INSTITUTE OF COMPUTER SCIENCE

Dictionary Learning for Spontaneous Neural Activity Modeling

E. Troullinou[1,3], G. Tsagkatakis[1], G. Palagina[2],
M. Papadopouli[1,3], S.M. Smirnakis[2], P. Tsakalides[1,3]

Institute of Computer Science, FORTH[1]
Harvard Medical School[2]
Computer Science Department, UoC[3]

Outline

I. Introduction
II. Proposed Approach
III. Experimental Results
IV. Conclusion

What - Why - How

Work on: → Neuronal Activity Modeling

Modeling? → Dimensionality Reduction

How? → Dictionary Learning

Why? → High dimensionality → Challenge in signal analysis
Insights into brain workings
Capture only significant information

Innovative Aspects

Sparse Signal Modeling
+
Dictionary Learning

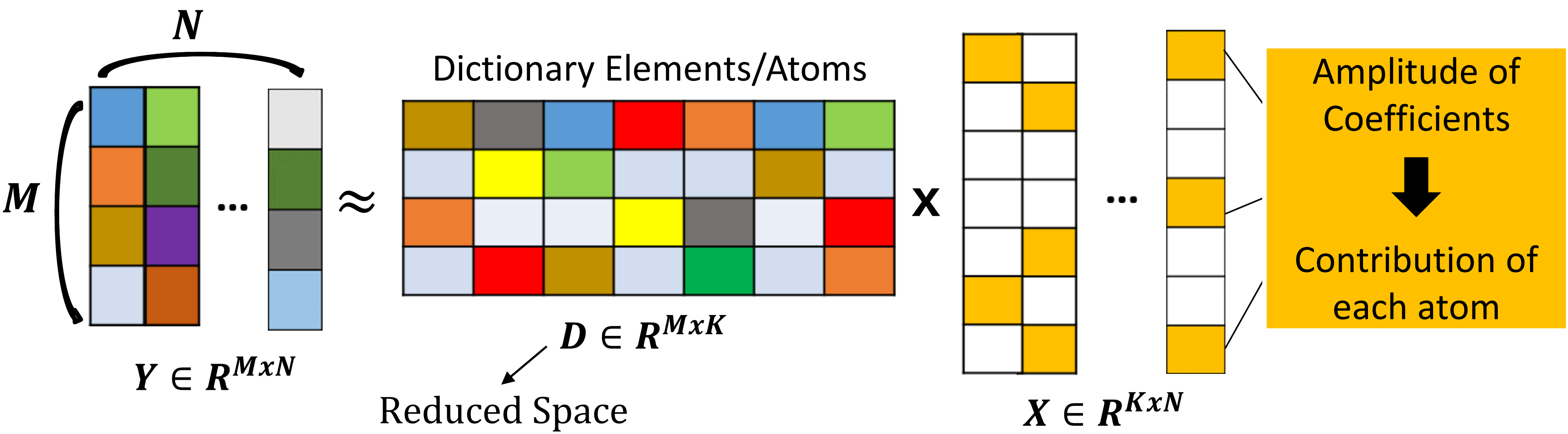
$\Big\}$ Real Neuronal Network Dataset

Generalization Capacity

Sensitivity to noise

$\Big\}$ Trained Dictionary

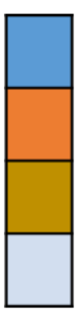
Sparse Signal Modeling



X (Coefficient Matrix) ➡ Sparse Matrix

Sparsity Level =
maximum # of non zero elements in every column (yellow boxes)

e. g.



=

1st raw 1st column

$$X_{11} * 1^{st}\ column\ of\ D + X_{61} * 6^{th}\ column\ of\ D$$

Goal: Go into a new reduced space that “summarizes” my input data!
Keep only useful information!

Why Imposing the Criterion of Sparsity?

1. Time Complexity
2. Avoid Overfitting

More General Examples

- If $\boldsymbol{Y}$ was an image, dictionary should capture:
- Important Edges
- Intensities

- If $\boldsymbol{Y}$ was a song, dictionary should capture:
- Basic notes and their combinations

Norms

Norms:

$$zero\ norm:\ \ \|x\|_0 = (i|x_i \neq 0)$$

$$l_1\ norm:\ \|x\|_1 = \sum_{i=1}^{R} |x|_i$$

$$l_2\ norm:\ \|x\|_2 = \ (\sum_{i=1}^{R} |x|_i^{\ 2})^{1/2}$$

$$l_p\ norm:\ \|x\|_p = (\sum_{i=1}^{R} |x|_i^{\ p})^{1/p}$$

$$Frobenius\ norm\ (matrix\ norm):\ \|x\|_F = \sqrt{\sum_{i=1}^{M} \sum_{j=1}^{N} \left|x_{ij}\right|^2}$$

