Adaptive Non-uniform Compressive Sampling for Time-varying Signals

Alireza Zaeemzadeh, Mohsen Joneidi, and Nazanin Rahnavard
Presenter: Alireza Zaeemzadeh
Communications and Wireless Networks (CWN) Lab
Electrical and Computer Engineering Department
University of Central Florida
http://cwnlab.eecs.ucf.edu/

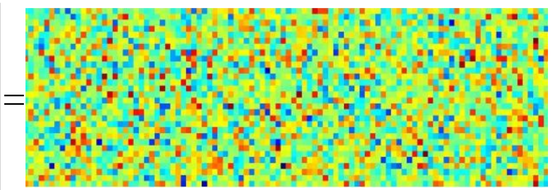
Compressive Sensing: An Overview

signal $\underline{x}$ → **Compressive Sensing (Encoding)** → $\underline{y} = \Phi\, \underline{x}$ → Store/Transmit → **Compressive Sensing Recovery (Decoding)** → $\hat{\underline{x}}$ N

$M = O(K\, log(N)) \ll N$

Measurement Matrix

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_M \end{bmatrix} =$$

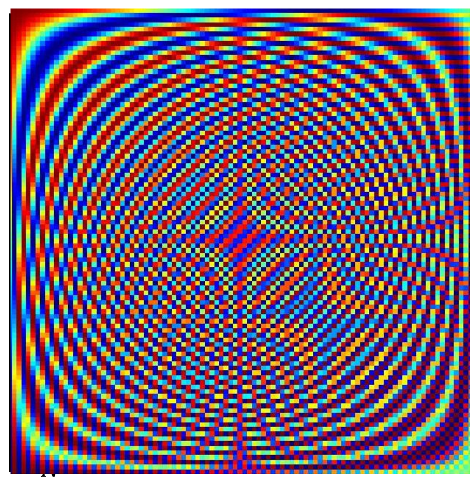


$\underline{y}_{M\times1}$

$\Phi_{M\times N}$

(Random Gaussian or Bernoulli-Gaussian)

$\Psi_{N\times N}$



$$\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_N \end{bmatrix}$$

$\underline{s}_{N\times1}$

k-sparse

$$\underbrace{\qquad\qquad}_{\underline{x}_{N\times1}}$$

$\underline{x} = \Psi_{N\times N}\underline{s}$

- Each row of Φ corresponds to one measurement $\qquad y_i = \sum_{i=1}^{N} \varphi_{ij} x_j$
- Reconstruction we solve $\qquad \hat{\underline{s}} = \arg\min \left\|\underline{s}\right\|_{l_1}$ subject to $\underline{y} = \Phi\Psi\underline{s}$

Conventional CS

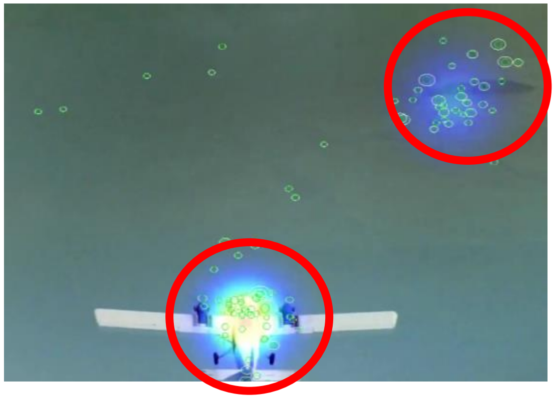
- Solely a simple sparsity model is considered for the underlying signal
- Entries of Φ are iid random variables:
  - i.i.d. zero mean Gaussian
  - i.i.d. Bernoulli {+1,-1}

- Identically distributed entries leads to *uniform* distribution of the sensing energy among all the signal coefficients

- All the signal coefficients are treated equally at sampling phase
- The underlying assumption is that all the coefficients are equally important!

