Center for Energy & Geo Processing

CeGP

KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

1963

Georgia Tech

Advanced Methods for Seismic Data Compression

M. Mohandes

Outline

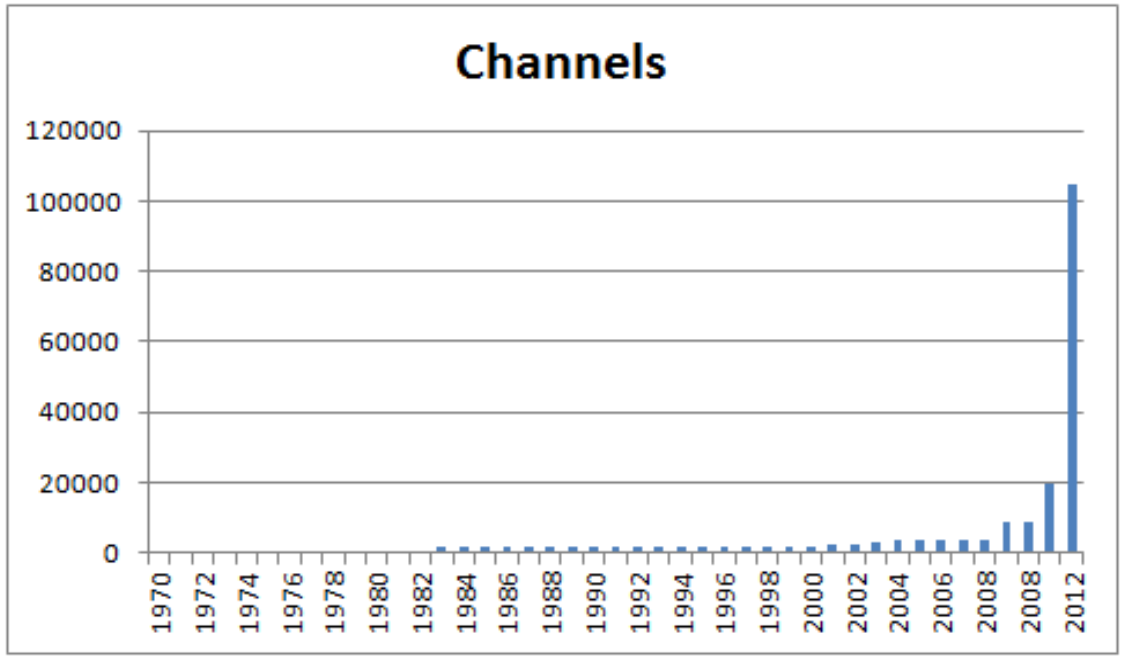
Compression Approaches:

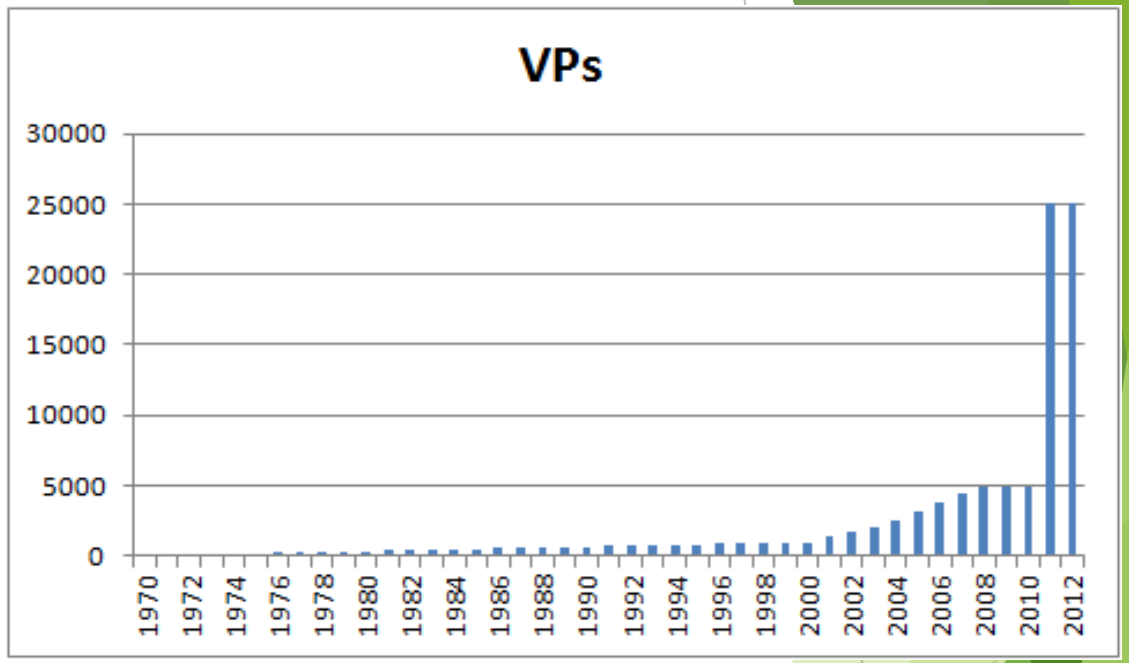
- Distributed PCA Compression
- Deep Learning Compression
- Model-based Compression
- Vector quantization for compression
- Rate Distortion curve

- Accomplishments
- Future works.

- Team Members: M. Deriche, S. Zummo, B. Liu, S. Muhammadi,
- PhD students: H. Nuha, H. Khan

