$ if $\mathcal{K}$ is Σ_K, or low-rank matrices (and others),
and $q = \frac{1}{4}$ otherwise. **[LJ, 17]**

Proposed Recovery Mechanism

Introduce *quantization consistency*

Quantized Orthogonal Matching Pursuit (Q-OMP)

- Sparse signals

$$\min \|\mathbf{y} - \mathcal{Q}(\mathbf{\Phi s})\|_2^2 \ \text{ s.t. } \ \|\mathbf{s}\|_0 \leq K$$

- Compressible signals

$$\min \|\mathbf{y} - \mathcal{Q}(\mathbf{\Phi D x})\|_2^2 \ \text{ s.t. } \ \|\mathbf{x}\|_0 \leq K$$

Q-OMP

Algorithm 1: Quantized Orthogonal Matching Pursuit (Q-OMP)

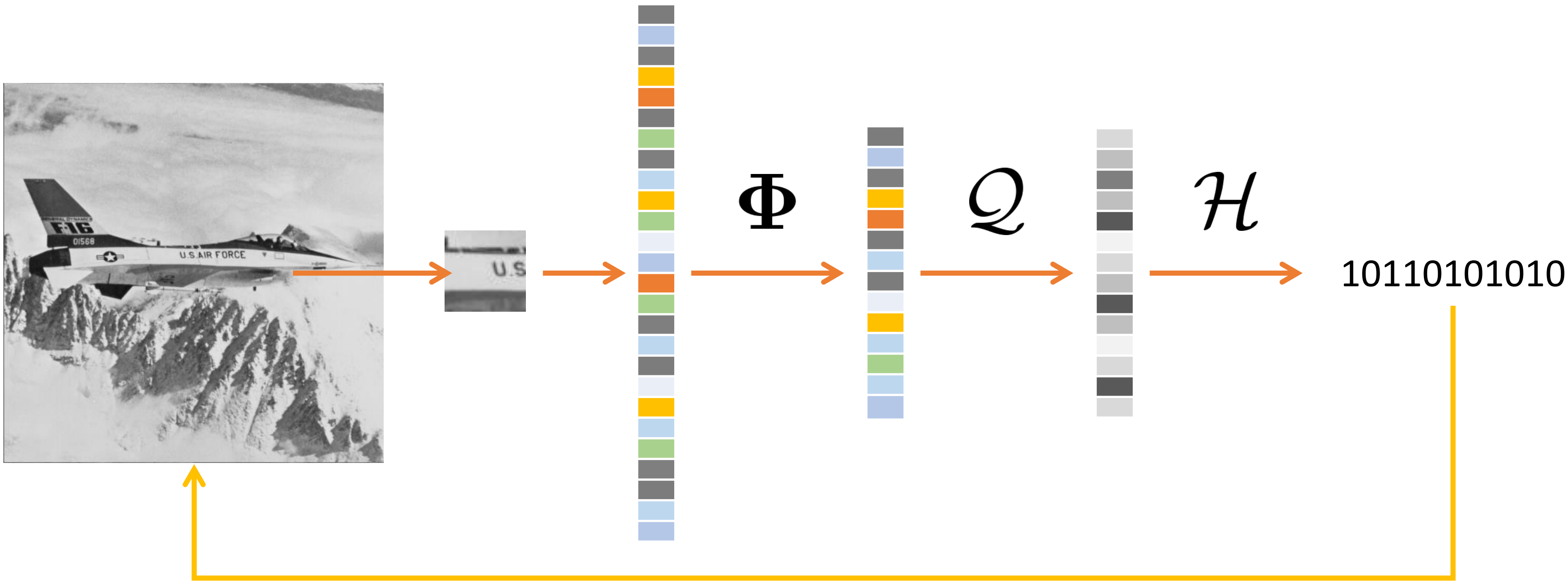
Input: The measurements $\mathbf{y}$,
The sensing matrix $\mathbf{\Phi}$,
The dictionary of examples $\mathbf{D}$,
The error tolerance *threshold* and/or maximum number of iterations k.

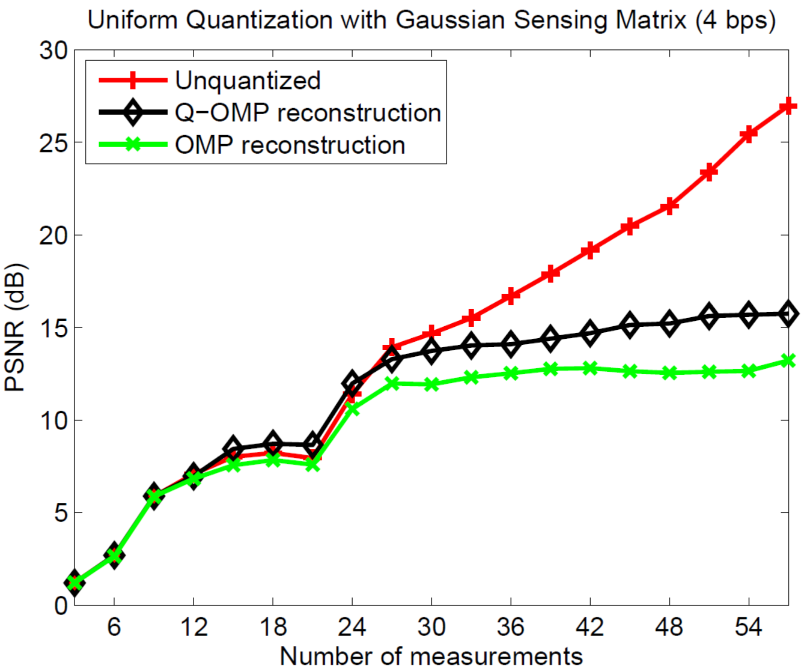
Output: The sparse representation coefficients $\hat{\mathbf{s}}$.

