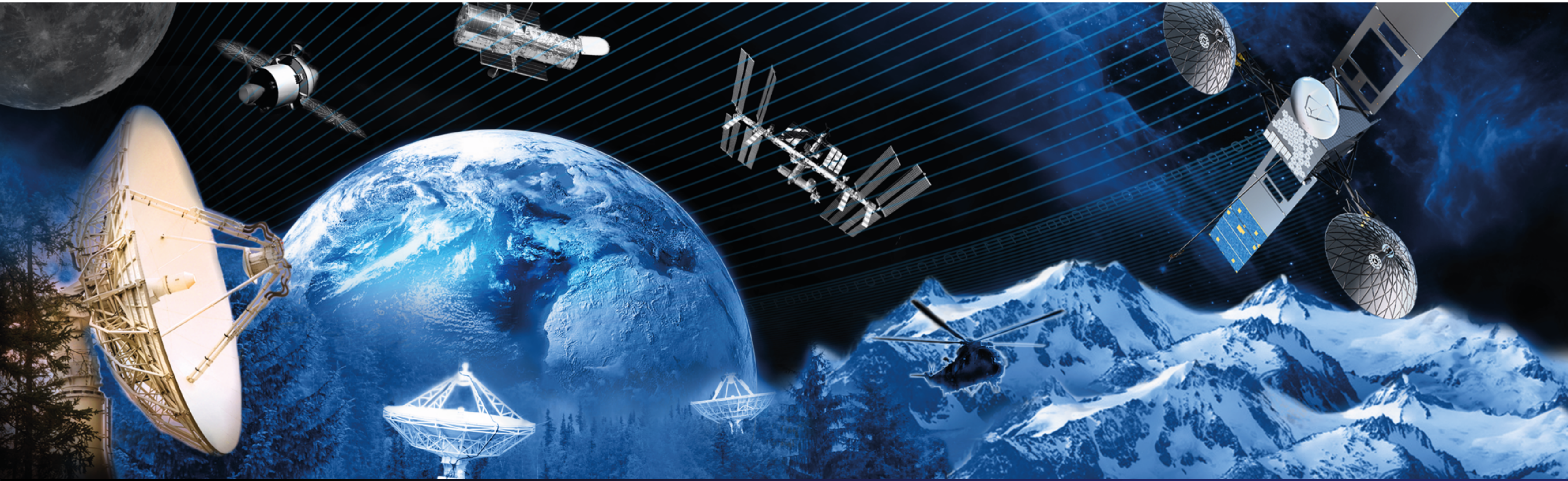
National Aeronautics and Space Administration

Exploration & SPACE Communications

More than you ever imagined...

Machine Learning Algorithms for Error Correction in Space Optical Communications Systems