Motivation for Non-uniform CS

- In many application, different coefficients might have different importance levels
- Specifically we might have a region of interest (ROI) and uniform sampling is not efficient
- Intelligent and adaptive CS techniques are needed to leverage signal model or side information about the coefficient



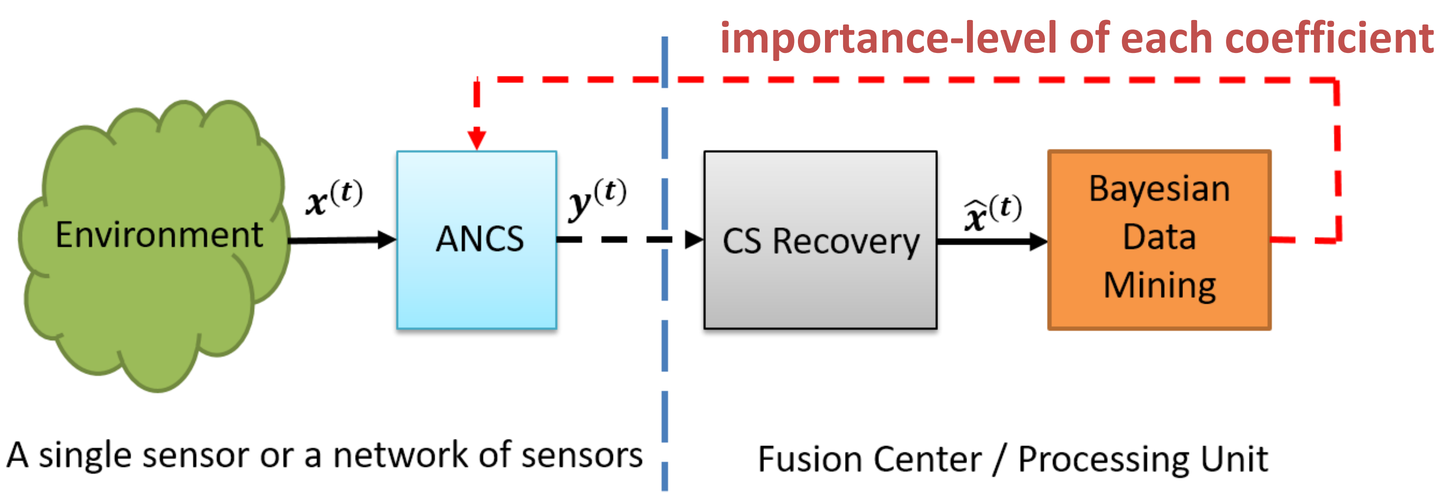
Salient area in a sequence of video frames.
Photo source: The DIEM Project
[online: https://thediemproject.wordpress.com/]

Related Work

Adaptive CS	Sequentially designs the measurement matrix to focus sampling into a desired subspace of signal.	
	Not concerned with recovering time-varying signals.	
	Adaptive CS (Iwen et al. 2011), Near-optimal adaptive CS (Malloy et al. 2014), constrained adaptive sensing (Davenport et al. 2016), sequentially designed CS (Haupt et al. 2012), compressive binary search (Davenport et al. 2012), info-greedy sequential adaptive CS (Braun et al.), sequential CS (Malioutov et al. 2010)	
Dynamic CS	The information from the previous observations are utilized to modify the recovery step.	
	Not concerned with the sampling step.	
	Modfied CS (Vaswani et al. 2010) and Kalman filtered CS (Vaswani 2008), Dynamic CS via AMP (Ziniel et al. 2013), Time Correlated CS (Shahrasbi et al. 2011), Bayesian online recovery (Wijewardhana, et al.)	

Our Contribution

- The main idea of Non-uniform CS (NCS) is to sample the crucial part of the information with more energy than the rest.
- Due to dynamic nature of the problem, *soft importance level information* is advantageous.



Adaptive Non-uniform Compressive Sampling (ANCS)

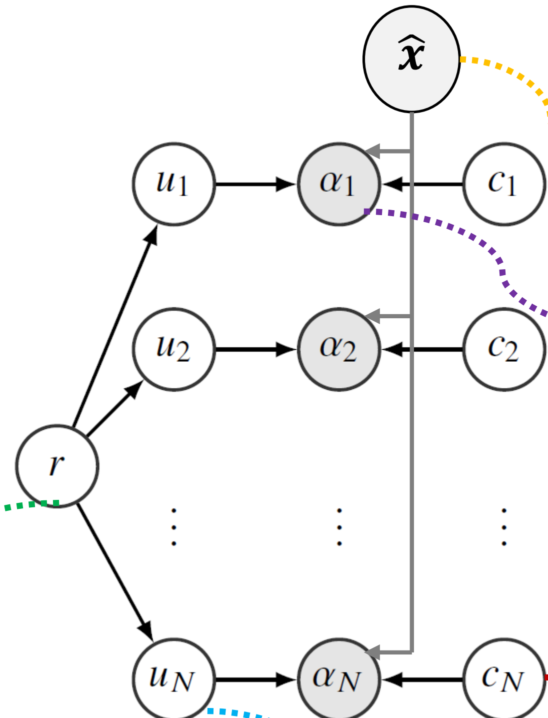
Steps of ANCS

Setting up the **probability model**.