Seismic Data

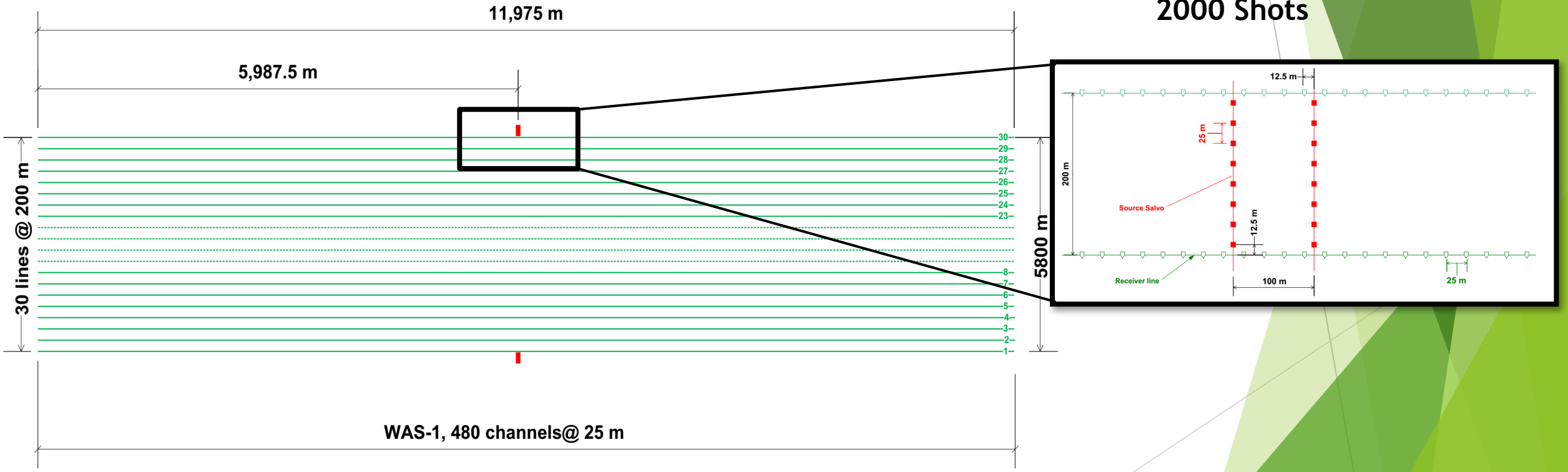




- The scale of seismic survey has been increasing drastically
- More seismic data are generated

A Typical Seismic Survey

Source Point interval	**25 m**	Source Line interval	**100m**	Channels per line	**480**
Receiver Point interval	**25 m**	Receiver Line interval	**200m**	30 RLs x 480 CH =	**14,400**



Seismic data & compression

- In one swath, there 30 Rows x 480 CH = 14,400 CHs
- More than 2,000 shots are generated per day
- More than 14,400 x 2,000=28,800,000 traces are recorded per day
- Several TBs data needed to be stored and transferred per day
- Compression is highly needed

Approaches for Seismic Data Compression

- **Distributed PCA (DPCA) Based Compression**

  Optimal transformation to reduce the dimensionality of the data

- **Deep Machine Learning Based Compression;**

  Recognition of patterns using neural networks.

- **Model-based Compression**

  Decomposition as wavelets, then parameters estimation of the wavelets.

- **Vector Quantization for compression**

Principal Component Analysis (PCA)

PCA is a linear transformation to decrease the dimension of the data.

Let X be the original data (vector) and Y be the compressed data (vector):

$$Y = P^T X$$

$$\dim(Y) < \dim(X)$$

where P is called transform basis, constructed by the most important n eigenvectors (Principal Components, PCs) of the covariance matrix of the original data.

$$P = [p_1, p_2, \dots, p_n]$$

DPCA Motivation:

- The local PCA (LPCA) considers the seismic data of individual sensors, requiring each sensor to generate a transform basis.
- A practical seismic sensor network usually consists of tens of thousand of sensors.
- DPCA generates a universal global transform basis for all sensors, leading to a higher compression ratio and lower computation cost.

DPCA Methodology (1):

Sensor i records N_i traces:

$$X_j^{(i)} \in R^n, \quad j=1,2,\ldots,N_i \tag{1}$$

which are regarded as N_i realizations of an unknown local PDF $f_i(x)$:

$$\left\{X_j^{(i)}\right\}_{j=1}^{N_i} \sim f_i(x) \tag{2}$$

The first two moments of $f_i(x)$ can be estimated as

$$\mu_i = \frac{1}{N_i}\sum_{j=1}^{N_i} X_j^{(i)} \quad \Sigma_i = \frac{1}{N_i-1}\sum_{j=1}^{N_i}(X_j^{(i)} - \mu_i)(X_j^{(i)} - \mu_i)^T \tag{3}$$

DPCA Methodology (2)

Suppose that there are K sensors in the network, the global PDF of all recorded traces is

$$f(x) = \Sigma_{i=1}^{K} \omega_i f_i(x), \text{ where } \omega_i = \frac{N_i}{\sum_{i=1}^{K} N_i} \quad (4)$$

The first two moments of the mixture model are

$$\mu = \Sigma_{i=1}^{K} \omega_i \mu_i \quad (5)$$

$$\Sigma = \Sigma_{i=1}^{K} \omega_i \Sigma_i + \Sigma_{i=1}^{K} \omega_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (6)$$

DPCA Methodology

Equivalence:

Centralized: $$\Sigma = \frac{1}{N-1}\sum_{i,j=1}^{K,N_i}(X_j^{(i)} - \mu)(X_j^{(i)} - \mu)^T \quad (7)$$

Distributed: $$\Sigma = \Sigma_{i=1}^{K}\omega_i\Sigma_i + \Sigma_{i=1}^{K}\omega_i(\mu_i - \mu)(\mu_i - \mu)^T \quad (8)$$

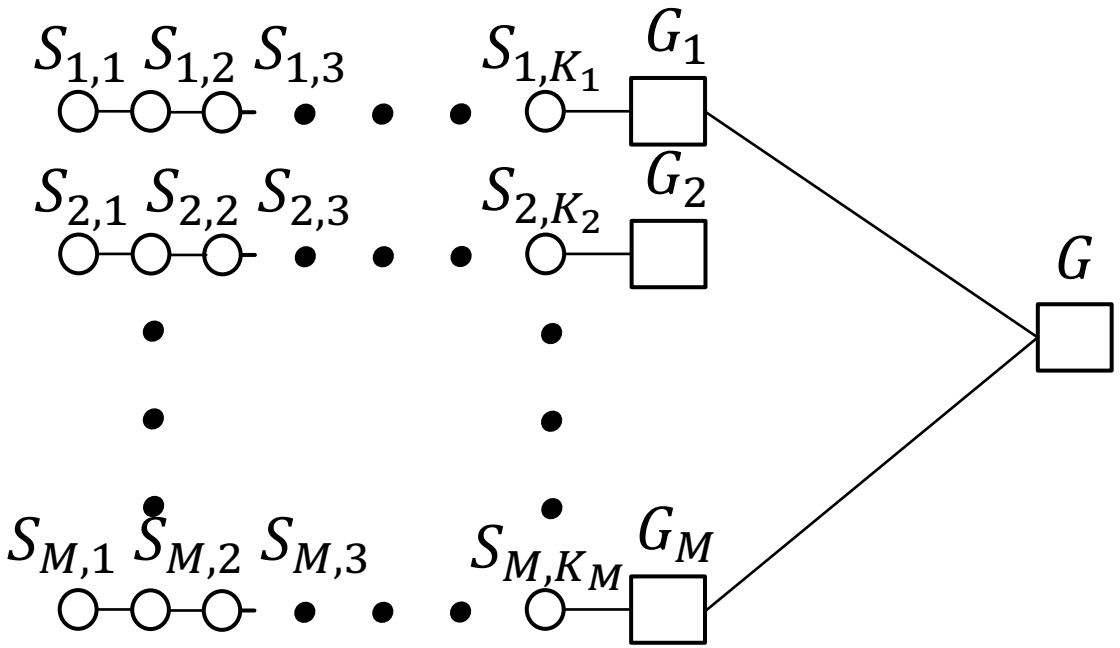
DPCA Scheme:

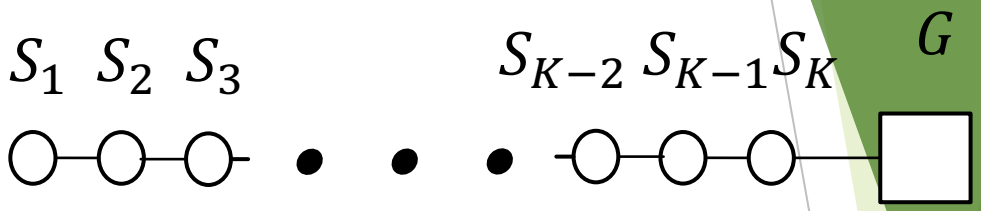
1. Sensor $S_i \Rightarrow \{N_i, \mu_i, \Sigma_i\} \Rightarrow$ fusion center G
2. Fusion centre G calculates Σ to determine global PCs $\{p_h\}_{h=1}^{k}$
3. Fusion centre G$\Rightarrow \{p_h\}_{h=1}^{k} \Rightarrow$ the sensors

DPCA Highlights:

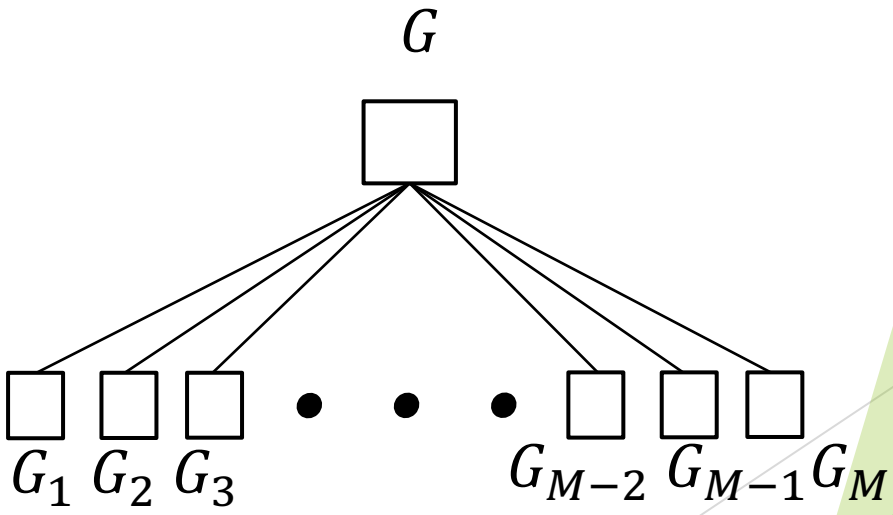
- Sending the local parameters to the fusion centers leads to a lower communication cost;
- DPCA generates a global transform basis for all sensors, leading to a higher compression ratio and lower computation cost.

Seismic network decomposition



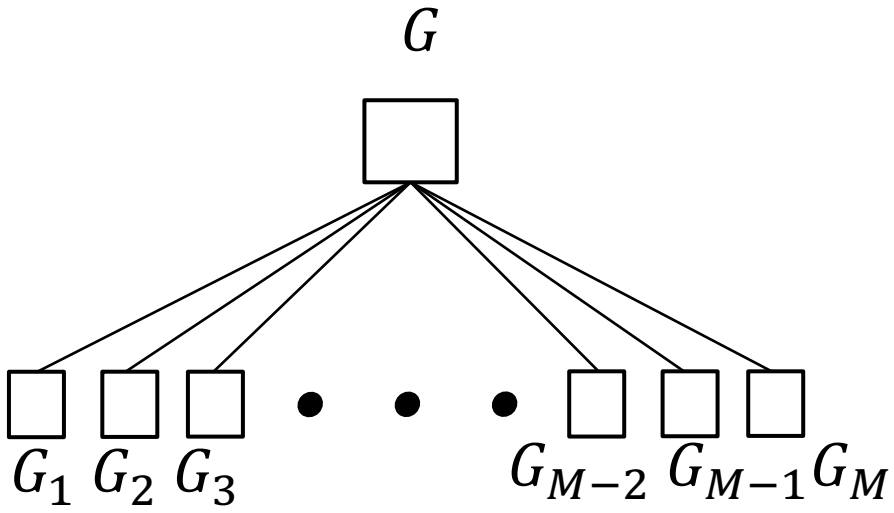


(a) *Cascade Connection*



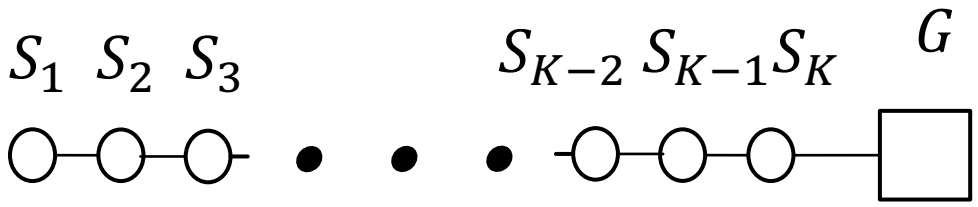
(b) *Star Connection*

Star connection



- Local center $G_r \Rightarrow \{N_r, \mu_r, \Sigma_r\} \Rightarrow$ fusion center G
- Fusion center G determines and sends back the global transform basis to the local centers

