Automated Data Collection and Transmission System for Subsurface CO_2 Monitoring

DE-SC0019854

Dr. Salah A. Faroughi

Petrolern LLC

U.S. Department of Energy

National Energy Technology Laboratory

Carbon Management and Oil and Gas Research Project Review Meeting – Carbon Storage

August 2 - 11, 2021

Presentation Outline

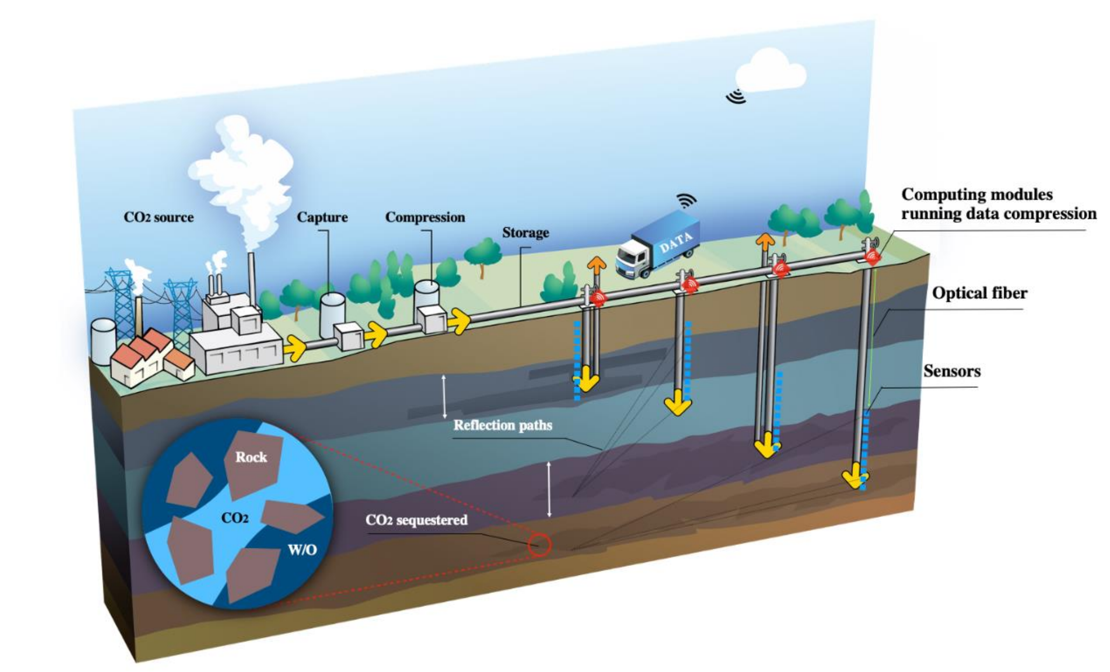
- Introduction
  - Problem Statement
  - Goals and Objectives
  - Benefit to Program
- Technical Status
  - Applied Methodologies
  - Results Set 1
  - Results Set 2
  - Next Steps
- Project Summary
- Accomplishment to Date
- Lessons Learned

Introduction

Problem Statement

- Large volumes of data generated by advanced subsurface monitoring sensors for CO_2 MVA in carbon storage sites presents a challenge in terms of transmitting to the surface, processing and storage (e.g., DAS: 4TB/day)

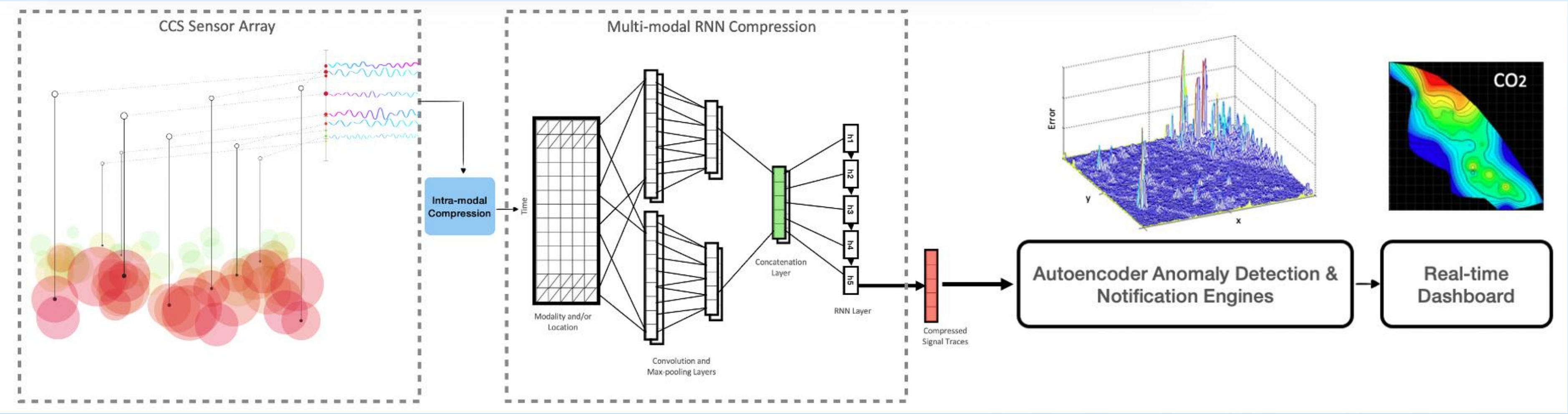
- Capability is needed to reduce the volumes of data by pre-processing the data to create smaller volumes of derivative data to enable real-time decision making.