Calculating the probability distributions of the desired variables.

Designing the measurement matrix

Bayesian Data Analysis: Probability Model (1)



Estimate of the signal

$\boldsymbol{\alpha_n} \in \{\mathbf{0}, \mathbf{1}\}$**:** Estimate of the ROI for coefficient n.

Observed Data

$\boldsymbol{r} \in [\mathbf{0}, \mathbf{1}]$**:** overall reliability

$\boldsymbol{u_n} \in \{\mathbf{0}, \mathbf{1}\}$**:** Coefficient-specific reliability

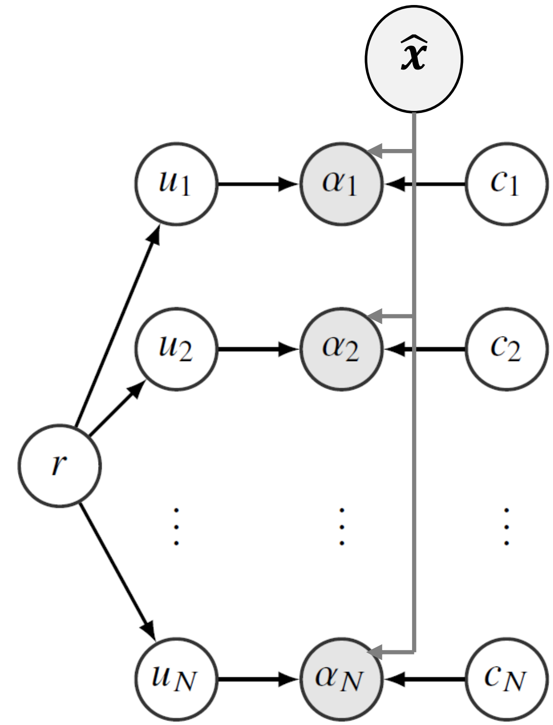
$\boldsymbol{c_n} \in [\mathbf{0}, \mathbf{1}]$ **:** Importance level of the n^{th} coefficient.

Hidden Variables

Bayesian Data Analysis: Probability Model (2)

- Prior information on hidden variables are modeled by:

$$\begin{aligned} r &\sim \text{Beta}(b^1, b^0) \\ u_n &\sim \text{Bernoulli}(r) \\ c_n &\sim \text{Beta}(\beta_n^1, \beta_n^0) \end{aligned}$$



These distributions can be exploited to obtain the joint *prior distribution* over all the hidden variables:

$$P(H) = \prod_{n=1}^{N} \mathbb{P}\{u_n|r\} \prod_{n=1}^{N} \mathbb{P}\{c_n|\beta_n^1, \beta_n^0\} \mathbb{P}\{r|b^1, b^0\}$$

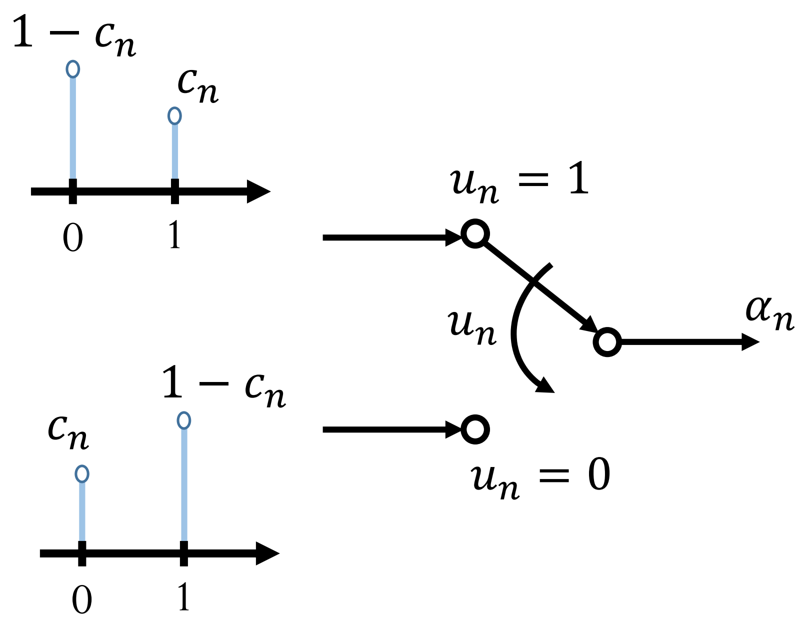
Bayesian Data Analysis: Probability Model (3)