Orthogonal Matching Pursuit (OMP)

$\boldsymbol{Optimization\ Problem - Sparse\ Coding}$:

$$\min_{x_i} \|\boldsymbol{y_i} - \boldsymbol{D}\boldsymbol{x_i}\|_2^2 \quad subject\ to \quad \|x_i\|_0 \ \leq\ T_0\ \ \forall i, \qquad where$$

y_i: input vector
D: trained dictionary
x_i: coefficient vector
$\|.\|_0$: zero norm
T_0: Sparsity Level

OMP (Orthogonal Matching Pursuit)

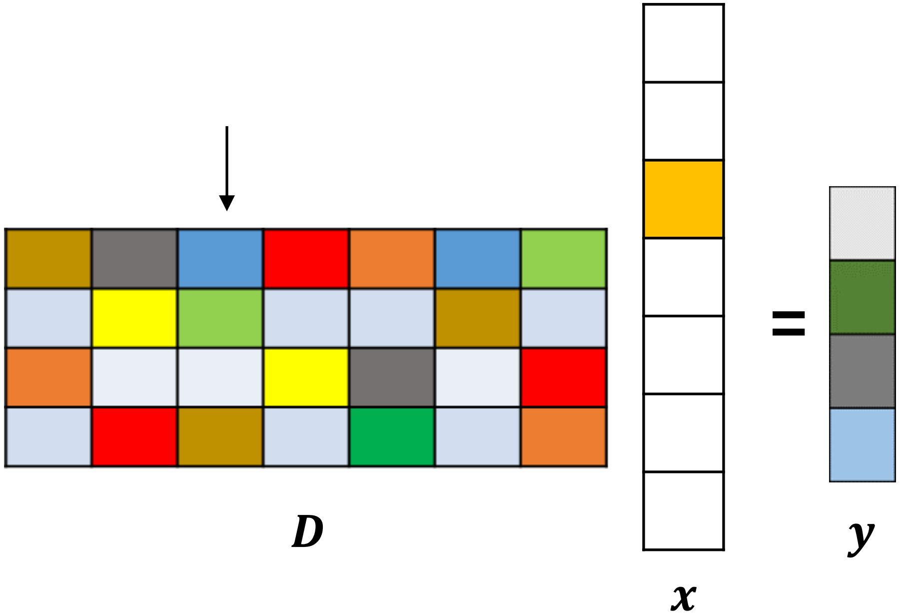
<u>Basic Idea:</u> Approximately represent a signal y as a weighted sum of finitely many functions d_j (dictionary elements) taken from D. For an approximation with N dictionary elements:

$$y = \sum_{j=1}^{N} x_j d_j \quad , \qquad where$$

x_j is the scalar weighting factor (coefficient) for d_j

OMP Visualization

1. Set the residual $\mathrm{r}_1 = y$
   Iteratively: Set j=1
2. Find an unselected atom that best matches the residual $\left\| r^j - Dx \right\|$
3. Get the coefficient of x
4. Recalculate the residual from matched atoms $r^{j+1} = r^j - Dx$
5. Repeat until $\left\| r^j \right\| \leq \in$
6. j = j+1



$$r_1 = y$$
$$r_2 = r_1 - x_j d_j = y - x_j d_j$$
$$r_3 = r_2 - x_{j'} d_{j'} = y - x_j d_j - x_{j'} d_{j'}$$
$$\vdots$$
$$r_n = y - \sum_{j=1}^{N} x_j d_j$$

Dictionary Learning

The Dictionary can be one of the following:

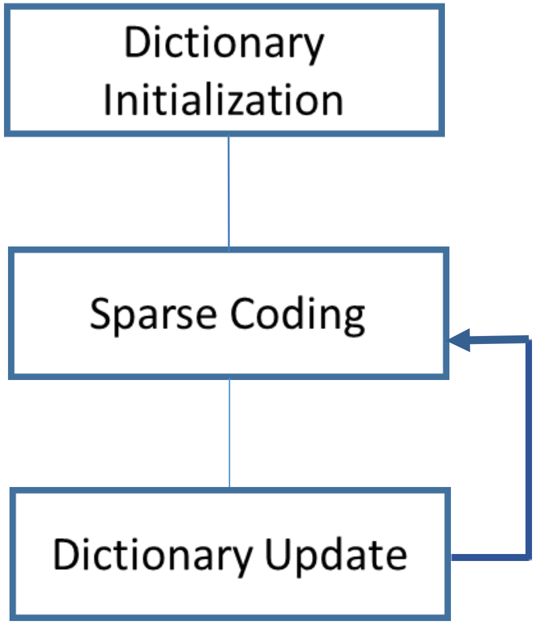
- Parametric: Fourier signals, wavelets, etc.
- Trained: Learning from randomly selected input examples