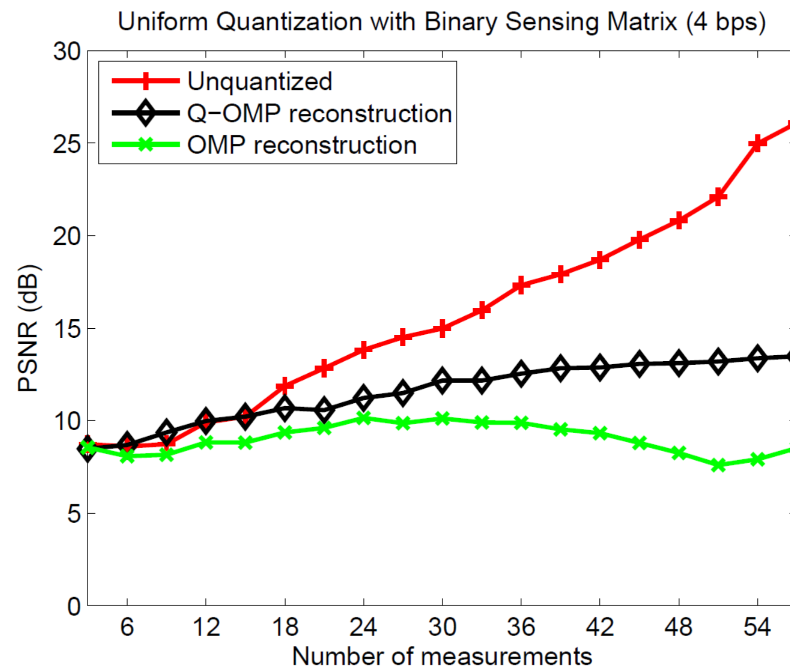
1: **initialization** $T^0 = \emptyset, \mathbf{r}^0 = \mathbf{y}$
2: **while** $error \geq threshold$ or $k \leq iteration\_lim$ **do**
3: $T^k = T^{k-1} \cup \arg\max_j |\mathcal{Q}(< \mathbf{r}^{k-1}, (\mathbf{\Phi D})_j >)|$.
4: $\hat{\mathbf{s}}_{T^k} = \arg\min_s \|y - \mathcal{Q}((\mathbf{\Phi D})_{T^k}\mathbf{s})\|_2$.
5: $\mathbf{r}^k = \mathbf{y} - (\mathbf{\Phi D})_{T^k}\hat{\mathbf{s}}_{T^k}$.
6: **set** $k \longleftarrow k+1$
7: **end while**

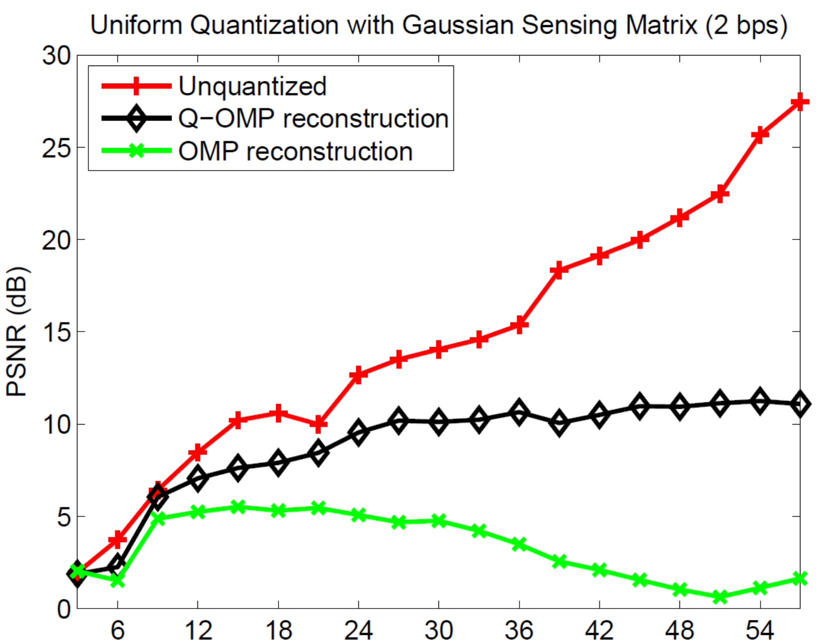
Experimental Setup

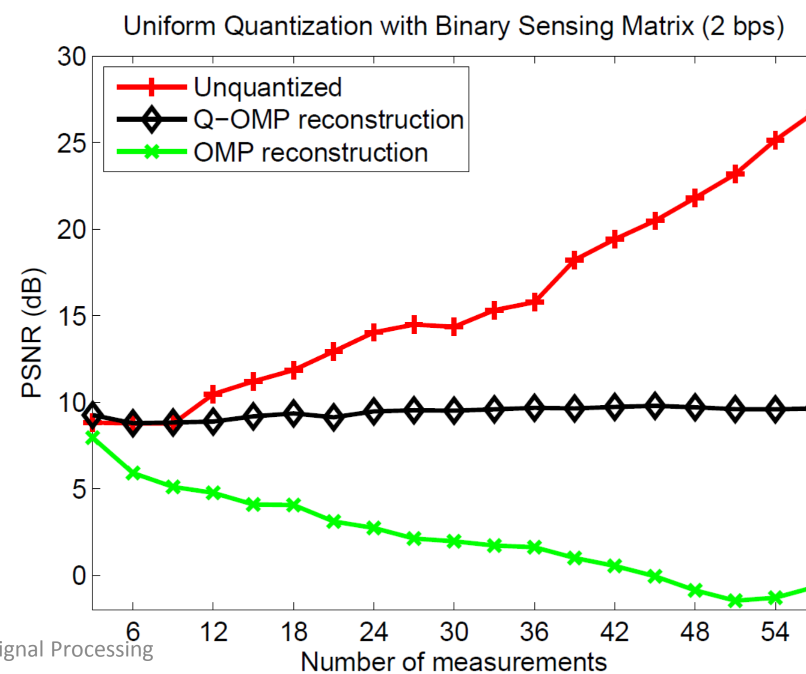
Sensing Matrices	Quantization
✓ Gaussian	✓ Uniform
✓ Binary	✓ Optimal