Cascade connection



- Direct method:
  - Sensor S_i sends whatever it receives from S_{i-1} and its own statistics $\{N_i, \mu_i, \Sigma_i\}$ to S_{i+1}

A better method for cascade connection is developed

Cascade connection

Rewrite global statistic (5) and (6) as :

$$\mu = \frac{1}{\mathrm{N}} \sum_{i=1}^{K} N_i \mu_i$$

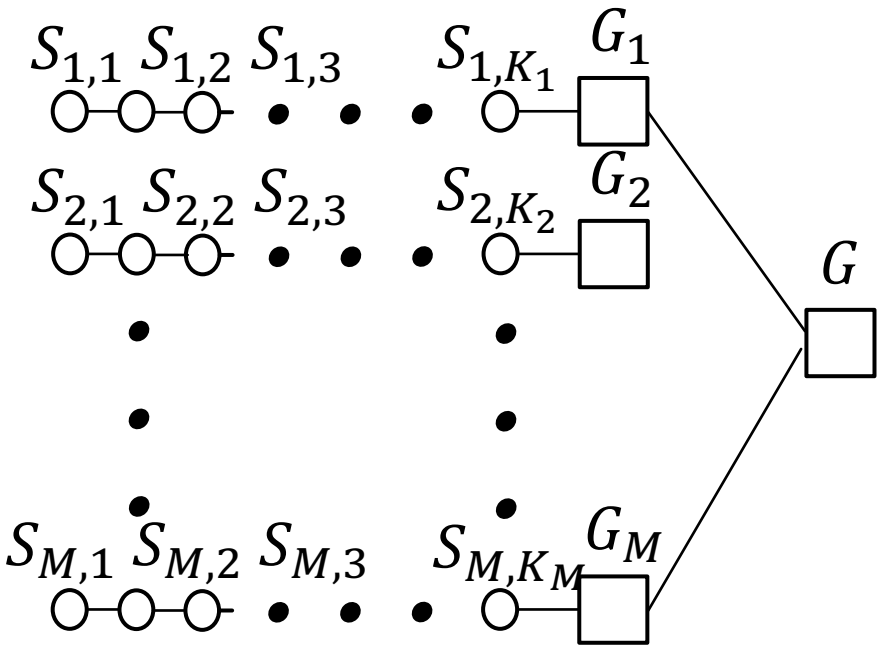
$$\Sigma = \frac{1}{N}\left(\sum_{i=1}^{K} N_i \Sigma_i + \sum_{i=1}^{K} N_i \mu_i \mu_i^T + \left(\sum_{i=1}^{K} N_i \mu_i \right) \mu^T - \mu \left(\sum_{i=1}^{K} N_i \mu_i \right)^T \right)$$

A better scheme for cascade connection:

1. $S_{i-1} \Rightarrow \left\{ \sum_{j=1}^{i-1} N_j , \sum_{j=1}^{i-1} N_j \mu_j , \sum_{j=1}^{i-1} N_j \Sigma_j , \sum_{j=1}^{i-1} N_j \mu_j \mu_j^T \right\} \Rightarrow S_i$
2. $S_i \; updates \left\{ \sum_{j=1}^{i} N_j , \sum_{j=1}^{i} N_j \mu_j , \sum_{j=1}^{i} N_j \Sigma_j , \sum_{j=1}^{i} N_j \mu_j \mu_j^T \right\}$
3. $S_i \Rightarrow$ the updated package $\Rightarrow S_{i+1}$

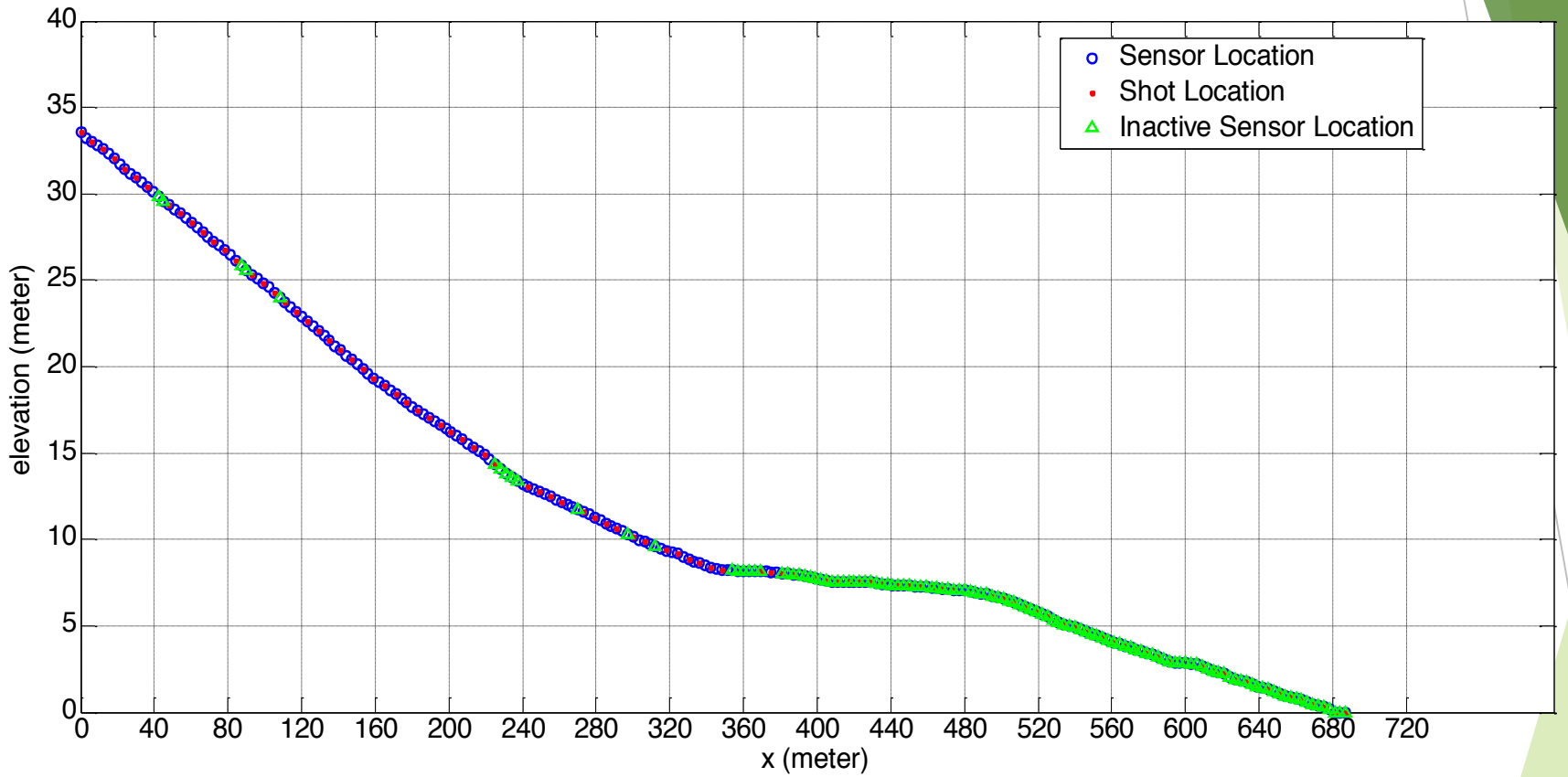
Benefit: The size of data package does not increase throughout the transmission.