Introduction

Goals and Objectives

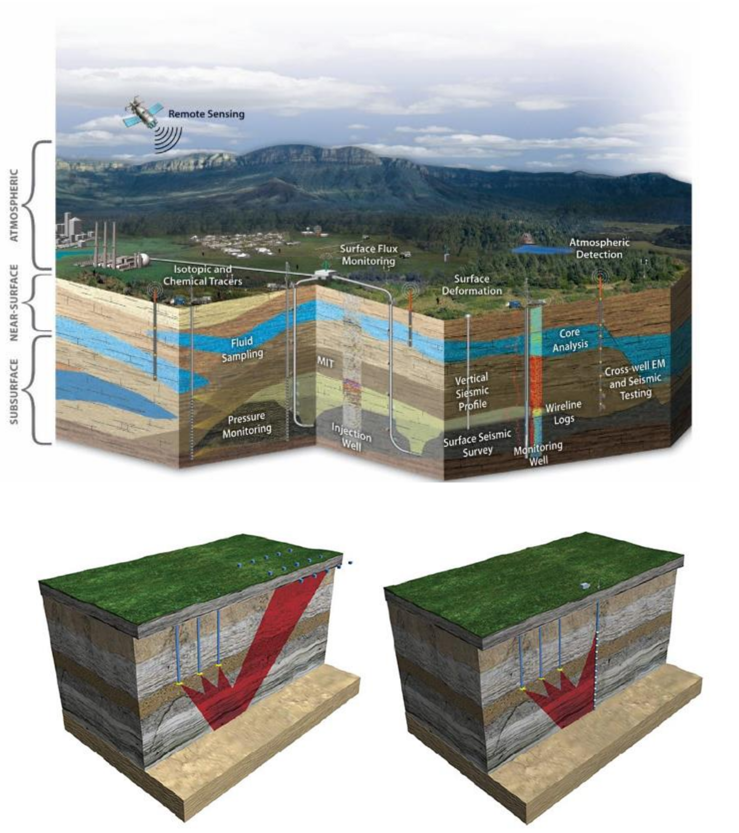
- Design and develop an automated data compression system for adaptive data acquisition and smart segmentation to address the issues of high latency, inadequate bandwidth, and limited storage
- Account for the interdependency between data using a multimodal compression method to first remove temporal redundancy, and then spatial redundancy (**compression ratio between 10-100**).



Introduction

Benefit to Program

- The benefits of an efficient data compression toolset for the CCS include:
  - Enabling large scale data collection required for CO_2 MVA.
  - Reducing the cost of data transmission, processing and storage.
  - Enabling efficient and real-time decision making.
  - Possibility to apply the developed technology to other datatypes such as seismic surveys.

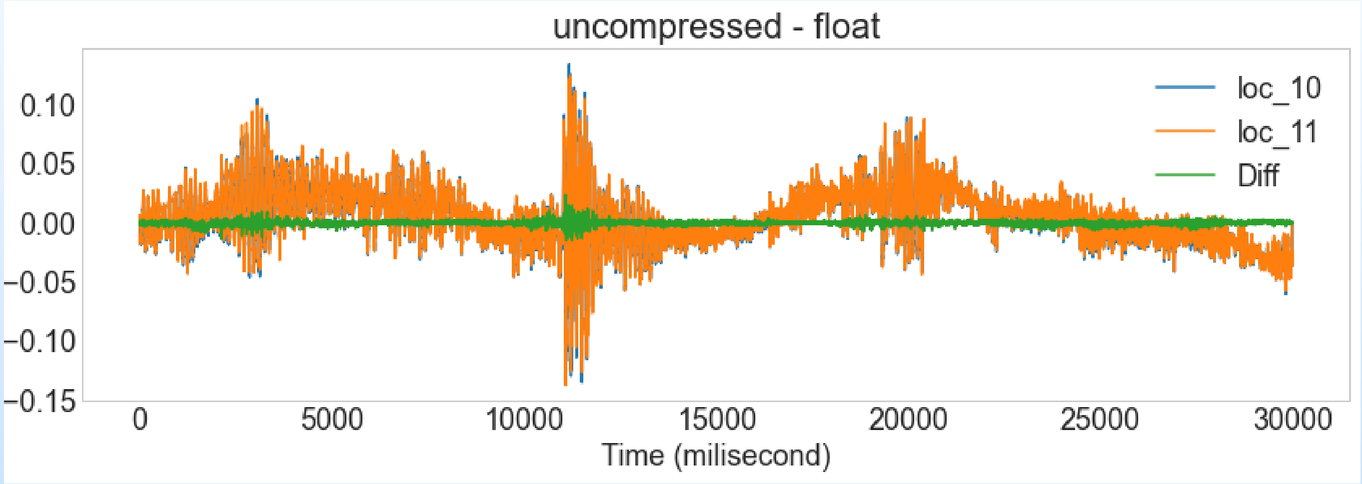


https://www.netl.doe.gov/coal/carbon-storage/advanced-storage-r-d/monitoring-verification-accounting-and-assessment

Technical Status

Applied Methodology

- Real-valued time series data is non-stationary and float (over large alphabet usually 32 bits)



- Traditional compression methods are not suitable
  - Correlations across parallel sensors are not leveraged
  - They usually do not offer a trade-off between compression gain and error rate.