K-SVD Algorithm

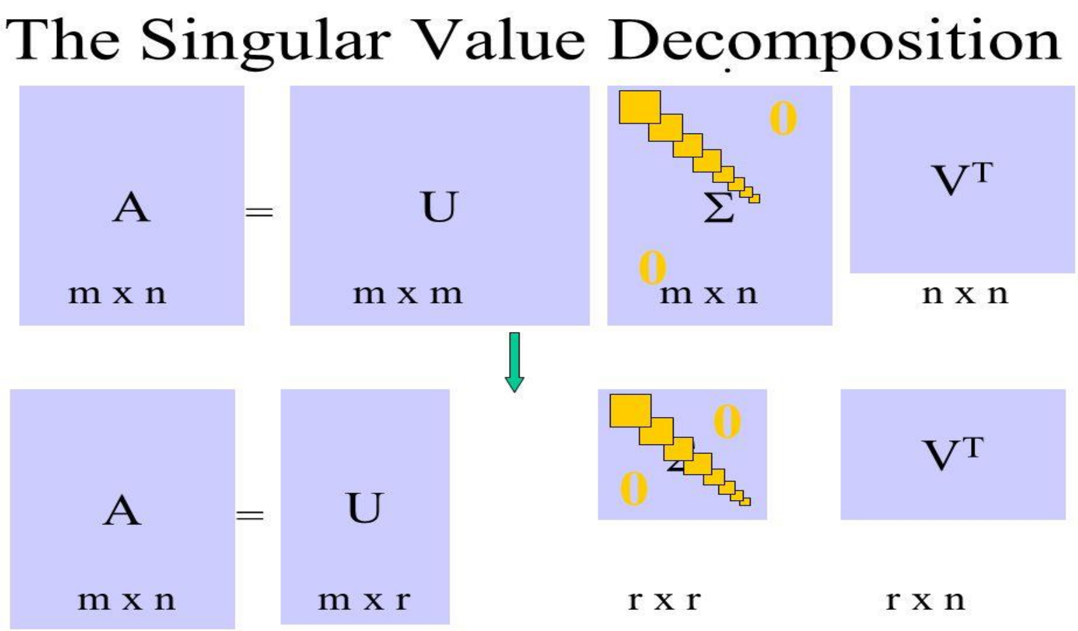
For Sparsity Level T_0, K-SVD solves the following:

$$\min_{D,X} \|Y - DX\|_F^2 \text{ subject to } \|X_i\|_0 \leq T_0 \quad \forall i, \text{ where}$$

Y: Input Signal
D: Trained Dictionary
X: Coefficient Matrix
$\|.\|_F$ denotes the Frobenius norm
T_0: sparsity level



Singular Value Decomposition (SVD)



A singular value and a pair of singular vectors of a matrix A are a nonnegative scalar σ and two nonzero vectors u and v s.t.:

$$Au = \sigma v$$
$$A^H v = \sigma u$$

A^H denotes the complex conjugate transpose of a matrix

U is an mxm orthogonal matrix
Σ is a diagonal mxn matrix with non-negative real numbers in the diagonal (singular values of A)
V is an nxn orthogonal matrix
V^T is the transpose of V