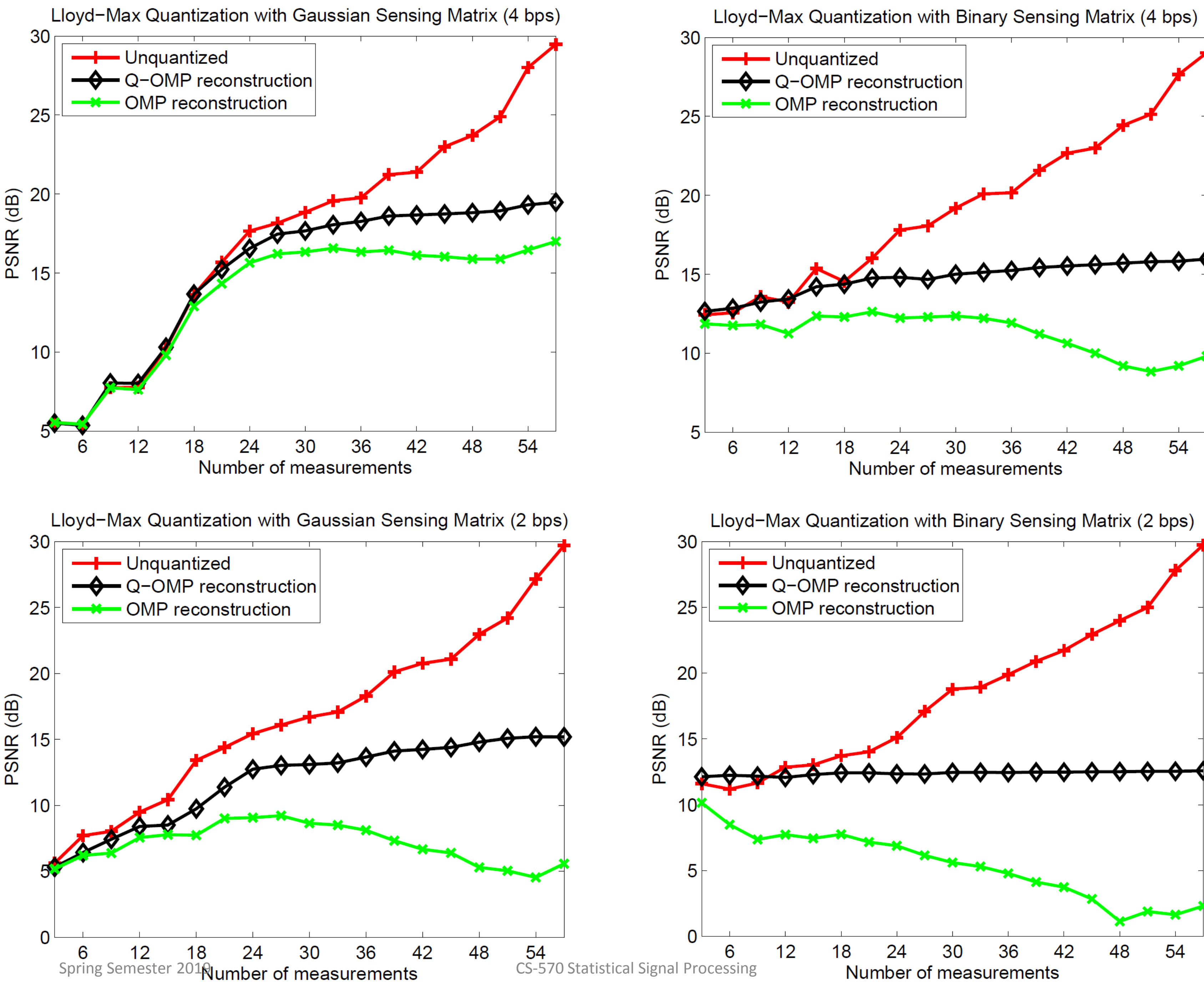



Uniform Quantization with Gaussian Sensing Matrix (4 bps)
Unquantized
Q−OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements
0
5
10
15
20
25
30
6
12
18
24
30
36
42
48
54


Uniform Quantization with Binary Sensing Matrix (4 bps)
Unquantized
Q−OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements
0
5
10
15
20
25
30
6
12
18
24
30
36
42
48
54


Uniform Quantization with Gaussian Sensing Matrix (2 bps)
Unquantized
Q−OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements
0
5
10
15
20
25
30
6
12
18
24
30
36
42
48
54


Uniform Quantization with Binary Sensing Matrix (2 bps)
Unquantized
Q−OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements
0
5
10
15
20
25
30
6
12
18
24
30
36
42
48
54


Lloyd−Max Quantization with Gaussian Sensing Matrix (4 bps)
Lloyd−Max Quantization with Binary Sensing Matrix (4 bps)
Lloyd−Max Quantization with Gaussian Sensing Matrix (2 bps)
Lloyd−Max Quantization with Binary Sensing Matrix (2 bps)
Unquantized
Q−OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements

Discussion

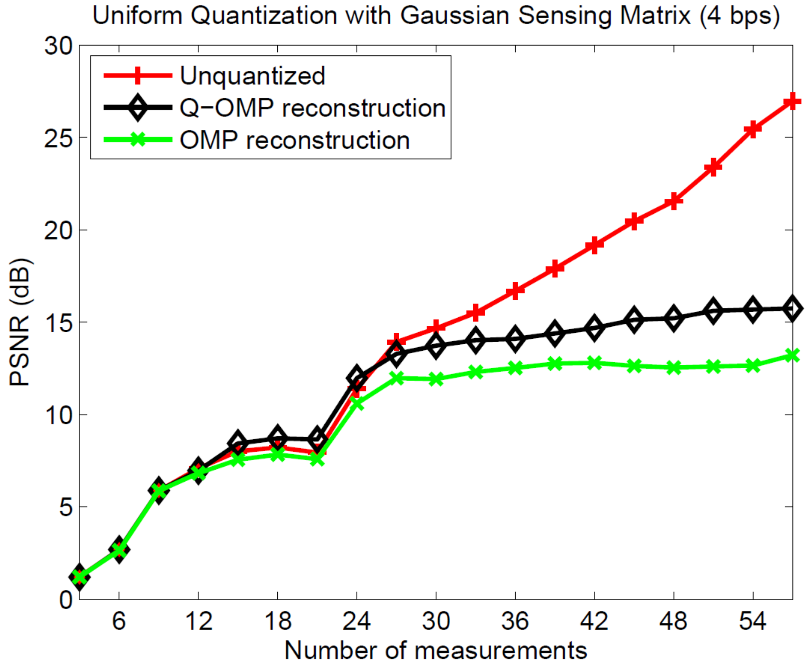
4 BPM		OMP	Q-OMP
Uniform	Gaussian	✓	✓
	Binary	✗	✓
Optimal	Gaussian	✓	✓
	Binary	✗	~

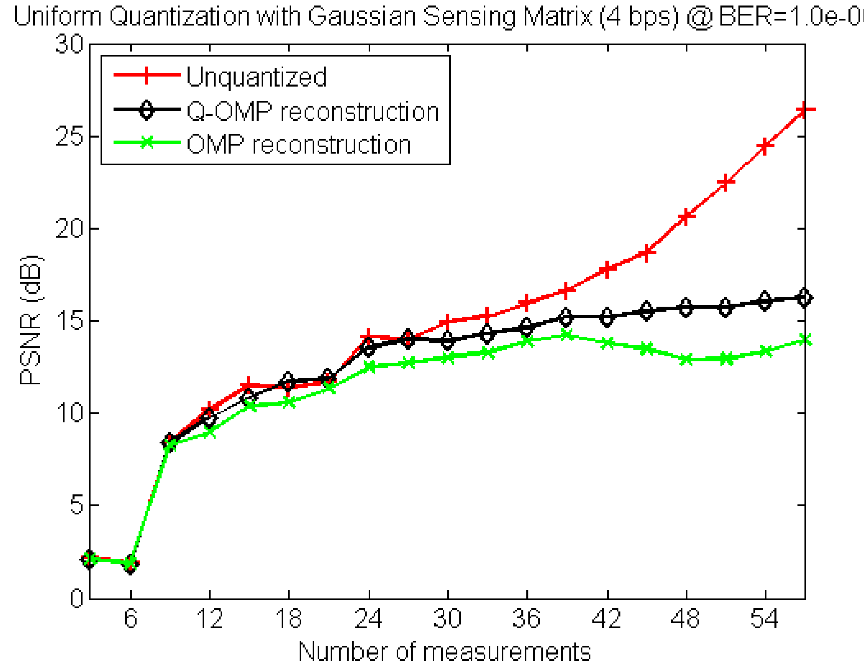
2 BPM		OMP	Q-OMP
Uniform	Gaussian	✗	✓
	Binary	✗	✓
Optimal	Gaussian	✗	✓
	Binary	✗	~