Technical Status

Applied Methodology

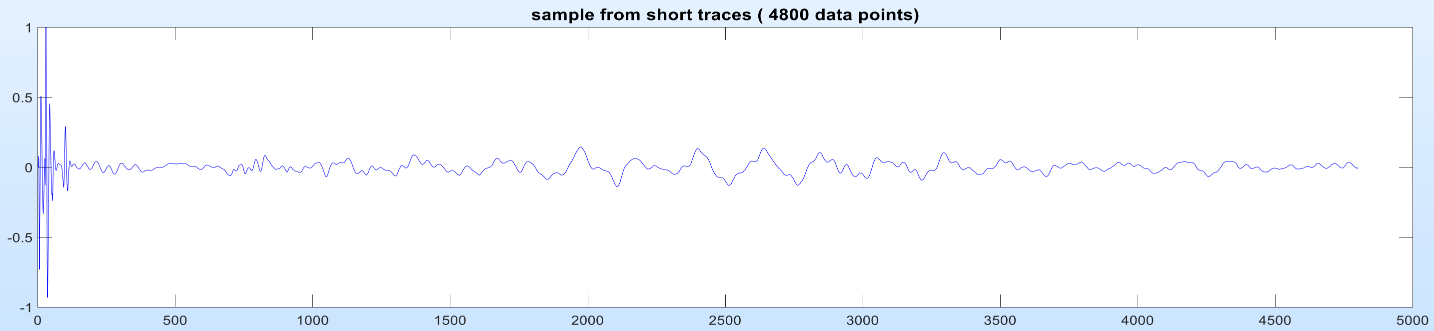
- First Level compression methods:
  - **Lossless:** reduces bits by identifying and eliminating statistical redundancy.
  - **Lossy:** reduces bits by removing less important information

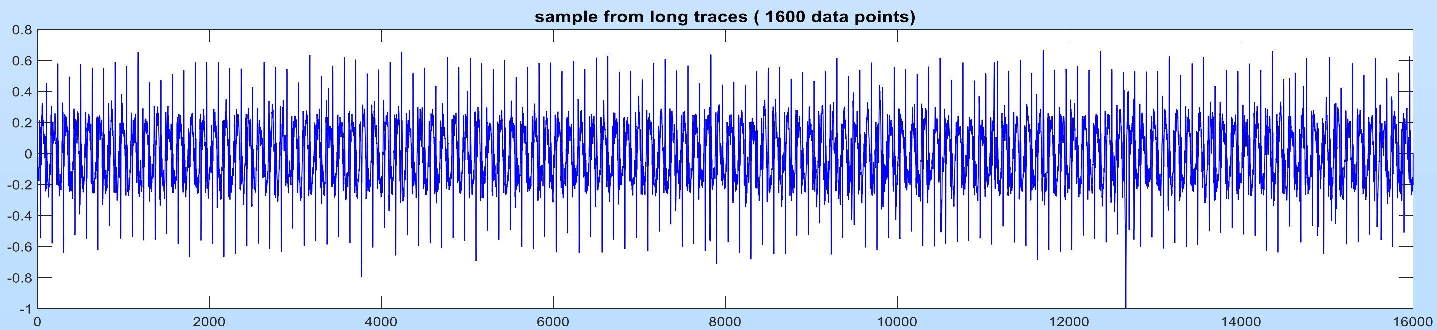
- Second level compression methods
  - **Predictive Coding:** use a predictive filter to remove temporal correlation, and then compress the residual using a lossy/lossless universal coding
    - **Pros:** Show good performance for low noise signals and are lightweight
    - **Cons:** Do not learn from the past
  - **Model-based Learning:** represent data using well-established approximation models (e.g., Neural Network)
    - **Pros:** Learn the statistics of signal and usually higher compression gain
    - **Cons:** Memory intensive and have initial learning cost

Technical Status

Results Set 1 – SEGY

- SEG-2 seismic datasets acquired in the CASSM experiment (Frio-2 GCS pilot)
- All files correspond to the single source continuous active source seismic monitoring
- Includes:
  - 6953 short traces with 4800 data points per trace
  - 1513 long traces with 16000 data points per trace

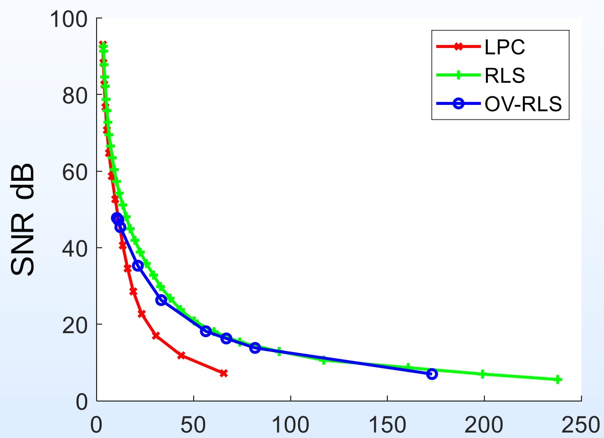




PETROLERN

Technical Status

Results Set 1 – SEGY



Short traces

Long traces

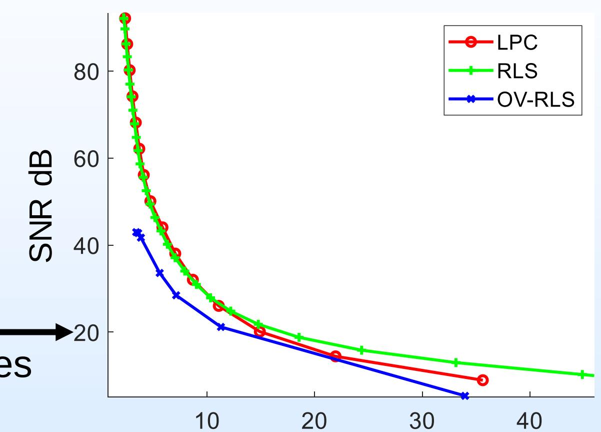


Table 1. Compression gain for short traces (ST) and long traces (LT).