We can define the probability distribution for observed data given the hidden variables $P(D|H)$ as:

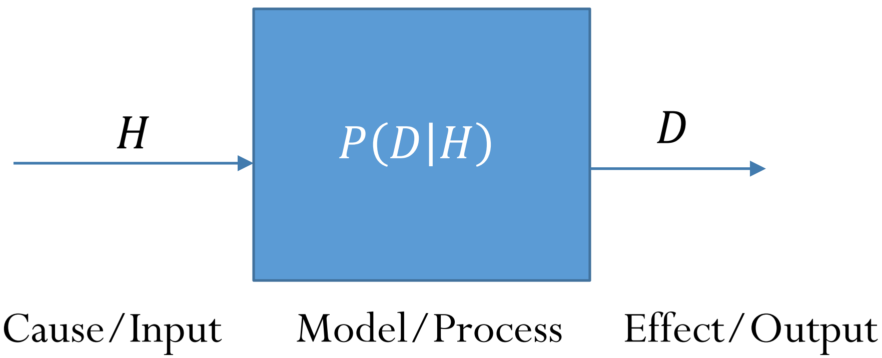
$$\boxed{\alpha_n^{(t)} \sim \; u_n \,\text{Bernoulli}(c_n) + (1 - u_n)\,\text{Bernoulli}(1 - c_n)}$$

Distribution with true parameters (reliable data)

Distribution with incorrect parameters (faulty data)



Bayesian Data Analysis: Inference



Inference: Given the effect/output find the cause/input.

- We wish to obtain $P(H|D)$.
- $P(H|D)$ can be formulated using the prior information $P(H)$ and the sampling distribution $P(D|H)$.

$$P(H|D) \propto P(D|H)P(H)$$

Bayesian Data Analysis: Inference (1)

- To handle the intractable integrals of the inference procedure, *variational inference* is often employed.
- The posterior distribution is approximated by a family of distributions, for which the inference procedure is tractable.

$$P(H|D) \approx Q\,(H)$$

- We aim to find the most similar approximation that minimizes the KL divergence:

$$\hat{Q}\,(H) = arg \min_{Q(H)} D_{KL}\,(P(H|D)||Q\,(H))$$

Bayesian Data Analysis: Inference (2)

- The posterior distribution is assumed to be fully factorized over all the hidden variables. i.e.

$$Q\ (H) = \prod_{n} \mathbb{Q}\{c_n|\hat{\beta}_n^1, \hat{\beta}_n^0\}\mathbb{Q}\{u_n|\tau_n\}\mathbb{Q}\{r_n|\hat{b}^1, \hat{b}^0\}$$

- By decoupling the variables, the inference procedure becomes tractable. And the update rules can be derived in closed form.

Generating $\boldsymbol{\Phi}_{ANCS}$

- The importance levels of the coefficients, inferred by the Bayesian method, is utilized to distribute the energy among the columns of the sensing matrix.