DPCA Summary



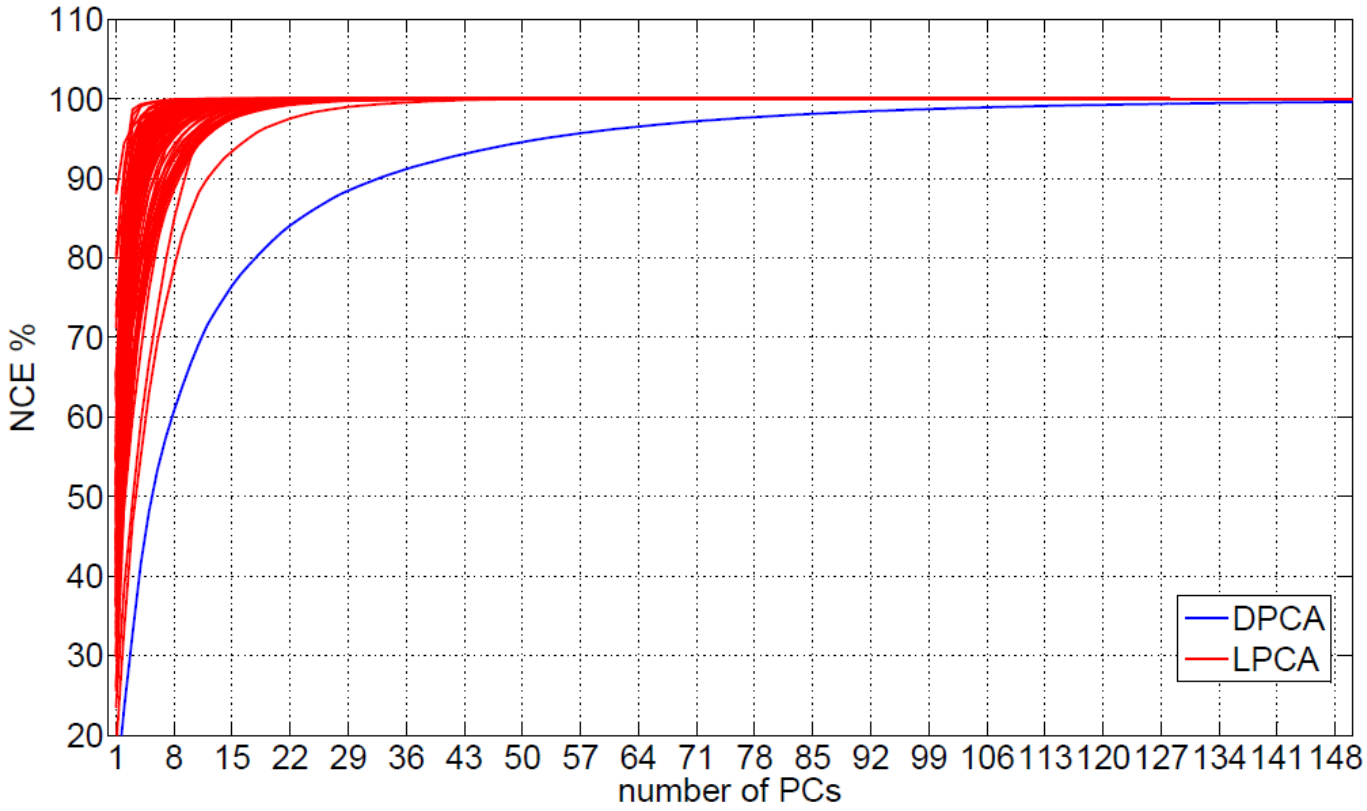
- for each line, the sensors pass an updated statistics to the local center G_r
- local centers G_r calculate the μ_r and Σ_r of lines
- local centers G_r send $\{N_r, \mu_r, \Sigma_r\}$ to center G
- center G calculates the global μ and Σ
- center G finds and sends back PCs
- Sensors project their data and send them to center G

UTAM Seismic Data



- Total 10,000 traces, 4,000 time samples
- 100 sensors (3 m sensor interval), 100 shots
- LPCA & DPCA preserve 99% NCE (20dB of SNR)

DPCA Results on UTAM real dataset:



k_D =109 PCs, k_L =12.30 PCs. Compression ratio r_D = 27:1 , r_L = 7:1

LPCA requires 1341 PCs totally, DPCA requires 109 PCs.