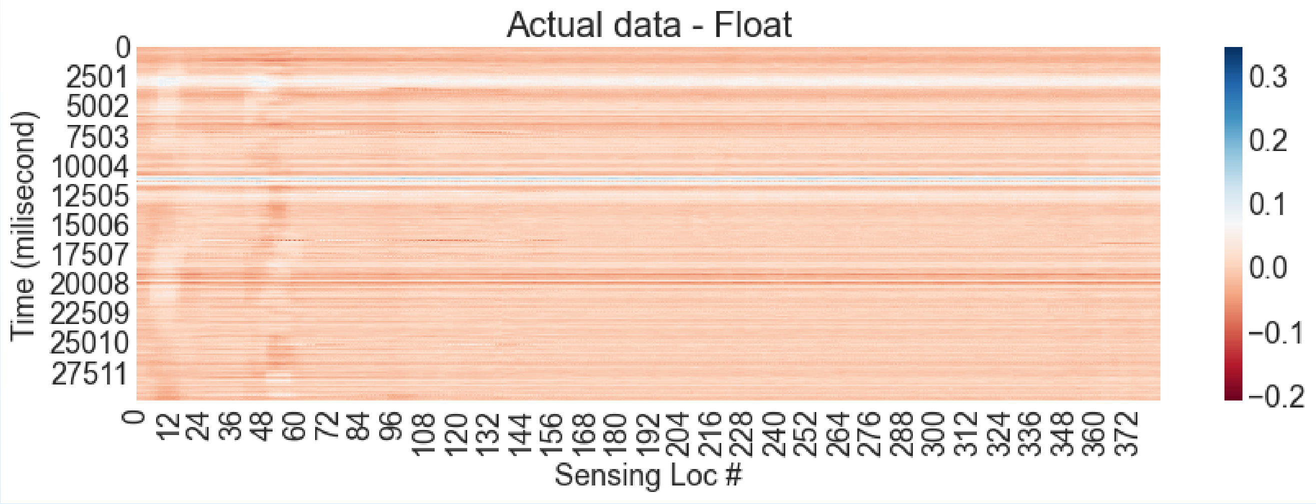
Methods	SNR 20 dB		SNR 30 dB		SNR 40 dB		SNR 50 dB	
	ST	LT	ST	LT	ST	LT	ST	LT
LPC+CTW	27	15	17	9	12	6	15	5
OV+RLS+CTW	53	13	28	7	17	4	8	2
RLS+CTW	54	17	32	9	20	6	15	5
Multi Trace RLS + CTW	55	24	33	13	22	7	18	6

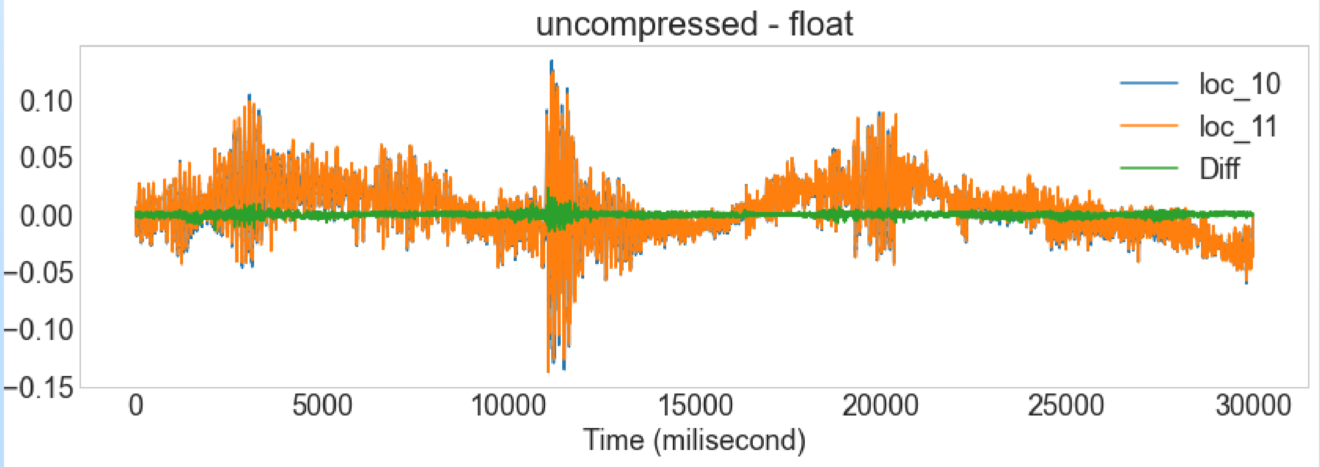
- Predictive Coding via Linear Predictive Filter + CTW
- Predictive Coding via RLS Predictive Filter + CTW
- Multi-Trace Predictive Coding via RLS predictive filter + CTW
- Oversampling + RLS + CTW **(Context Tree Weighting)**

Technical Status

Results Set 2 – DAS in H5

- Vertical monitoring at Brady Geothermal Field: deltaX = 1m, deltaT = 1ms, size = 1TB/day