Dictionary Update

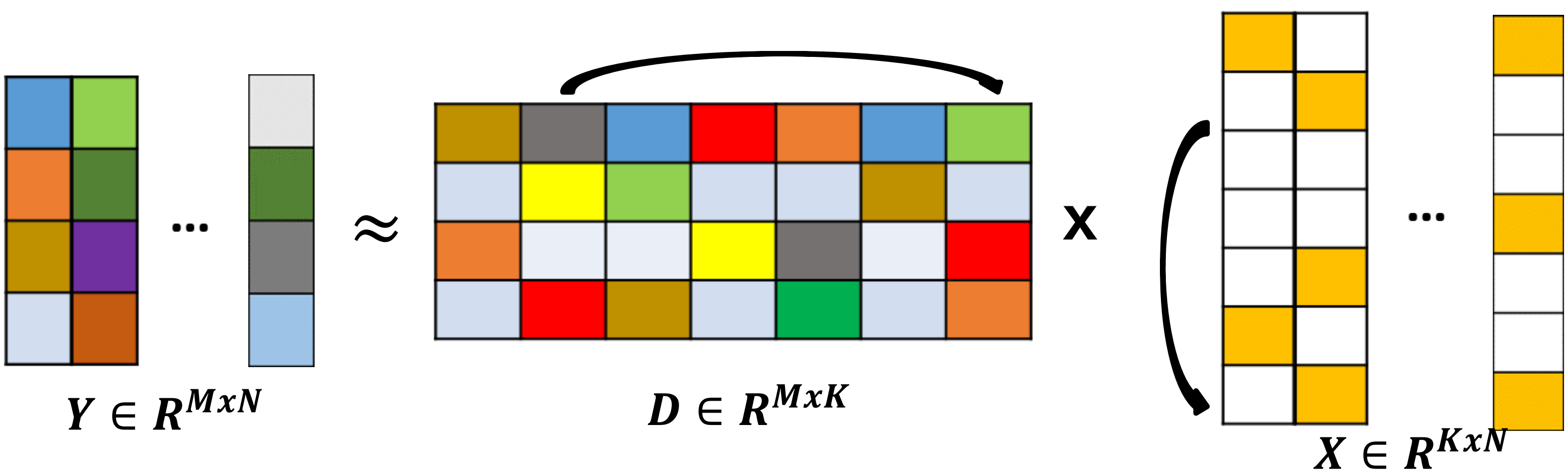
For each column k = 1, 2, . . . , K in D update it by

- Define the group of examples that use this atom, $\omega_\kappa = \{i | 1 \le i \le N, x_T^k(i) \neq 0\}$
- Compute the overall representation error matrix, E_k, by $E_k = Y - \sum_{j \neq k} d_j x_T^j$