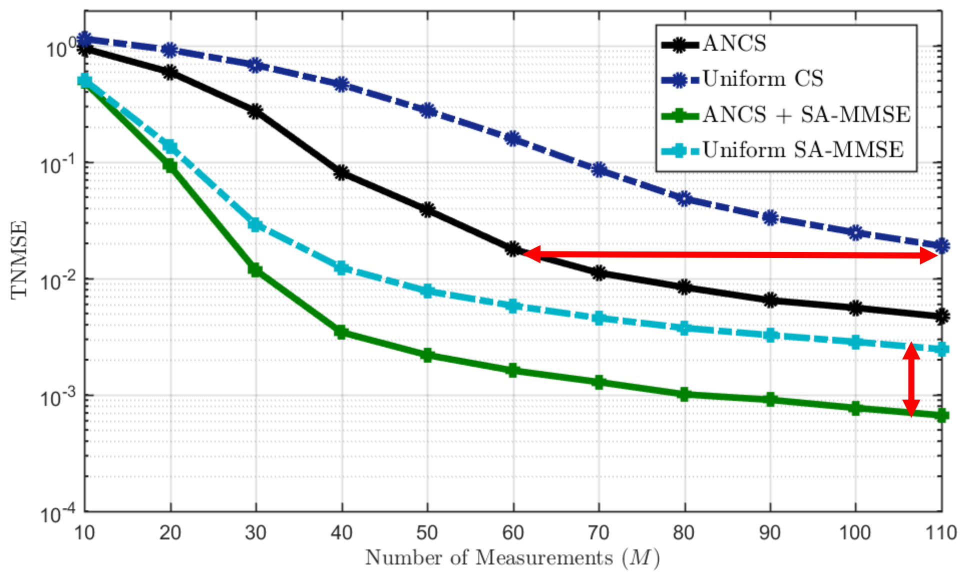
$$\|\boldsymbol{\varphi}_n\|_2 \propto \bar{c}_n$$

- $\bar{c}_n$ is a point estimate of the random variable c_n, (mean, median, or mode).
- We also assume that the available sensing energy is E_{avail}.

$$\|\boldsymbol{\varphi}_n\|_2 = \sqrt{E_{avail}}\,\frac{\bar{c}_n}{\sqrt{\sum_n \bar{c}_n{}^2}}$$

Results: Sparse Signal in Canonical Domain

- N = 200, SNR = 20 dB
- T = 30 time steps
- Window size: W = 5
- $p_{01} = 0.02, \lambda = 0.1,$



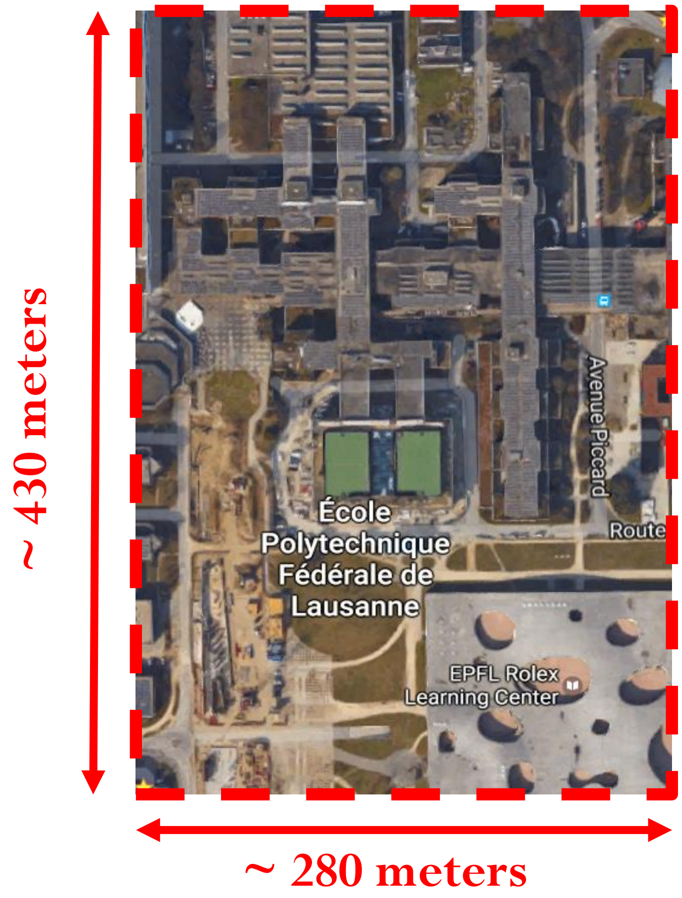
Needs less number of measurements.

Decreases the lower bound.

ANCS: The proposed non-uniform sampling + L1 minimization as the recovery algorithm.
Uniform CS: Gaussian matrix + L1 minimization as the recovery algorithm.
SA-MMSE: Support-aware minimum mean square error estimator as the recovery algorithm (lower bound).
TNMSE: Time-averaged normalized mean square error (averaged over T = 30 time steps).