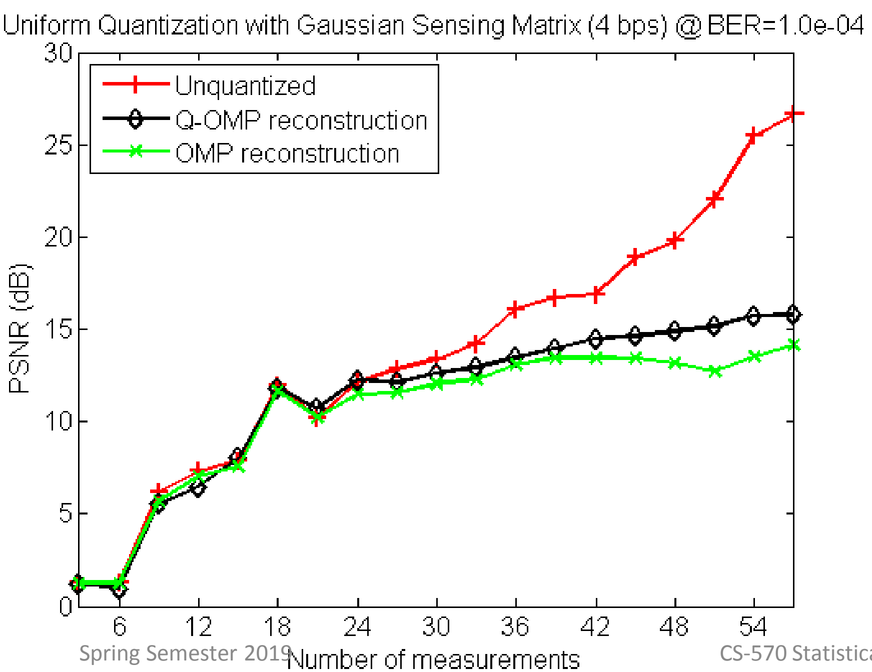
FORTH Institute of Computer Science

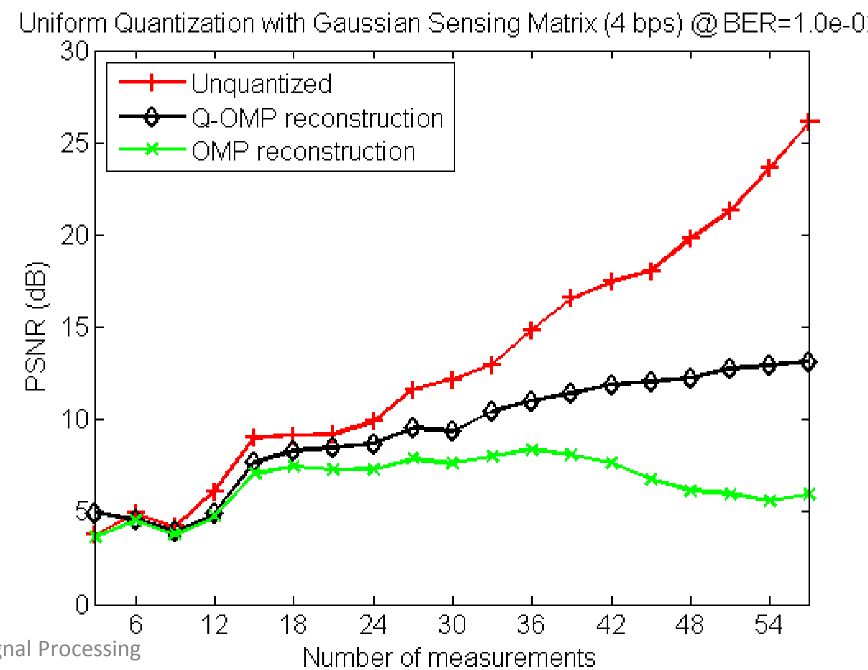
Source coding

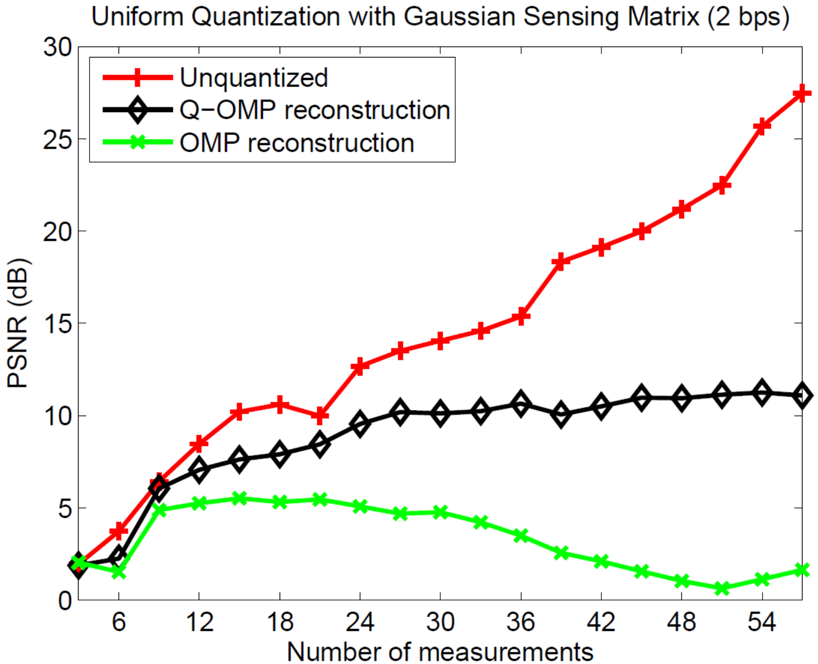
- Examine performance under realistic conditions
- Lossless coding: $\mathcal{H}$
  - Convert indices to binary vectors
  - Entropy coding
    - Huffman coding(codewords, prob.) -> binary code (JPEG)
    - Arithmetic coding (message) - > binary (JPEG2000)
  - Introduce bit errors
    - Storage
    - Transmission

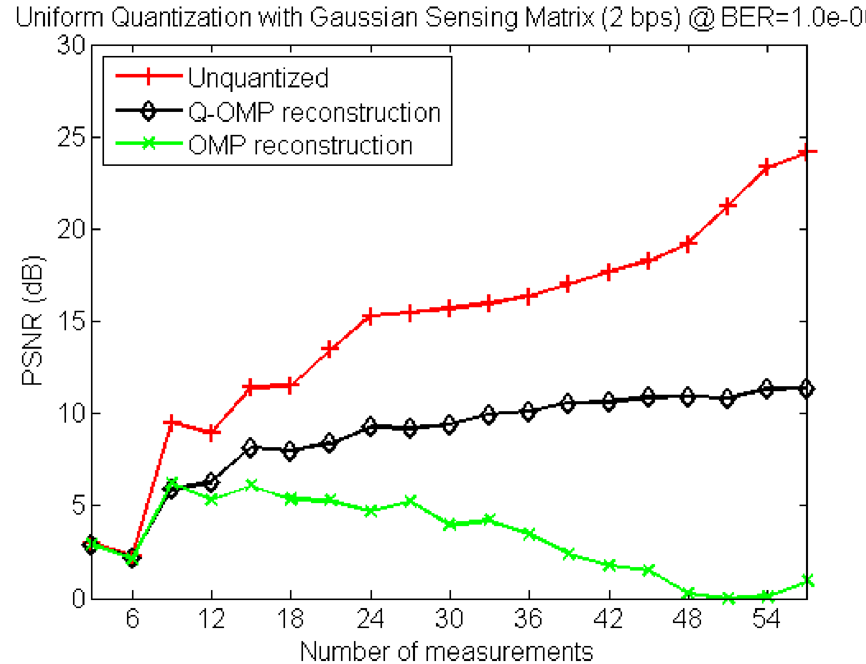

Uniform Quantization with Gaussian Sensing Matrix (4 bps)
Unquantized
Q−OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements

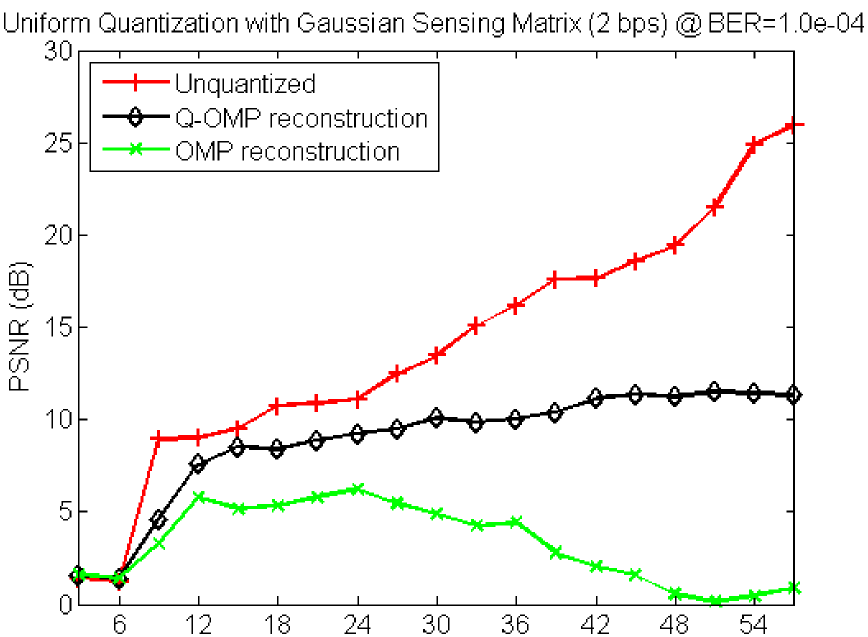

Uniform Quantization with Gaussian Sensing Matrix (4 bps) @ BER=1.0e-06
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements

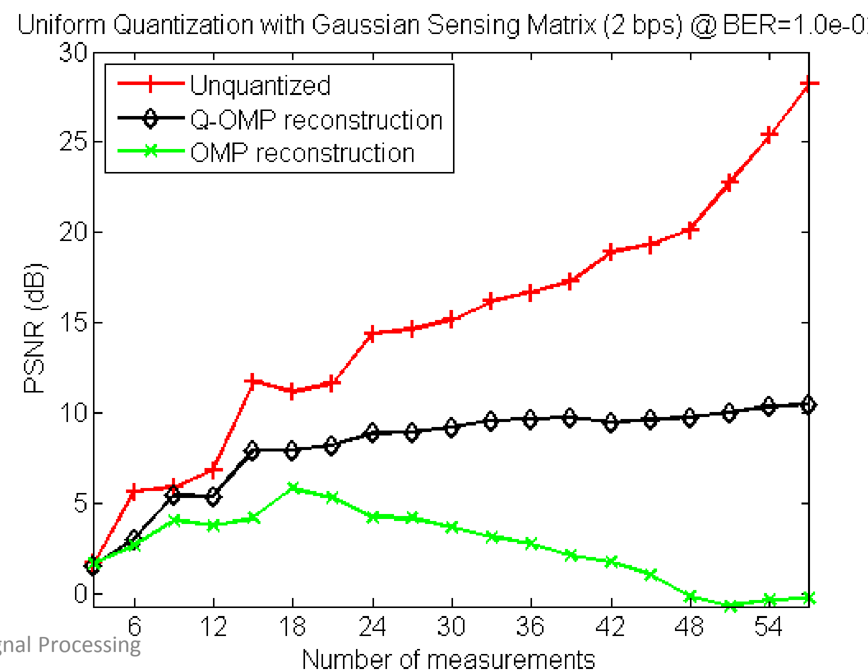

Uniform Quantization with Gaussian Sensing Matrix (4 bps) @ BER=1.0e-04
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements

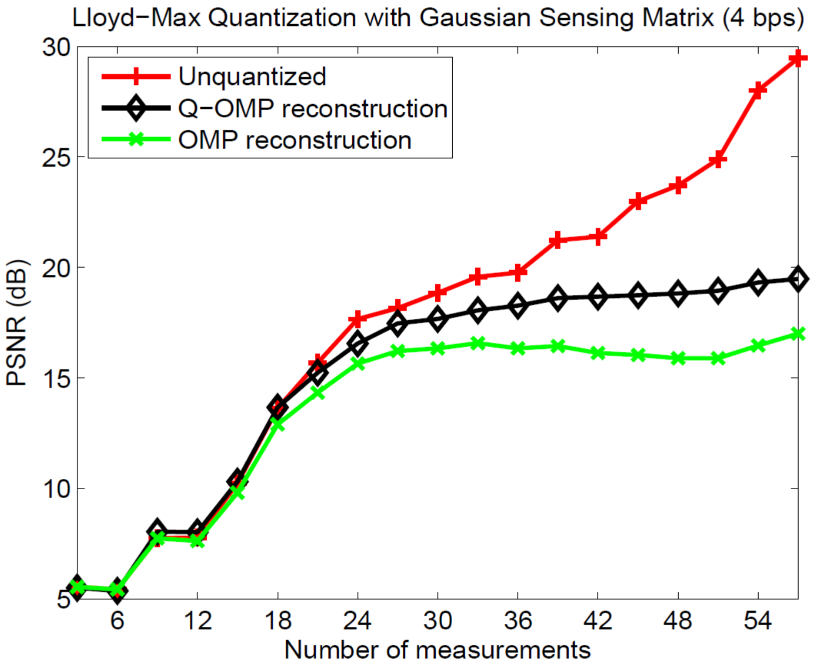

Uniform Quantization with Gaussian Sensing Matrix (4 bps) @ BER=1.0e-02
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements

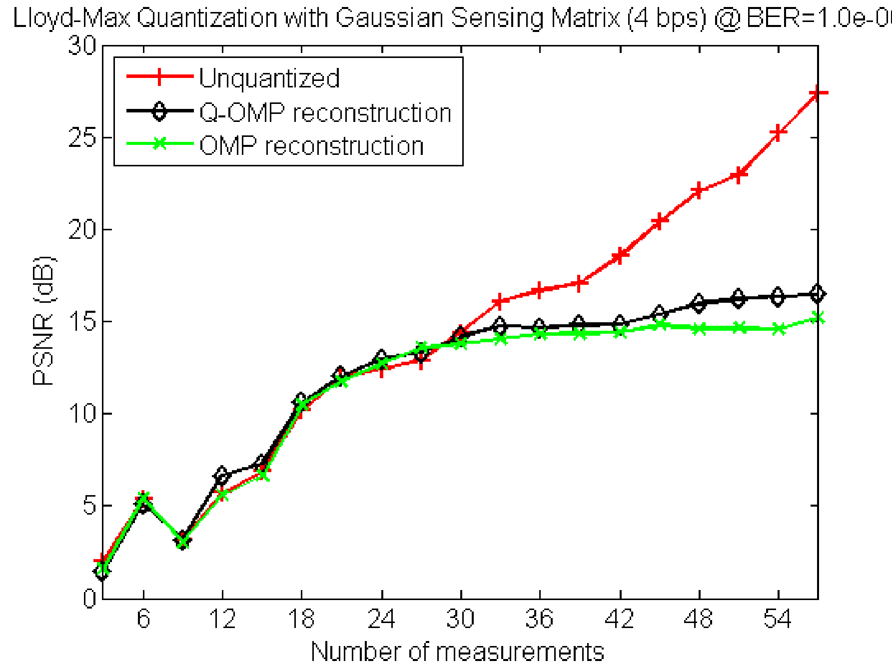

Uniform Quantization with Gaussian Sensing Matrix (2 bps)
Unquantized
Q−OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements
0
5
10
15
20
25
30
6
12
18
24
30
36
42
48
54

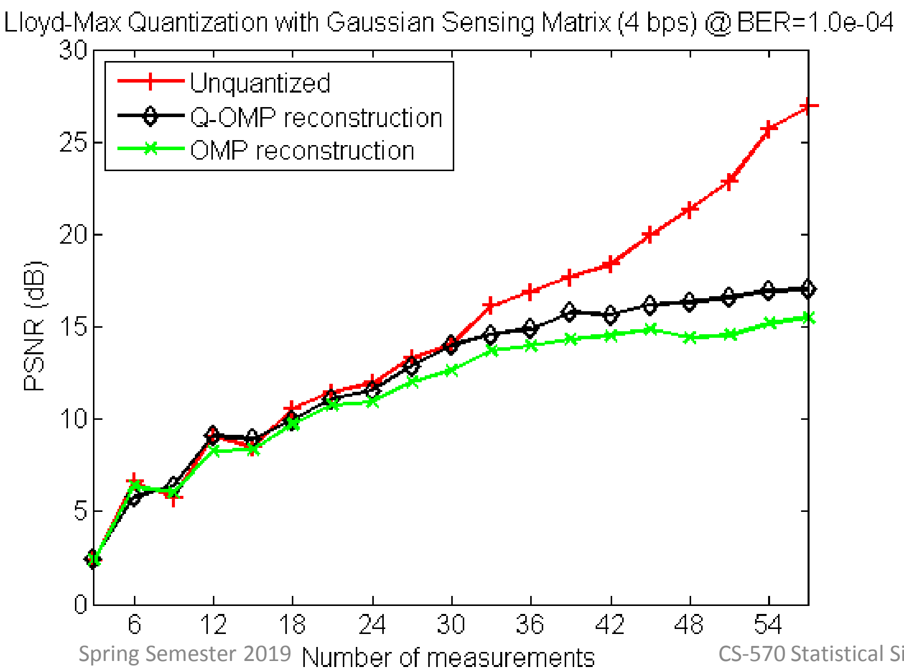

Uniform Quantization with Gaussian Sensing Matrix (2 bps) @ BER=1.0e-06
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements
0
5
10
15
20
25
30
6
12
18
24
30
36
42
48
54

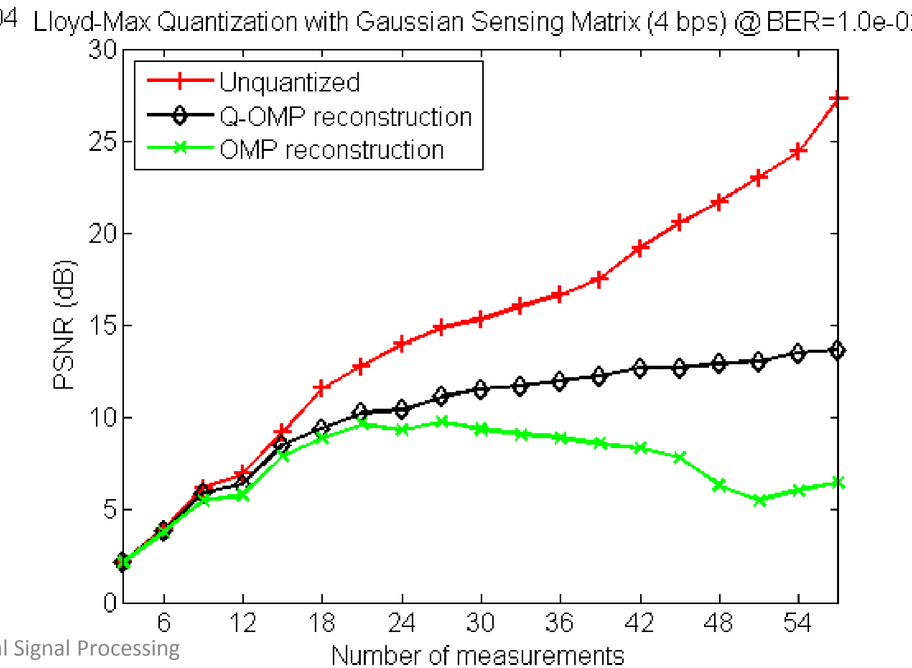

Uniform Quantization with Gaussian Sensing Matrix (2 bps) @ BER=1.0e-04
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements
0
5
10
15
20
25
30
6
12
18
24
30
36
42
48
54

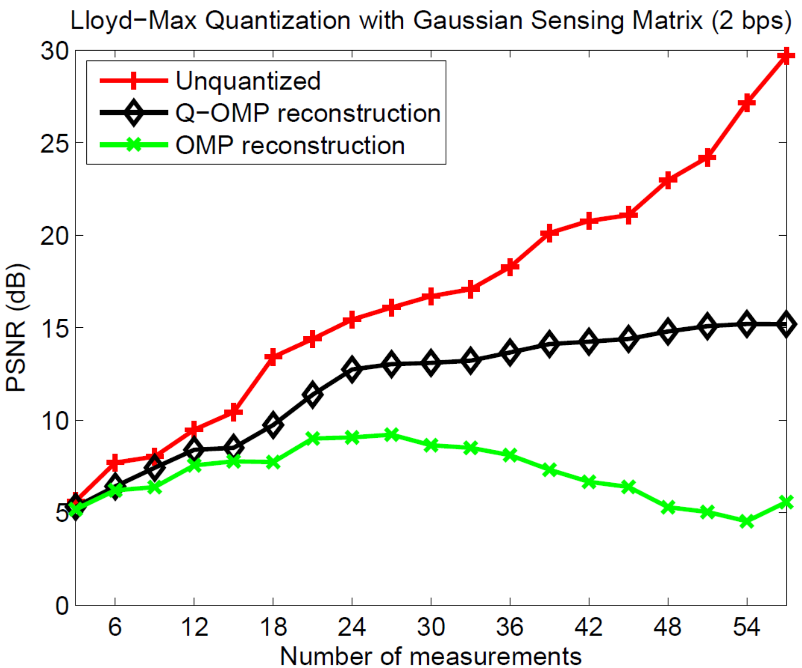

Uniform Quantization with Gaussian Sensing Matrix (2 bps) @ BER=1.0e-02
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements
0
5
10
15
20
25
30
6
12
18
24
30
36
42
48
54