Technical Status

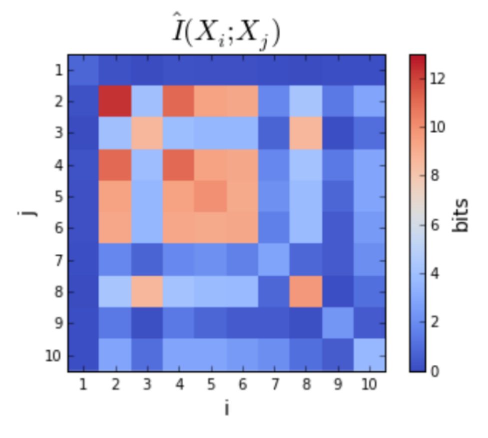
Results Set 2 – DAS in H5

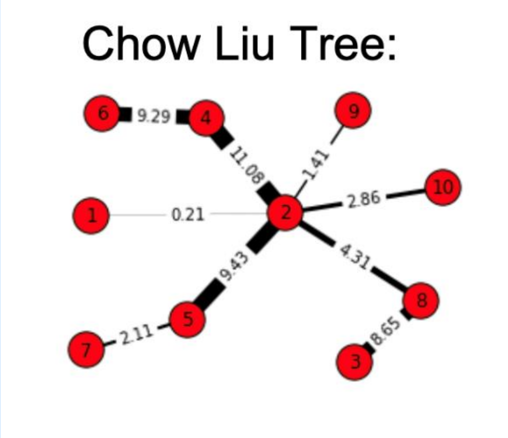
- Conventional Lossless Compression

30-sec DAS Data	88 M	CR
G-zip	33 M	2.7X
LZ2	42 M	2.1X

- These methods do not account for Cross-location dependencies!

- **Solution:** account for both temporal and spatial dependencies based on probabilistic assumptions, e.g. a Chow-Liu Tree.





- Compressing X_i given X_j:
- For each possible x_j, store $P_{X_i|X_j}(\cdot \mid x_j)$

Expected compression ratio:

15X to 20X (lossless)

Ingber, A., Pavlichin, D., Weissman, T., "Compressing Tabular Data via Pairwise Dependencies"

PETROLERN

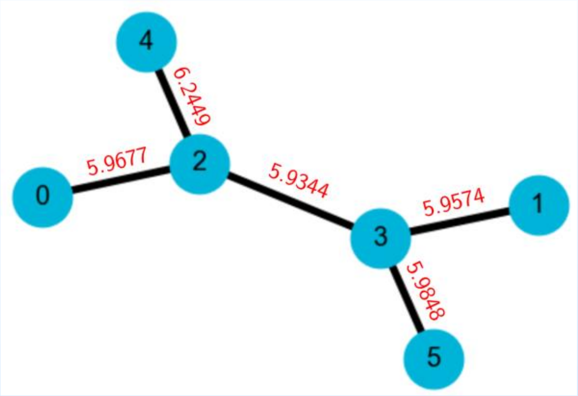
Technical Status

Results Set 2 – DAS in H5

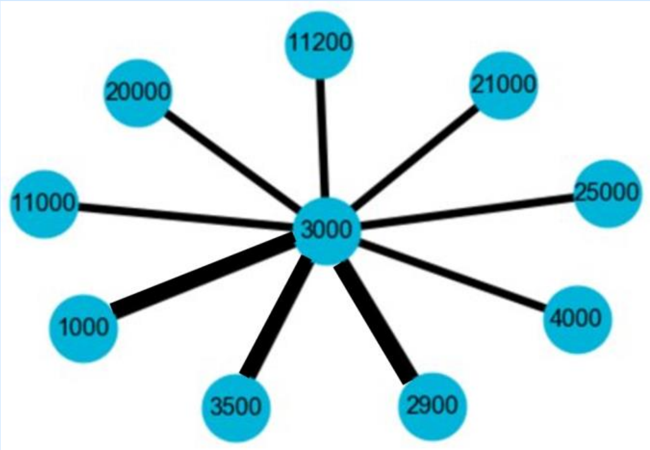
- To portray dependencies between 6 nodes using Chow-Liu Tree

$$P(X_0, X_1, X_2, X_3, X_4, X_5) = P(X_0)P(X_1)P(X_2|X_0,X_1) P(X_3|X_1)P(X_4|X_2)P(X_5|X_2)$$

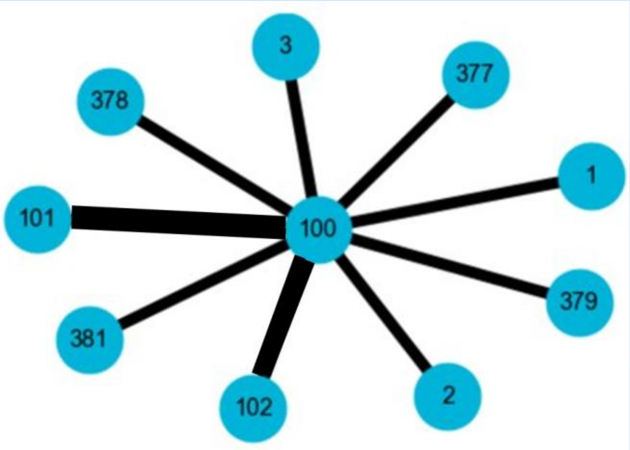
Total bits saved through joint distribution ≈ 30 bits