Results: EPFL data set

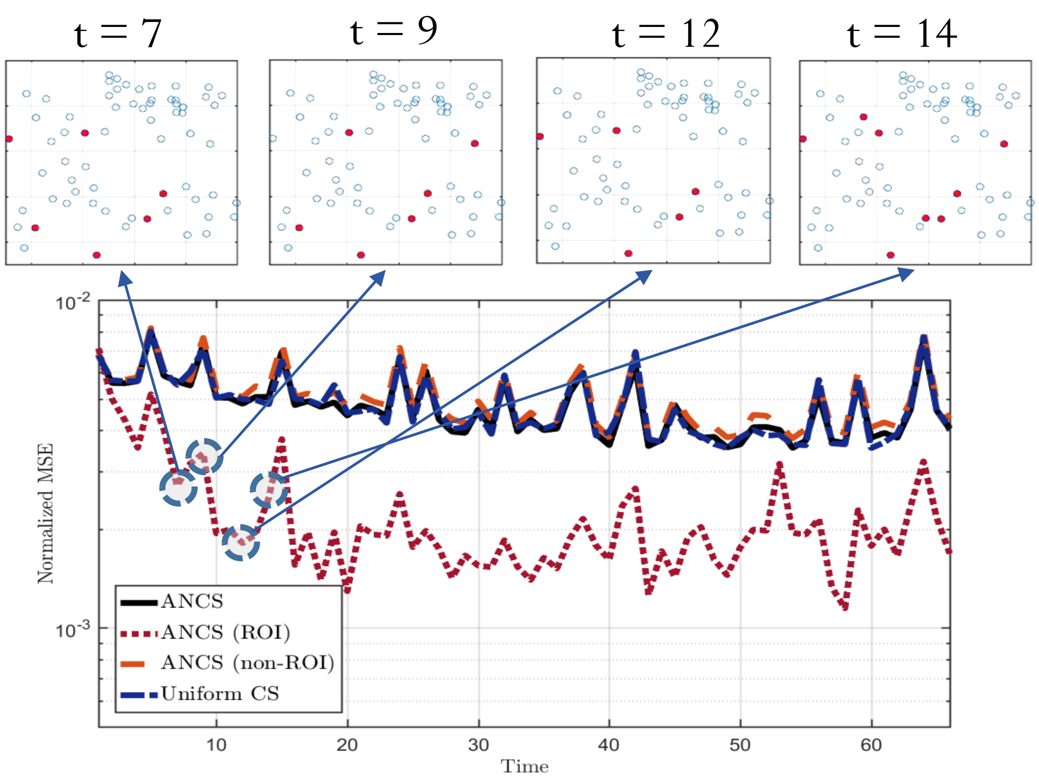
- 97 sensors deployed in a heterogeneous urban environment.
- For our numerical experiments, 80 most active sensors are sampled minutely.
- Ambient temperature, surface temperature, and relative humidity of the sensors are normalized and stacked in a vector (N = 240).



Results: EPFL data set

- M = 140
- T = 66 minutes
- W = 5
- Sparsifying matrix: DCT

- ROI is defined as the sensors that:
  - Has the largest humidity (top %25) AND
  - One of their temperatures is also in top %25.



Future Research Directions

- Finding the *optimal* energy distribution and analyzing the lower bound of the error.
- Employing higher level statistics of the inferred distributions.
- Exploiting spatial correlation, as well as temporal correlation, e.g. camera sensor or spectrum sensor networks.

Thank you!