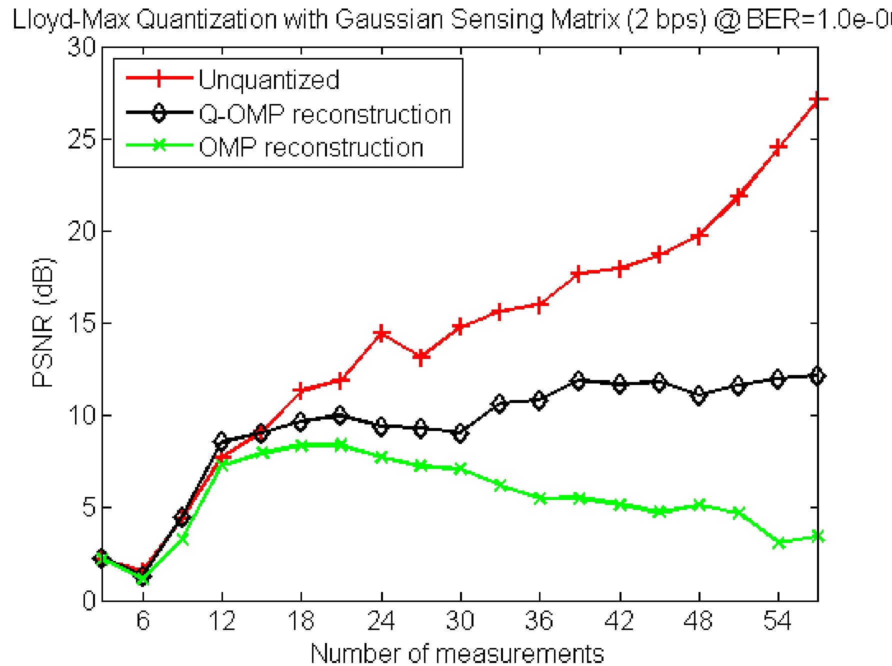

Lloyd−Max Quantization with Gaussian Sensing Matrix (4 bps)
Unquantized
Q−OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements

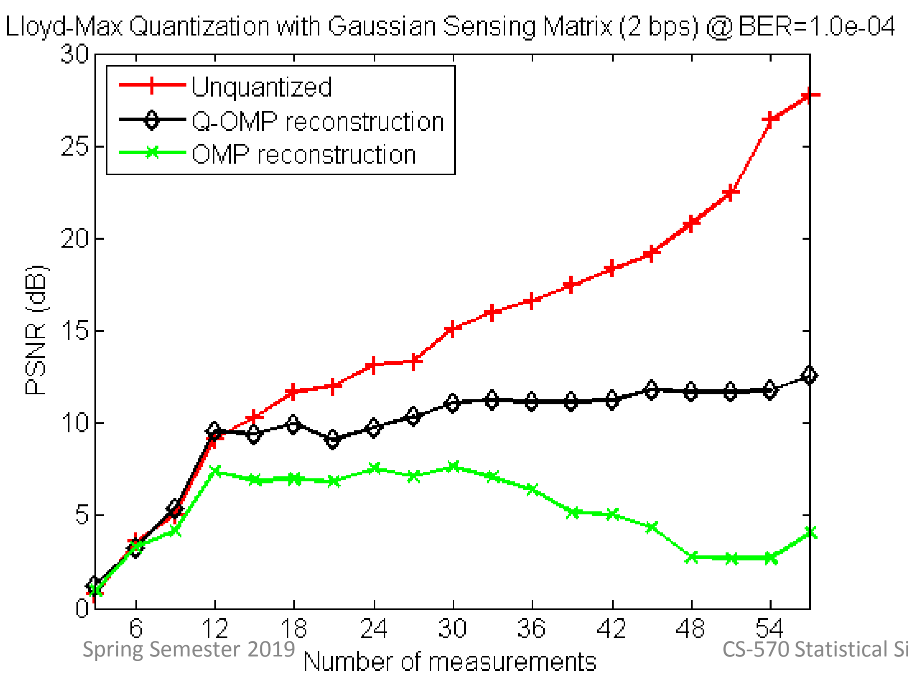

Lloyd-Max Quantization with Gaussian Sensing Matrix (4 bps) @ BER=1.0e-06
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements

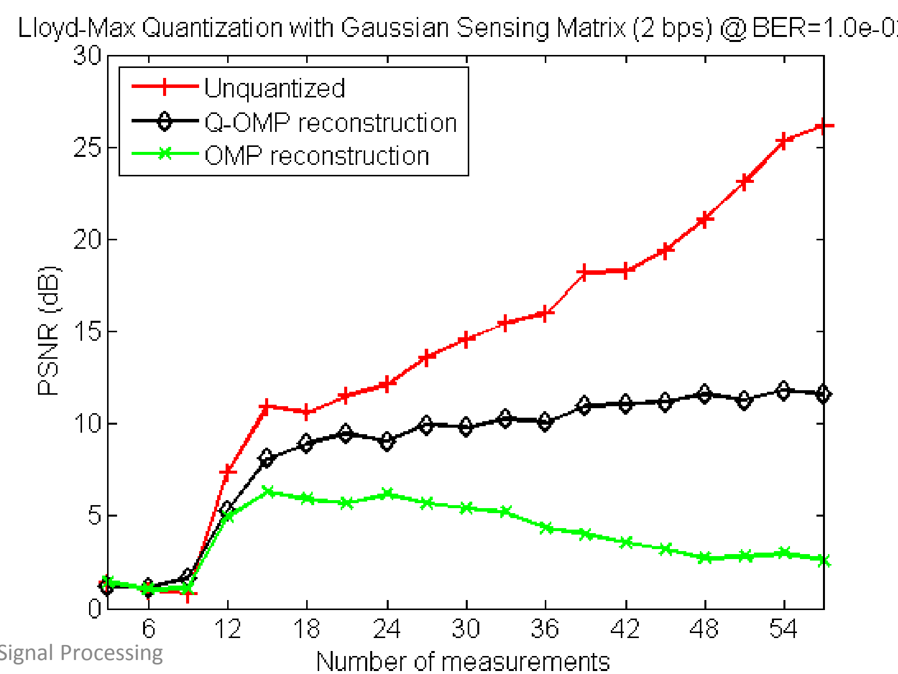

Lloyd-Max Quantization with Gaussian Sensing Matrix (4 bps) @ BER=1.0e-04
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements


Lloyd-Max Quantization with Gaussian Sensing Matrix (4 bps) @ BER=1.0e-02
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements


Lloyd−Max Quantization with Gaussian Sensing Matrix (2 bps)
Unquantized
Q−OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements


Lloyd-Max Quantization with Gaussian Sensing Matrix (2 bps) @ BER=1.0e-06
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements


Lloyd-Max Quantization with Gaussian Sensing Matrix (2 bps) @ BER=1.0e-04
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements


Lloyd-Max Quantization with Gaussian Sensing Matrix (2 bps) @ BER=1.0e-02
Unquantized
Q-OMP reconstruction
OMP reconstruction
PSNR (dB)
Number of measurements

Discussion

4 b.p.m		OMP	Q-OMP
Small	Uniform	✓	✓
	Optimal	✓	✓
Medium	Uniform	✓	✓
	Optimal	✓	✓
Large	Uniform	✗	✓
	Optimal	✗	✓

2 b.p.m		OMP	Q-OMP
Small	Uniform	✓	✓
	Optimal	✓	✓
Medium	Uniform	✓	✓
	Optimal	✓	✓
Large	Uniform	✗	✓
	Optimal	✓	✓

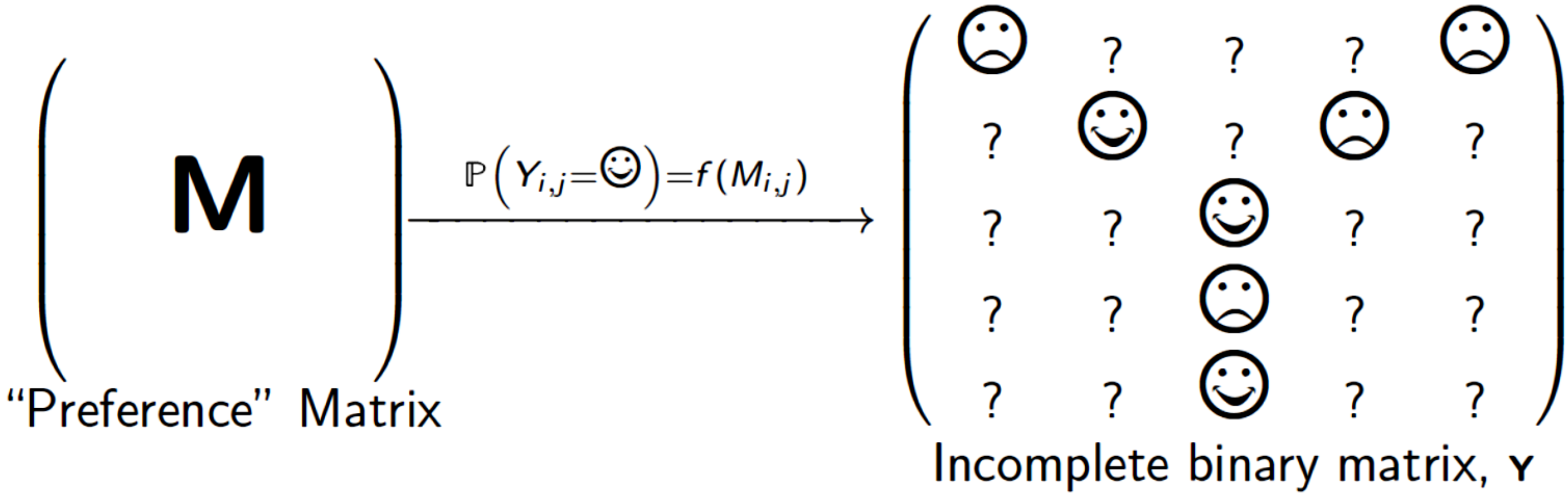
- Quantization is critical in CS
  - Little effect w.r.t uniform vs. scalar
  - Large effects on b.p.m & sampling scheme
  - Sensitive to BER
- Q-OMP
  - Large effect w.r.t optimality
  - Smaller effect on b.p.m & sampling scheme
  - Robust to BER

1-bit Matrix Completion

- Netflix challenge v2

	John	Anne	Scot	Mark	Alice
Chicago	☺	☹	?	?	?
Matrix	☹	?	☺	?	?
Star wars	?	?	☺	?	☹
Inception	?	☺	?	☹	?
Alien	☺	☹	?	?	?
Pulp Fiction	?	?	☹	?	☺

Generalized linear model



- $\mathbf{M}$ is unknown. $\mathbf{M}$ has (approximately) low rank.
- $f : \mathbb{R} \to [0,1]$ is a known function (e.g., the logistic curve).
- $\mathbf{M} \in \mathbb{R}^{d \times d}, \mathbf{Y} \in \{☺, ☹\}^{d \times d}$.
- $\Omega \subset \{1, 2, \ldots, d\} \times \{1, 2, \ldots, d\}$. You see $\mathbf{Y}_\Omega$.

FORTH Institute of Computer Science

Formulation

- Consider a $d \times d$ matrix **M** with rank *r*. Suppose we observe a subset Ω of entries of a matrix **Y** in the following way:

$$Y_{i,j} = \begin{cases} +1 & \text{if } M_{i,j} + Z_{i,j} \geq 0 \\ -1 & \text{if } M_{i,j} + Z_{i,j} < 0 \end{cases}$$

Observations Matrix Noise

- Generalized model: M is $d_1 x d_2$, and $f : \mathbf{R} \to [0,1]$

$$Y_{i,j} = \begin{cases} +1 & \text{with probability } f(M_{i,j}), \\ -1 & \text{with probability } 1 - f(M_{i,j}) \end{cases} \quad \text{for } (i,j) \in \Omega.$$

Equivalence when $f(x) = \frac{e^x}{1+e^x}$ and Z i.i.d logistic distribution

Estimation

- Log-likelihood function

$$\mathcal{L}_{\Omega,\mathbf{Y}}(\boldsymbol{X}) := \sum_{(i,j)\in\Omega} \left(\mathbb{1}_{[Y_{i,j}=1]} \log(f(X_{i,j})) + \mathbb{1}_{[Y_{i,j}=-1]} \log(1 - f(X_{i,j})) \right)$$

- Recovery of **M**

$$\widehat{\boldsymbol{M}} = \arg\max_{\boldsymbol{X}} \mathcal{L}_{\Omega,\mathbf{Y}}(\boldsymbol{X}) \quad \text{subject to} \quad \|\boldsymbol{X}\|_* \le \alpha\sqrt{r d_1 d_2} \quad \text{and} \quad \|\boldsymbol{X}\|_\infty \le \alpha.$$

- Error

Let f be the logistic function. Assume that $\frac{1}{d}\|\mathbf{M}\|_* \le \sqrt{r}$. Suppose the sampling set is chosen at random with $\mathbb{E}|\Omega| = m \ge d\log(d)$. Then with high probability,

$$\frac{1}{d^2}\sum_{i,j} d_H^2(f(\hat{M}_{i,j}), f(M_{i,j}))^2 \le C \min\left(\sqrt{\frac{rd}{m}}, 1\right).$$