Temporal dependencies of DAS



Spatial dependencies of DAS



PETROLERN

Project Summary

- Key Findings
  - Predictive coding can remove temporal redundancies with compression rate between 20X and 55X (Lossy for SEGY)
  - Lossless compression models on DAS data can achieve compression rate between 2X and 3X
  - Distribution trees-based compression that account for temporal and spatial dependencies together can achieve higher compression rates

- Next Steps
  - Compress lossless distribution trees with goal of achieving 15X - 20X
  - Combine predictive coding and distribution trees
  - Finalize the predictive coding compression algorithm for low powered devices (Raspberry PI)
  - Examine several model-based compression methodologies for cloud compression (RNN-based models)

Accomplishments to Date

- Many contacts has been made to gather the CO_2 sensory data from variety of MVA projects.
  - DTS
  - DAS
  - Seismic Surveys
- Several compression algorithm candidates have been tested.
  - Lossless & Lossy
  - Predictive Coding (e.g., Recursive Least Square + CTW)
  - Model-based Learning (e.g., NN, RNN, Sparse Dictionary Learning)
  - Universal and multipurpose methods (e.g., gzip, Lz2)
- Baseline python codes for most of the compression algorithms methods have been implemented.
- The evaluation of the candidate methods based on the two measure of compression gain and temporal/spatial complexity is being carried out using DAS and low and high noise seismic data.

Lessons Learned

- Without impacting the downstream, time-series monitoring data can be largely compressed as they contain a high-level of noise and a high-level temporal and spatial interrelationship.

- The data type for most of the time-series monitoring data is Float64 that is extremely harder for compression.

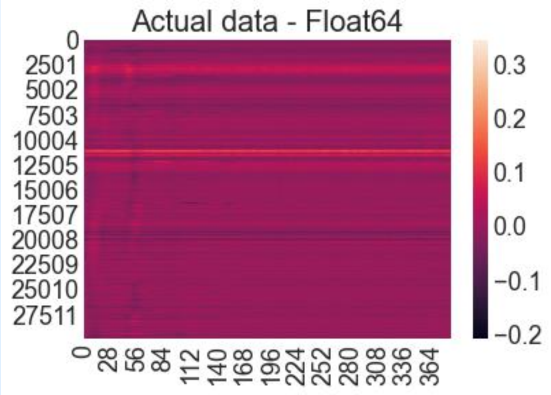
- We also plan to build a multimodal compression technique, no data can be found for which two modal of data are acquired simultaneously.

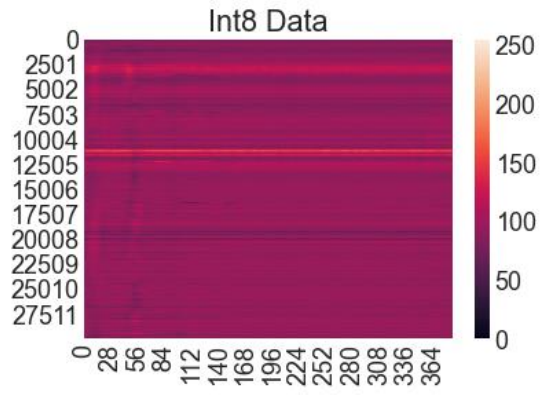
- To solve the multimodal data issue, we are now fully focused on fiber optics for which multimodal data can be acquired with same temporal and special discretization.

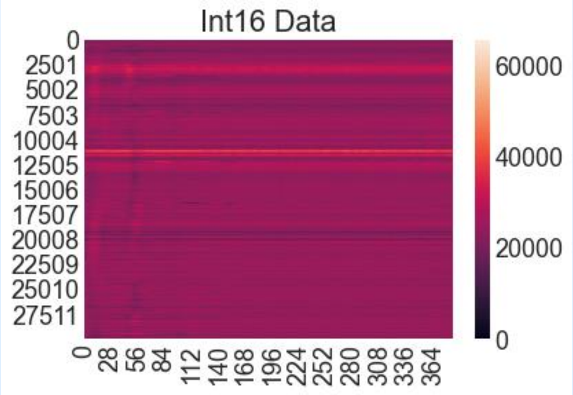
Appendix

– Float to Int mapping required for Chow Liu Tree

DAS Data (Float64)	88 M	Error
Int32	43 M	7.4e-11
Int16	22 M	4.8e-06
Int8	11 M	0.00125