References

1. A. Zaeemzadeh, M. Joneidi, and N. Rahnavard, "Adaptive Non-uniform Compressive Sampling for Time-varying Signalss,"(Accepted) Annual Conference on Information Sciences and Systems (CISS'17), Baltimore, MD, March 2017.
2. B. Shahrasbi and N. Rahnavard, "Model-Based Nonuniform Compressive Sampling and Recovery of Natural Images Utilizing a Wavelet-Domain Universal Hidden Markov Model," IEEE Transactions on Signal Processing, vol. PP, no. 99, p. 1, 2016.
3. B. Shahrasbi, A. Talari, and N. Rahnavard, "TC-CSBP: Compressive sensing for time-correlated data based on belief propagation," in 201145th Annual Conference on Information Sciences and Systems, pp. 1–6,IEEE, 3 2011.
4. M. L. Malloy and R. D. Nowak, "Near-Optimal Adaptive Compressed Sensing," IEEE Transactions on Information Theory, vol. 60, pp. 4001–4012, 7 2014.
5. J. Haupt, R. Baraniuk, R. Castro, and R. Nowak, "Sequentially designed compressed sensing," in 2012 IEEE Statistical Signal Processing Workshop (SSP), pp. 401–404, IEEE, 8 2012.
6. G. Braun, S. Pokutta, and Y. Xie, "Info-Greedy Sequential Adaptive Compressed Sensing," IEEE Journal of Selected Topics in Signal Processing, vol. 9, pp. 601–611, 6 2015.
7. M. A. Iwen and A. H. Tewfik, "Adaptive Strategies for Target Detection and Localization in Noisy Environments," IEEE Transactions on Signal Processing, vol. 60, pp. 2344–2353, 5 2012.
8. M. A. Davenport and E. Arias-Castro, "Compressive binary search," in 2012 IEEE International Symposium on information Theory Proceedings, pp. 1827–1831, IEEE, 7 2012.
9. U. L. Wijewardhana and M. Codreanu, "A Bayesian Approach for Online Recovery of Streaming Signals from Compressive Measurements," IEEE Transactions on Signal Processing, pp. 1–1, 2016.
10. J. Ziniel and P. Schniter, "Dynamic Compressive Sensing of Time-Varying Signals Via Approximate Message Passing," IEEE Transactions on Signal Processing, vol. 61, pp. 5270–5284, 11 2013.
11. C. F. Mecklenbrauker, P. Gerstoft, A. Panahi, and M. Viberg, "Sequential Bayesian Sparse Signal Reconstruction Using Array Data," IEEE Transactions on Signal Processing, vol. 61, pp. 6344–6354, 12 2013.

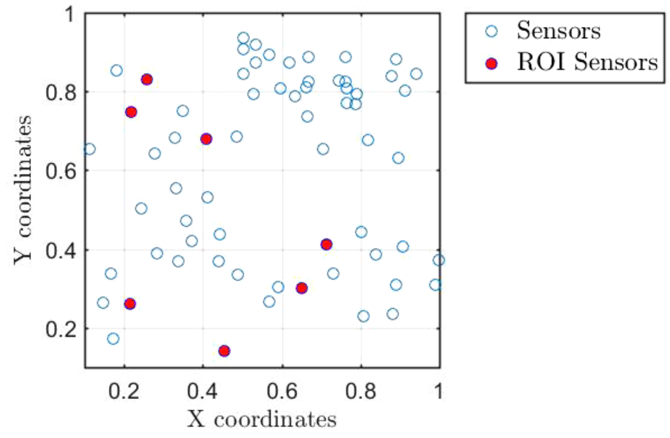
Results: Sparse Signal in Canonical Domain

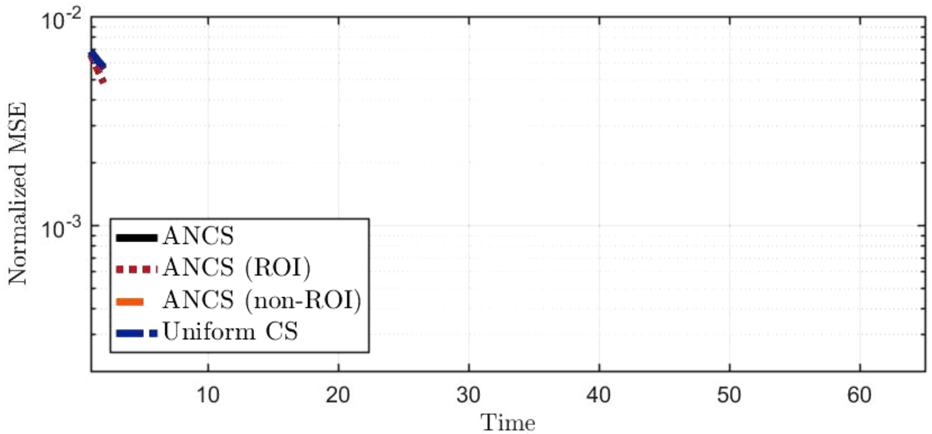
- A sparse signal contaminated with AWGN noise:

$$\boldsymbol{y}^{(t)} = \boldsymbol{\varphi}_{ANCS}^{(t)} \boldsymbol{x}^{(t)} + \boldsymbol{n}^{(t)}, \mathrm{t} = 1,2, \ldots, \mathrm{T}$$