Lower computation cost and higher compression ratio

Reconstructed Trace

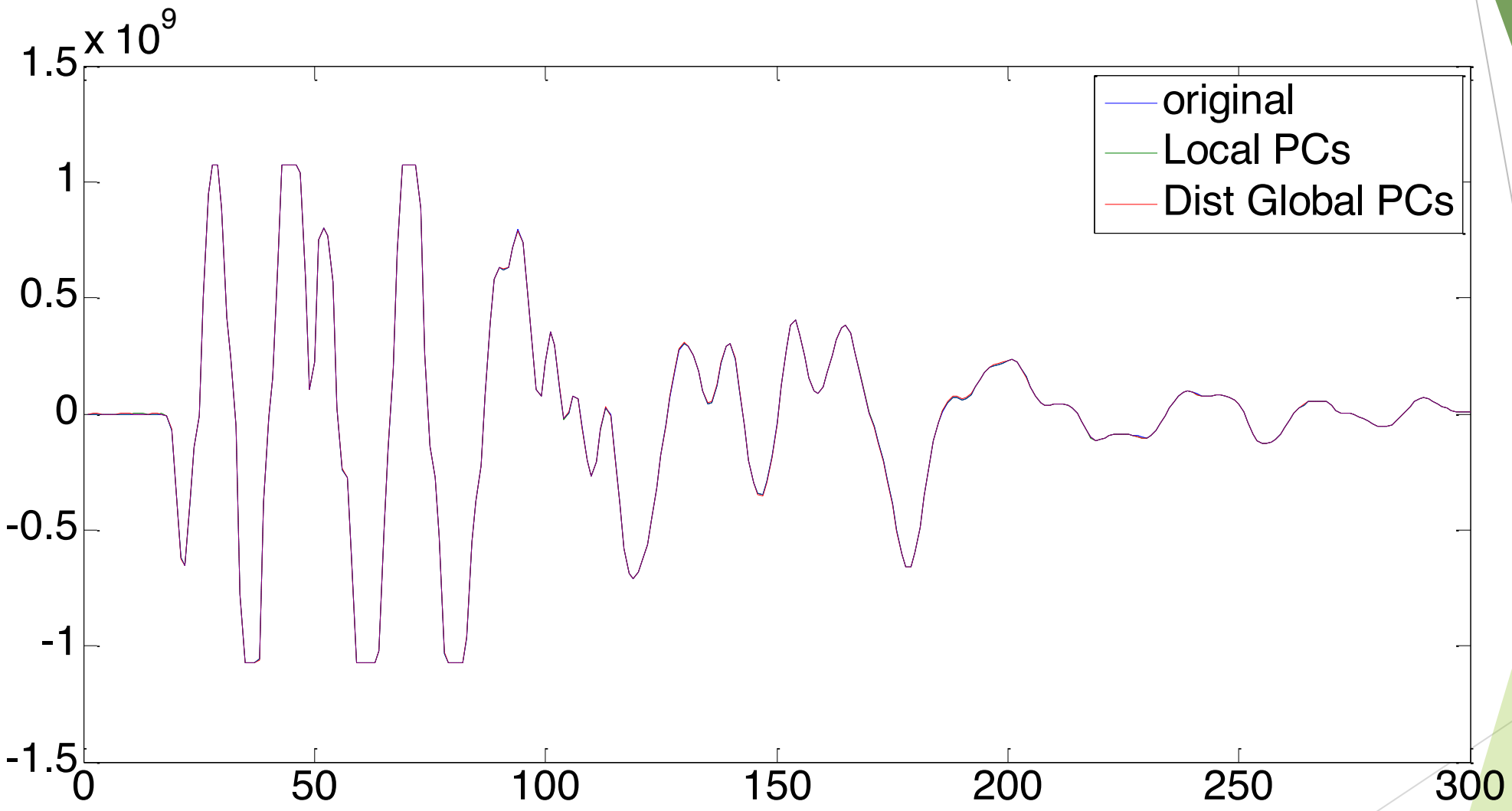


Figure: 5. A sample trace

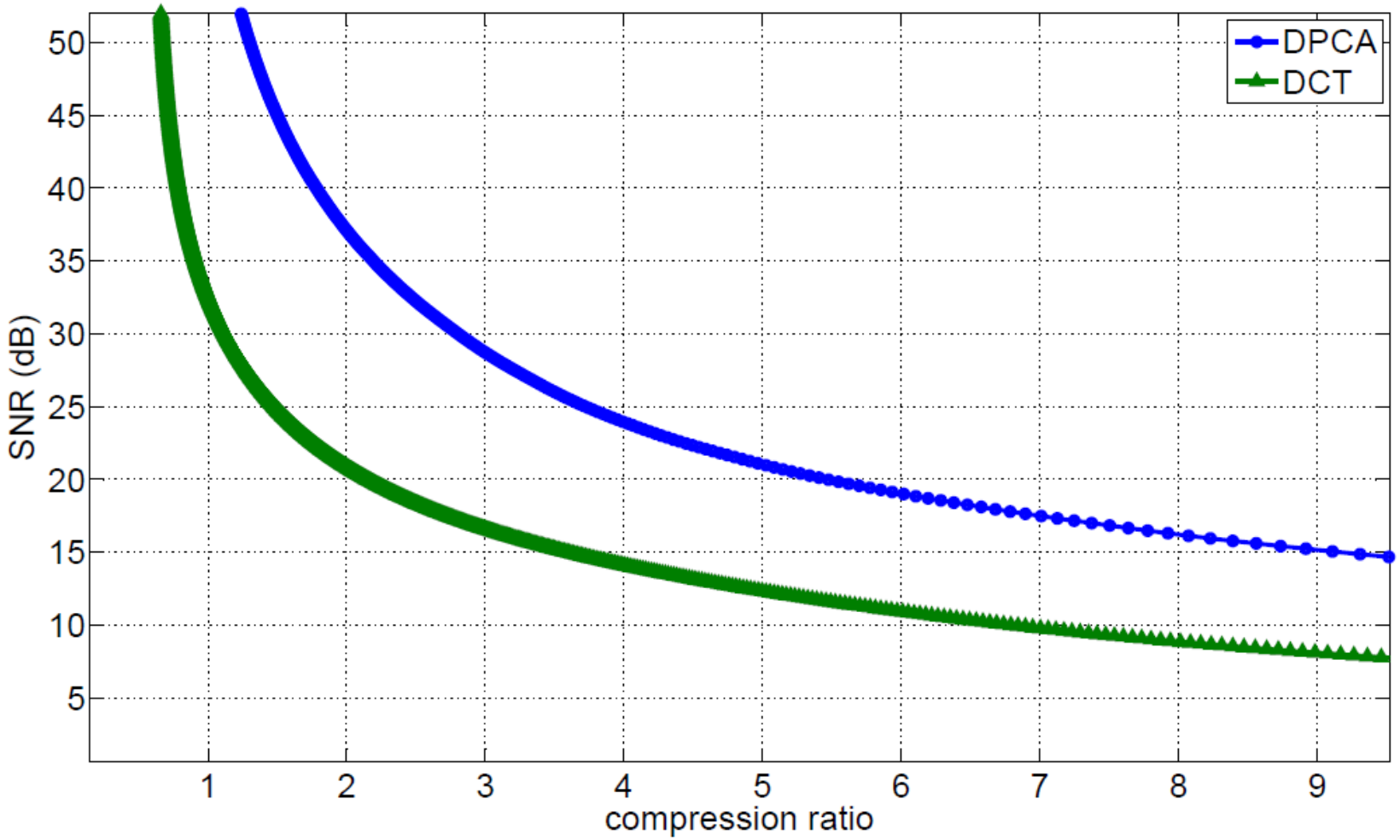
Comparison of DPCA with Discrete Cosine Transform (DCT) Compression