- Restrict E_k by choosing only the columns corresponding to ω_κ, and obtain E_k^R .
- Apply SVD decomposition $E_k^R = U \Sigma V^T$. Choose the updated dictionary column d_k' to be the first column of U.

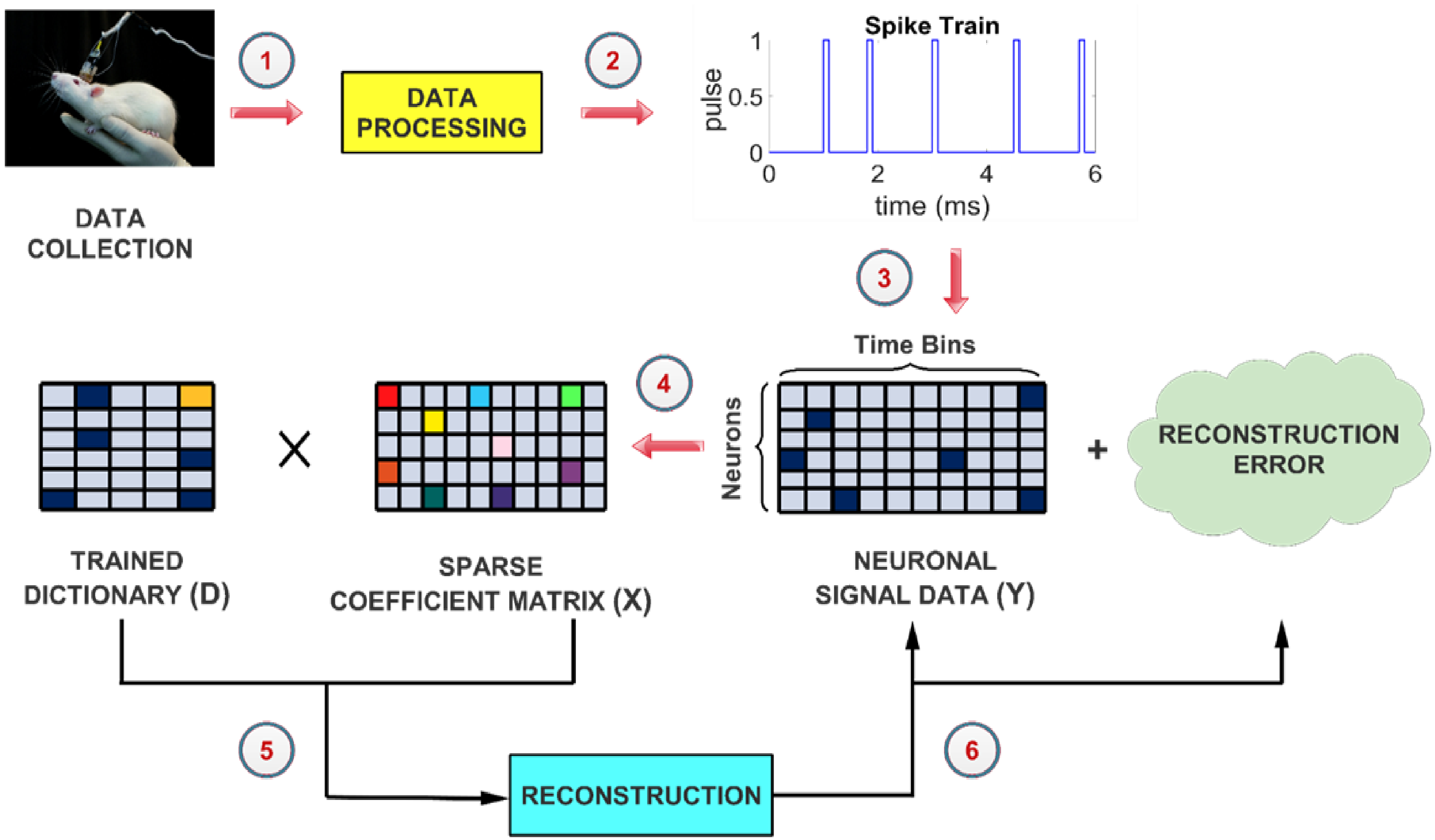
e.g. k=1 (1st Dictionary Element)