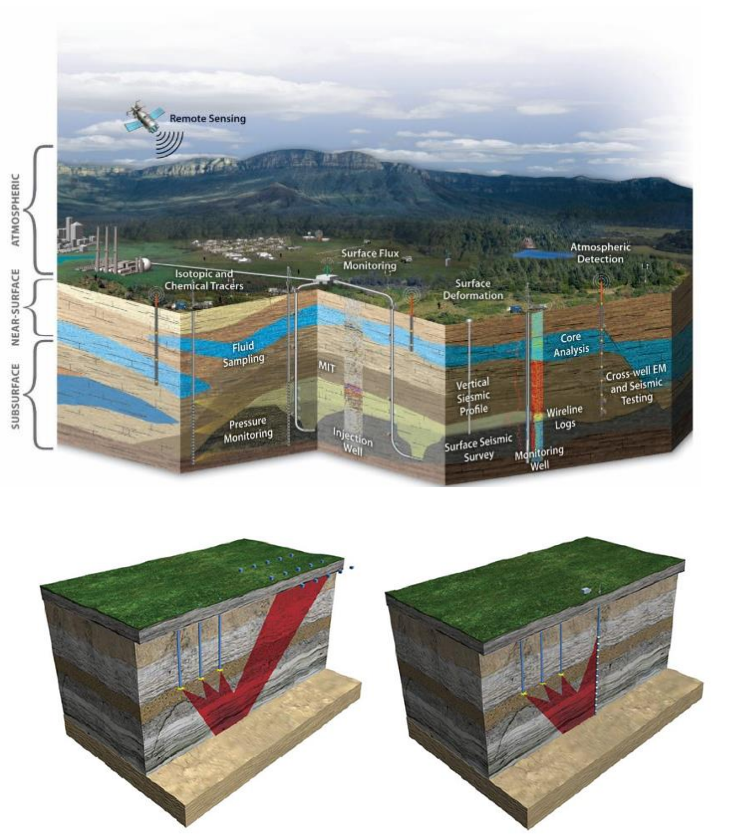






Benefit to the Program

- The benefits of an efficient data compression toolset for the CCS include:
  - Enabling large scale data collection required for CO_2 MVA.
  - Reducing the cost of data transmission, processing and storage.
  - Enabling efficient and real-time decision making.
  - Possibility to apply the developed technology to other datatypes such as seismic surveys.



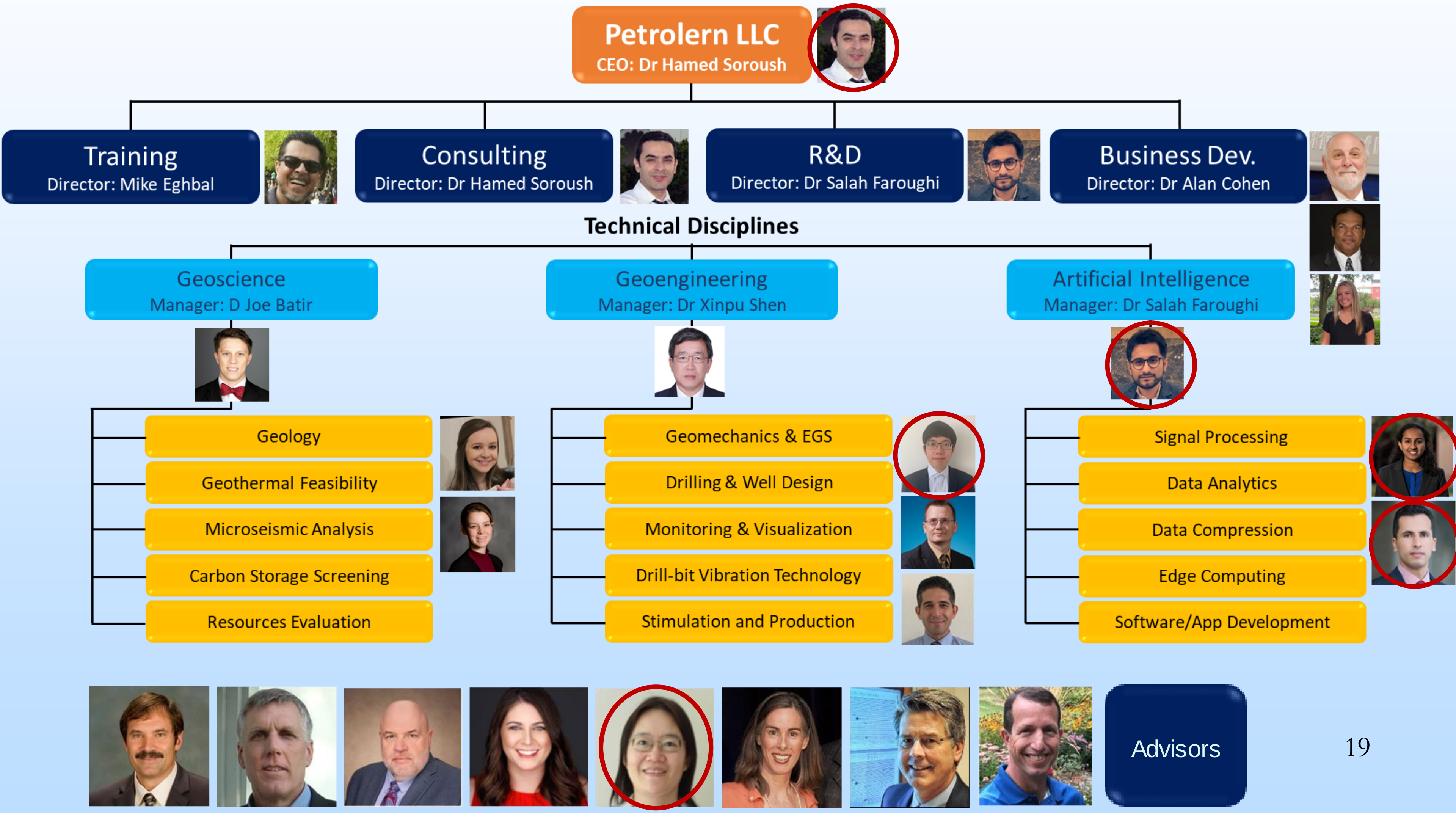
https://www.netl.doe.gov/coal/carbon-storage/advanced-storage-r-d/monitoring-verification-accounting-and-assessment