- Two random processes describe the signal:
  1. Support of the signal ($s_n = 0 \; or \; 1$) described by a binary Markov chain process
     - defined by $p_{01} = P\{s_n^{(t)} = 1 | s_n^{(t-1)} = 0\}$ and $\lambda = P\{s_n^{(t)} = 1\}$
  2. Amplitude of the coefficients a_n, described by a Gaussian process with zero mean
     - $a_n^{(t)} = (1 - \rho) a_n^{(t-1)} + \rho \, v_n^{(t)}$
     - $v_n^{(t)} = N(0, \sigma_L^2)$
     - $\rho = 1$ (No correlation) / $\rho = 0$ (No variation)
- $x_n^{(t)} = s_n^{(t)} \times a_n^{(t)}$

Results: EPFL data set

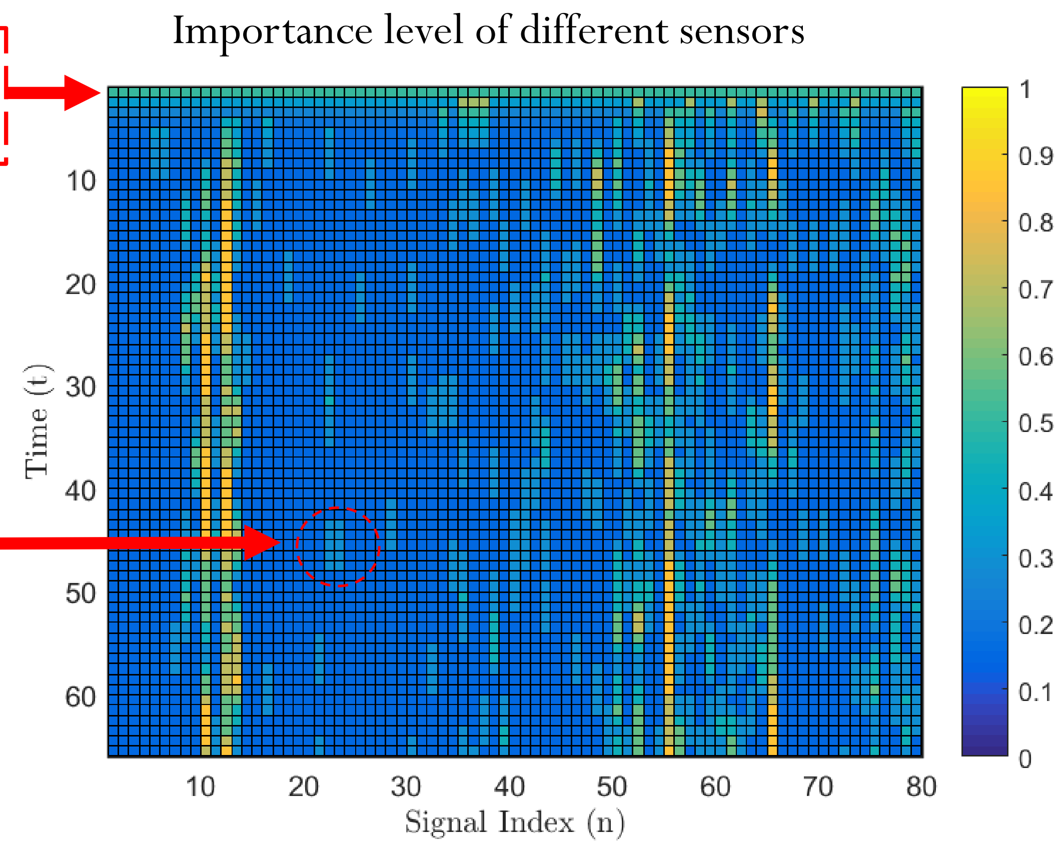




Results: EPFL data set

Starts with unbiased estimate of importance levels

False alarms can impact up to W = 5 time slots.



Error propagation, as well as computational complexity, are the main reasons that choosing large values for inference window should be avoided.