$$\omega_1 = \{1, N\}$$

$$E_1 = Y - \sum_{j \neq 1} d_j x_T^j$$



Proposed Dictionary Learning Framework



Dataset

Real Binary Dataset

1s are spiking events

0s are non-spiking events

Neural Activity of 9-day old mouse

29 min. spontaneous activity (11970 frames of 0.1451 sec.)

Time instances

Neurons id

	t_1	t_2	t_3	...	t_{11970}
1	0	1	0		0
.	1	0	1		0
.	0	0	0		0
.	0	1	0		1
183	1	0	0		0

Training Examples **Testing Examples**

Spiking Events: 0.36% of the dataset

K-SVD Performance - Parameters

K-SVD Performance in neuronal signal reconstruction in terms of:

i. Dictionary size

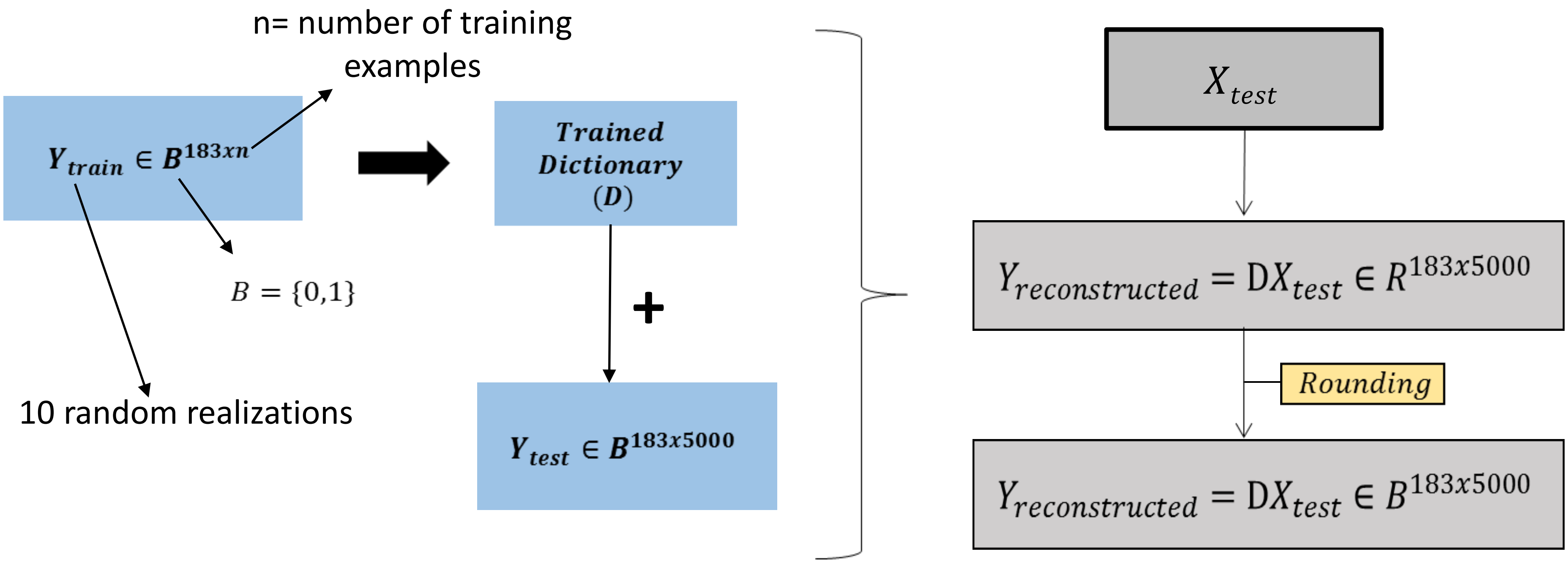
ii. Sparsity level

iii. Training size used for dictionary learning

Can we achieve a good reconstruction?

Can we model the data?

Experimental Setup

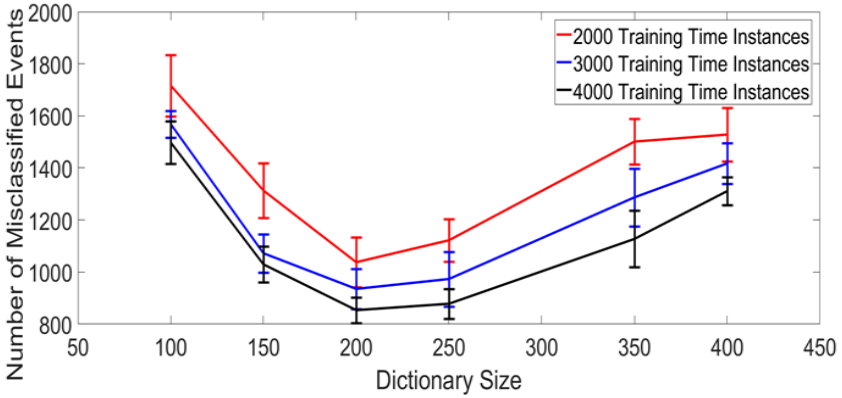


Number of misclassified Events = # $\{Y_{test} \neq Y_{reconstructed}\}$

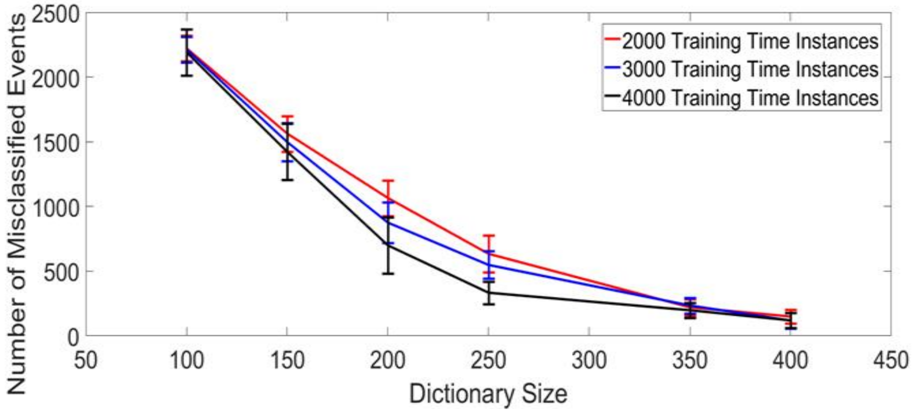
Impact of the Examined Parameters

Total number of events = 915000 (183x5000)