PETROLERN

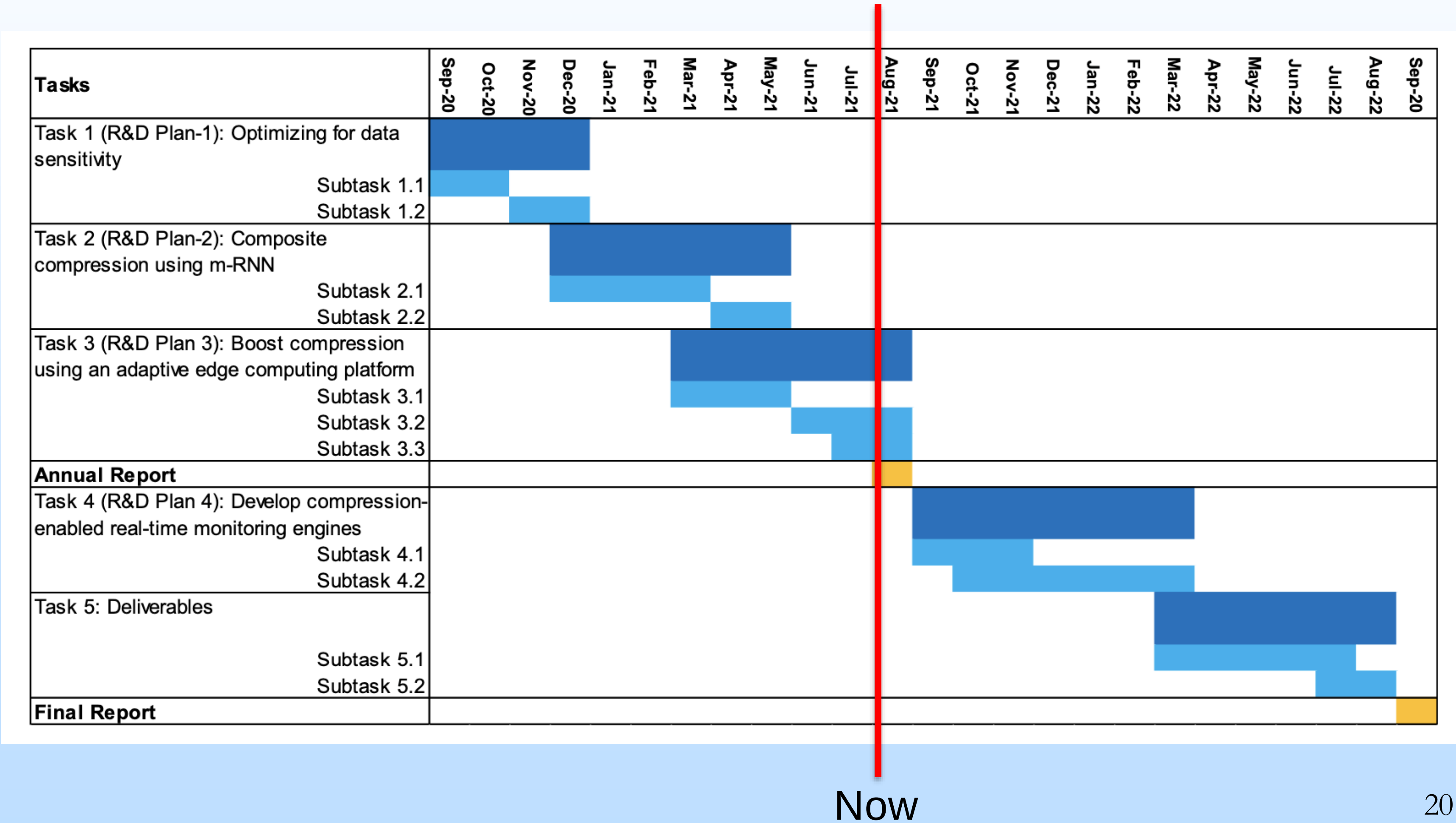
Project Overview

- Funding (DOE: $1,149,459 and Cost Share: $0)
- Overall Project Performance Dates: 08/24/2020 to 08/23/2022
- Project Participants:
  - Dr Salah Faroughi, PI
  - Dr Hamed Soroush, PO
  - Dr Ling Liu, Advisor
  - Dr Kamal Shadi, Consultant
  - Manju Murugesu, Intern
  - Dr Hao Kang, Scientist
- Overall Project Objectives: Account for the interdependency between data using a multimodal compression method to first remove temporal redundancy, and then spatial redundancy (compression ratio between 10-100).

Organization Chart



Gantt Chart


Tasks
Sep-20
Oct-20
Nov-20
Dec-20
Jan-21
Feb-21
Mar-21
Apr-21
May-21
Jun-21
Jul-21
Aug-21
Sep-21
Oct-21
Nov-21
Dec-21
Jan-22
Feb-22
Mar-22
Apr-22
May-22
Jun-22
Jul-22
Aug-22
Sep-20
Task 1 (R&D Plan-1): Optimizing for data sensitivity
Subtask 1.1
Subtask 1.2
Task 2 (R&D Plan-2): Composite compression using m-RNN
Subtask 2.1
Subtask 2.2
Task 3 (R&D Plan 3): Boost compression using an adaptive edge computing platform
Subtask 3.1
Subtask 3.2
Subtask 3.3
Annual Report
Task 4 (R&D Plan 4): Develop compression-enabled real-time monitoring engines
Subtask 4.1
Subtask 4.2
Task 5: Deliverables
Subtask 5.1
Subtask 5.2
Final Report
Now

Bibliography

– This is a SBIR project, and no paper is allowed for an extended amount of time.