- DCT is a transform-based compression method
- DCT projects the seismic data on frequency domain
- Low frequency coefficients are stored, and high frequency coefficients are discarded
- No need to store the transform basis

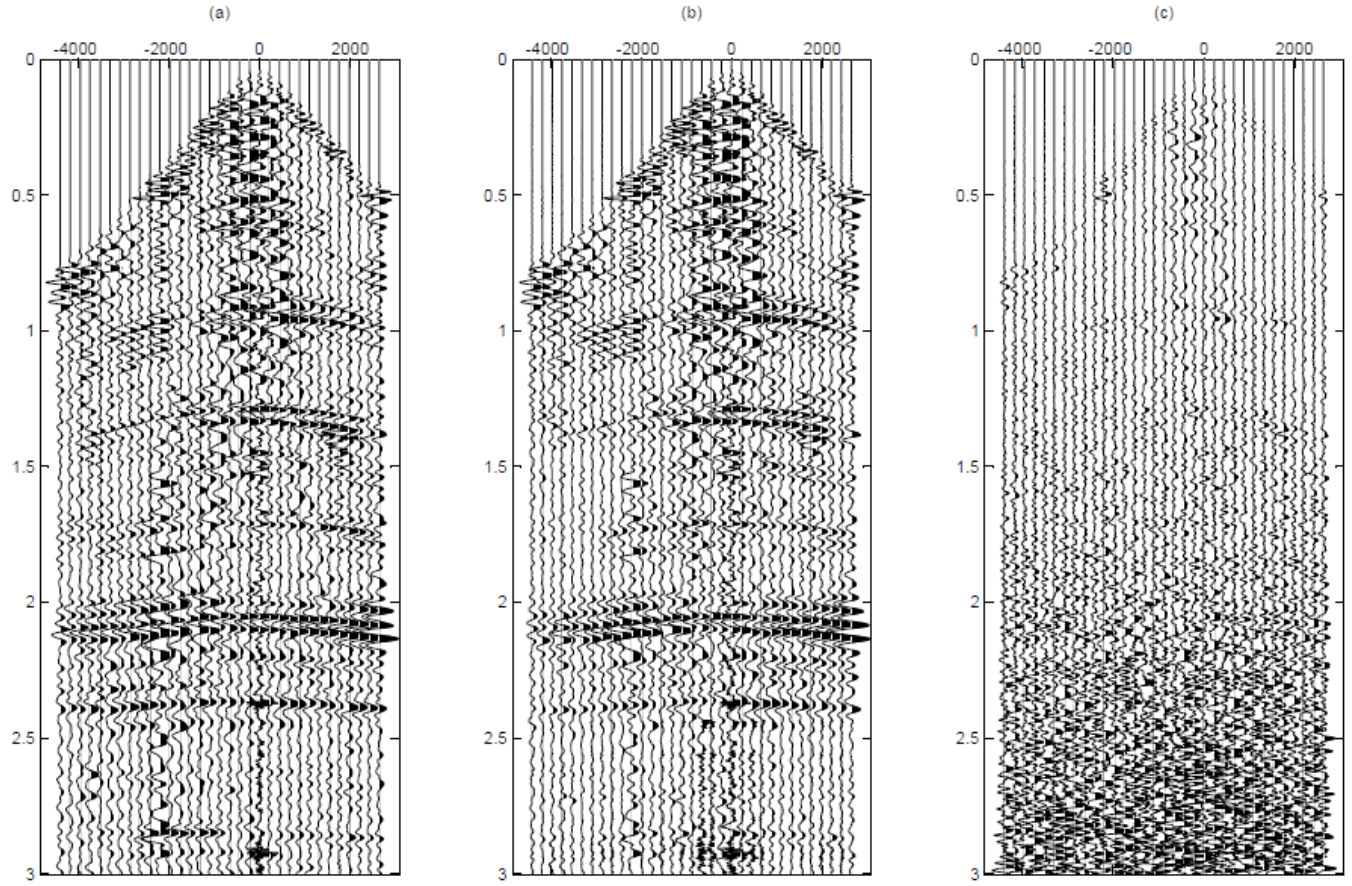
East Texas USA Database

- 10 sensors
- 19 shots
- 1501 time samples per trace

Performance of DPCA vs DCT:



Reconstructed traces DPCA vs DCT:



(a) Original (b) DPCA (c) DCT

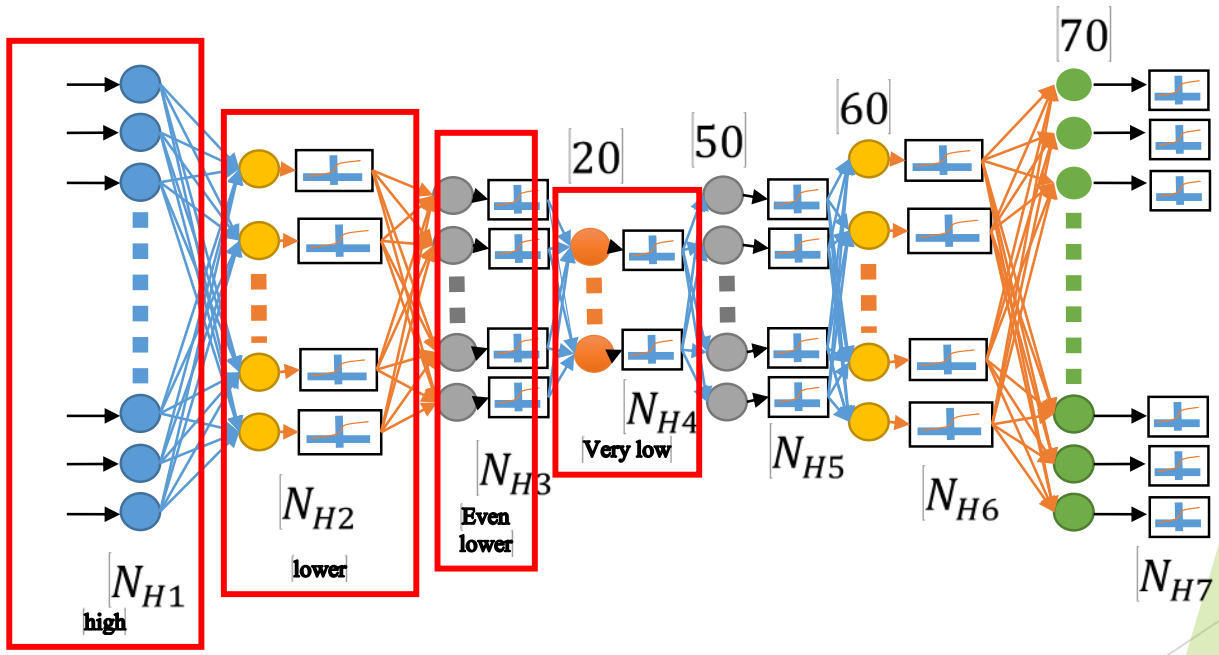
5:1 compression ratio

DPCA Conclusion:

- Higher Compression Ratio
- Less Computation Burden
- Theoretically Extendable
- Two-way communication of data

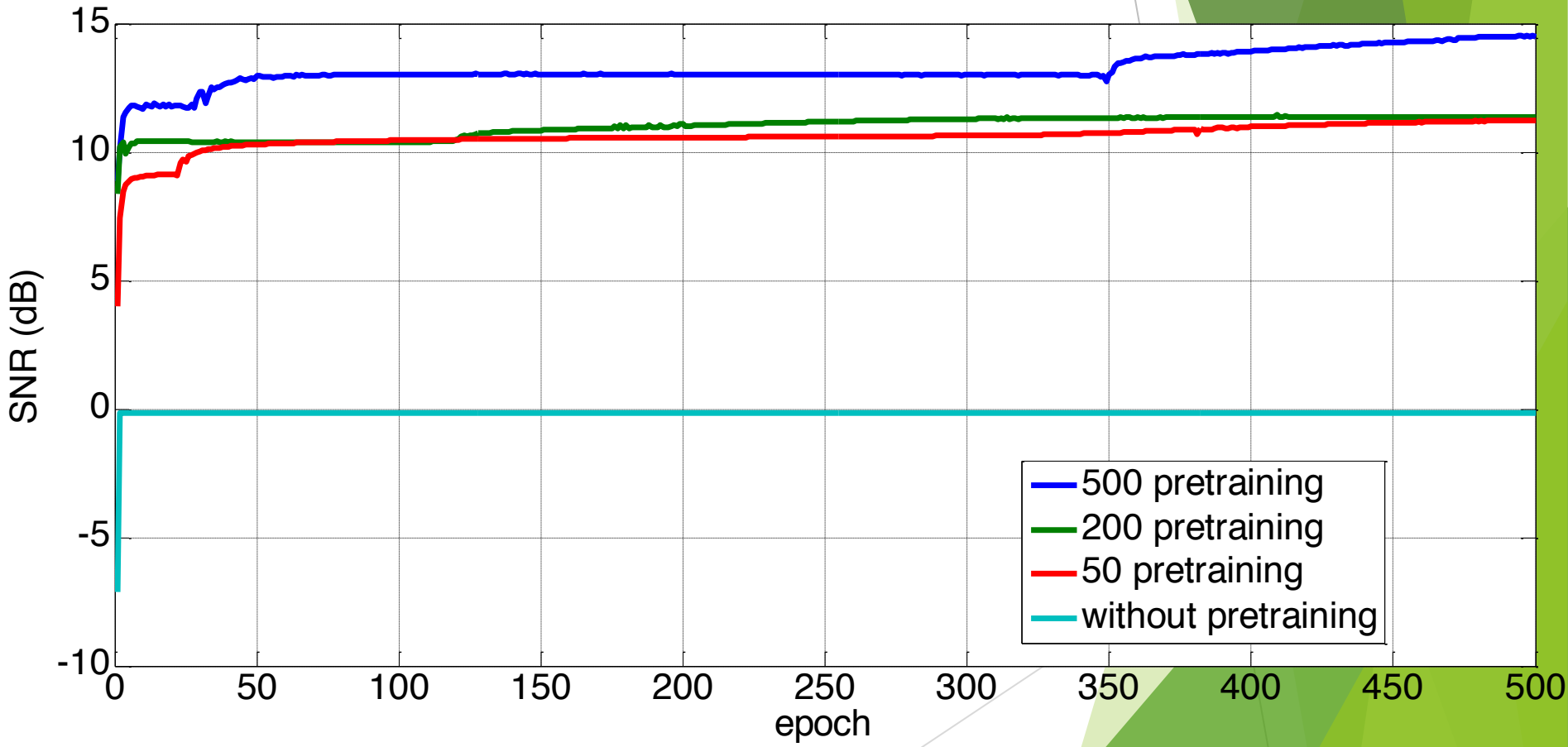
Deep Learning Compression

- Take parts of traces as input
- Uses RBM to pre-train different pairs of layers
- Train the network to achieve least reconstruction error using back propagation



Deep Learning Compression (Results):

- Traces are segmented with 100 length
- RBM pre-training boosts the fine tuning
- Compression ratio 5:1



Remarks on Deep Machine Learning

- Preliminary results indicate:
  - For compression ratio 5:1, SNR is 15
  - More comprehensive experimentations are needed

Accomplishments (Conferences)

- Near Lossless Seismic Data Compression Using Signal Projection Technique. 4th International Geoscience & Geomatics Conference, November 23-25, Bahrain, 2015
- Distributed principal component analysis for data compression of sequential seismic sensor arrays, SEG Annual Meeting, October 16-21, Dallas, 2016.
- Seismic Data Compression using Signal Alignment and PCA, 9th IEEE-GCC, Bahrain, 2017 (Accepted)

Accomplishments (Patents)

- A novel DPCA seismic compression (Submitted)
- A sequential PCA compression for seismic sensor networks (Submitted)

Accomplishments (Journals)

- Smart Seismic Sensor Network Data Compression using Distributed Principal Component Analysis. (Submitted to IEEE Trans. on geoscience and remote sensing)
- A Model-based seismic data compression, (To be submitted);
- Deep machine learning compression for seismic data, (To be submitted);
- A Literature Survey on Seismic Compression Techniques. (In preparation)

Thank You