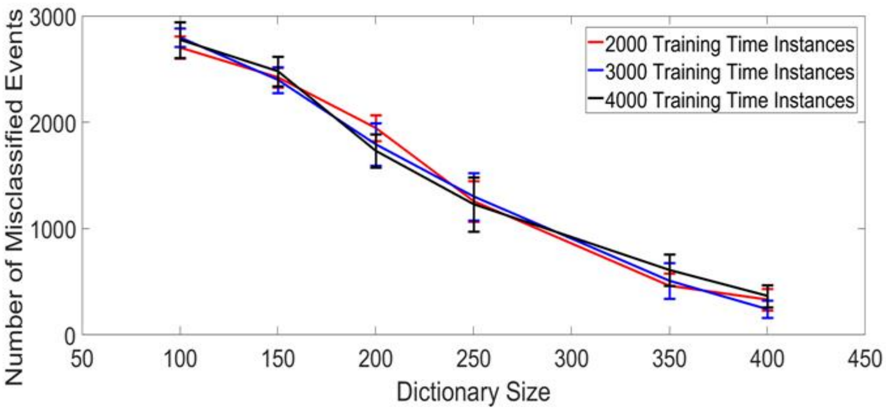
Sparsity Level = 4



Sparsity Level = 20



Sparsity Level = 50



Confusion Matrix of the Reconstructed Events

Dictionary Size = 400

Training Examples		2000		4000	
Sparsity	Predicted / Actual	0	1	0	1
4	0	99.982	0.017	1	0
	1	58.914	41.085	46.636	53.363
20	0	**99.997**	0.00241	**99.998**	0.00120
	1	19.668	**80.331**	18.491	**81.508**
50	0	99.982	0.017	99.998	0.00186
	1	32.639	67.36	52.82	47.179

Trained Dictionary – Sensitivity to noise

Trained Dictionary (D)

	t_1	t_2	t_3	…	t_m
n_1	0	0	1	…	0
.	0	1	0	…	0
.	0	1	0	…	1
.	1	0	0	…	0
n_{183}	0	0	1	…	0

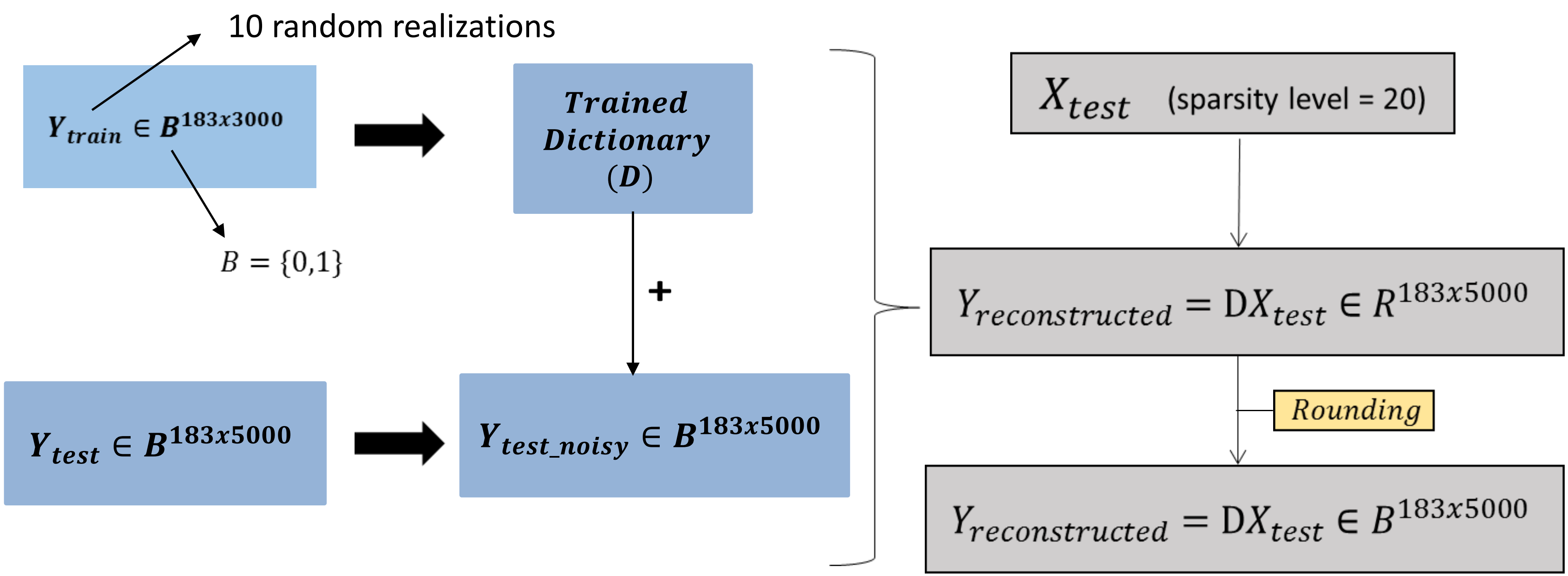
True Patterns

	t_1	t_2	t_3	…	t_m
n_1	1	0	1	…	0
.	0	1	1	…	0
.	0	1	1	…	1
.	0	1	0	…	0
n_{183}	1	0	1	…	0

Noisy Signal

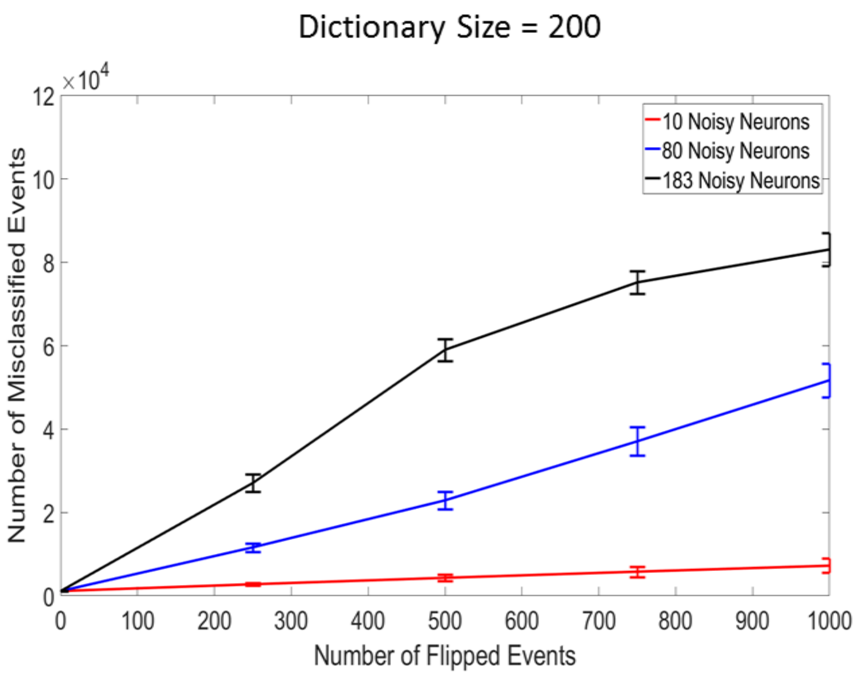
True Patterns and Noisy Signal both well modeled?

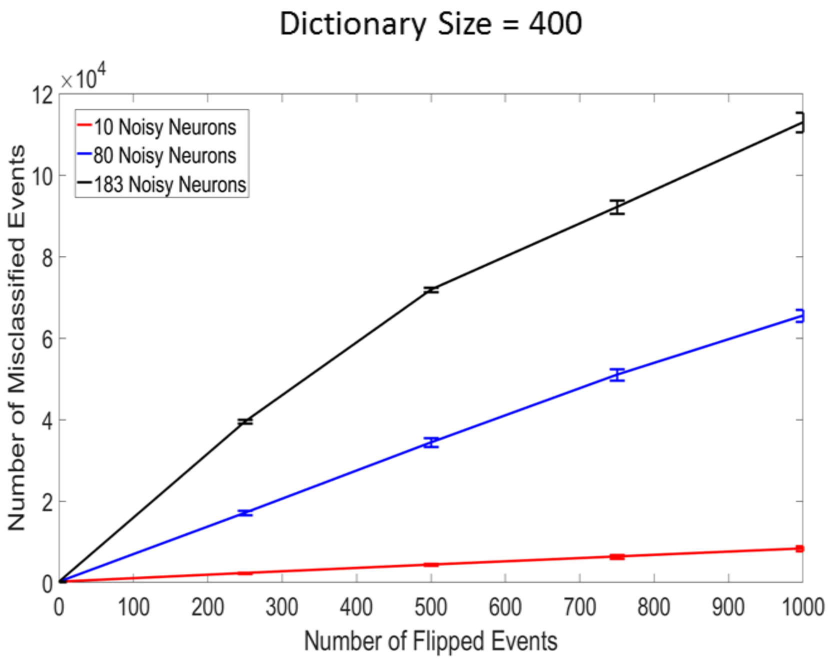
Experimental Setup



Number of misclassified Events = $\# \{Y_{test\_noisy} \neq Y_{reconstructed}\}$

Total number of events = 915000 (183x5000)





Number of misclassified Events = $\# \{ Y_{\text{test\_noisy}} \neq Y_{reconstructed} \}$

Conclusion – Concerns – Future Work

- Neuronal Signals – Dictionary Learning
  - Effectively represented in low-dimensional subspaces
  - Small training size is adequate
  - Average Sparsity level

- Dictionary – Sensitivity to noise
  - Random noise is not modeled

- Concerns - Future Work
  - Deal with other types of noise (e.g. Circularly Shifted Events)
  - Find a suitable metric to check the consistency of the dictionary
  - Focus on qualitative characteristics of the Dictionary
  - Adversarial Learning Methods

ACKNOWLEDGMENT

This work was partially funded by the **DEDALE** project, contract no. 665044, within the H2020 Framework Program of the European Commission.