By Naveed Naimipour, Haleh Safavi, Harry Shaw and Mojtaba Soltanalian



Outline

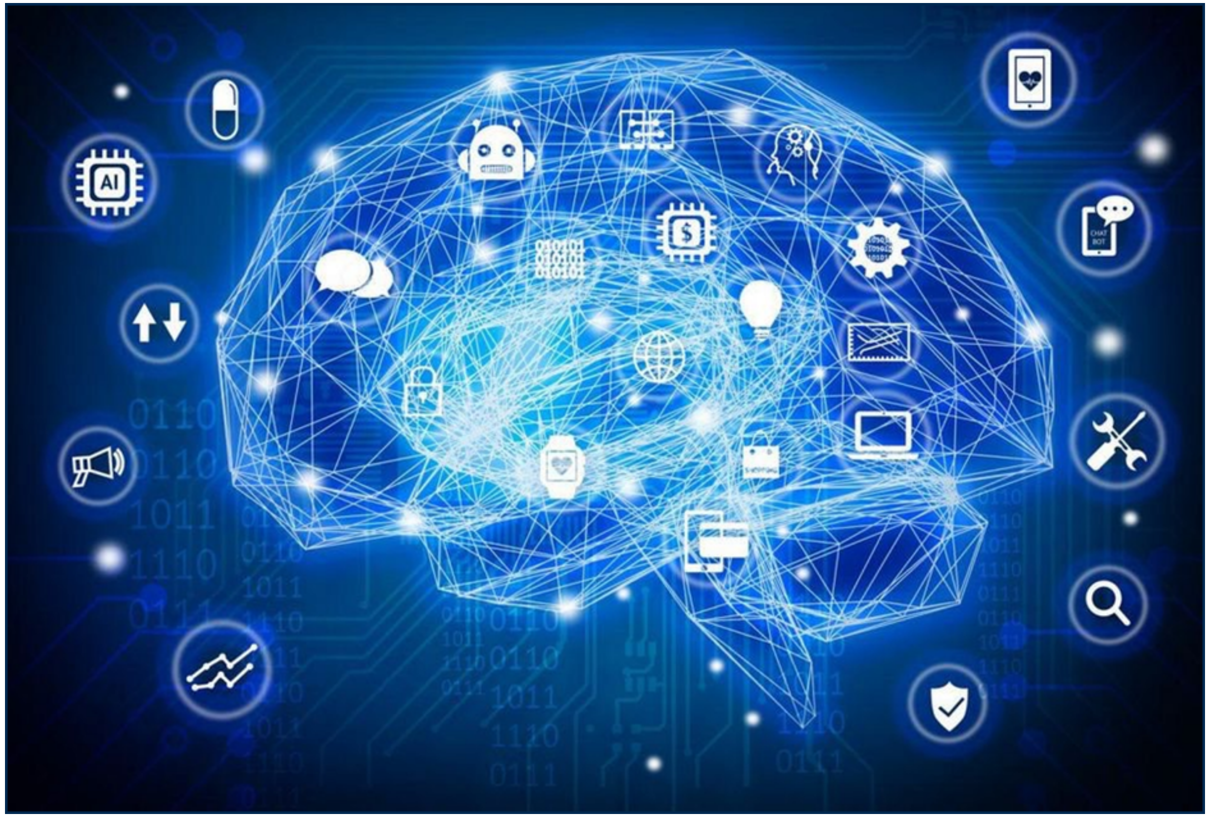
1. Introduction and Background
2. Clustering Programs
    - 2-Hard and 2-Soft Clustering
3. Proposed Systems
4. Results
5. Discussion
6. Future Work

Overview

Purpose: Exploring the applications of machine learning for FEC codes in optical communications.

Examining the concept of “smarter” error correction systems.

1. Proposing a system with potential real-time self-correction
2. Simulating schemes for system enhancement and real-time self-correction



Background

1. Advancement of ML algorithms for large scale real-time data has imminent applications in facilitating error correction systems
2. Determining the best performing FEC technique for various applications has continuously been a challenge
   - Especially relevant to NASA related missions through transitions of Reed-Muller codes, convolutional codes, concatenated convolutional and Reed-Soloman codes, turbo codes, and LDPC codes [1]
3. It was long believed that it would be too difficult to implement machine learning techniques onto FEC subsystems
   - Particularly true for optical communications systems
4. However, ML algorithms have slowly made their way into FEC systems and optical communications
   - Supervised learning techniques can assist with increasingly complex fiber-optic networks [2]
   - Implementation of the Metropolis-Hastings algorithm with Bayesian filtering for applications like system identification and carrier recovery [3]
5. Still, primary focus of ML has been on the optimization of overly complex systems
   - Most work being done on FEC is primarily language based [4]
6. To the best of our knowledge, there is no work being done on the real-time application of ML algorithms for error correction

Background

What if we took an alternate approach?

1. Utilize the 2-Hard and 2-Soft Clustering techniques introduced in [5] to…
   - Assist with the processing of large swaths of FEC data
   - Contribute to the error correction of large data sets
     - FEC codes and conventional clustering typically struggle to accomplish this efficiently

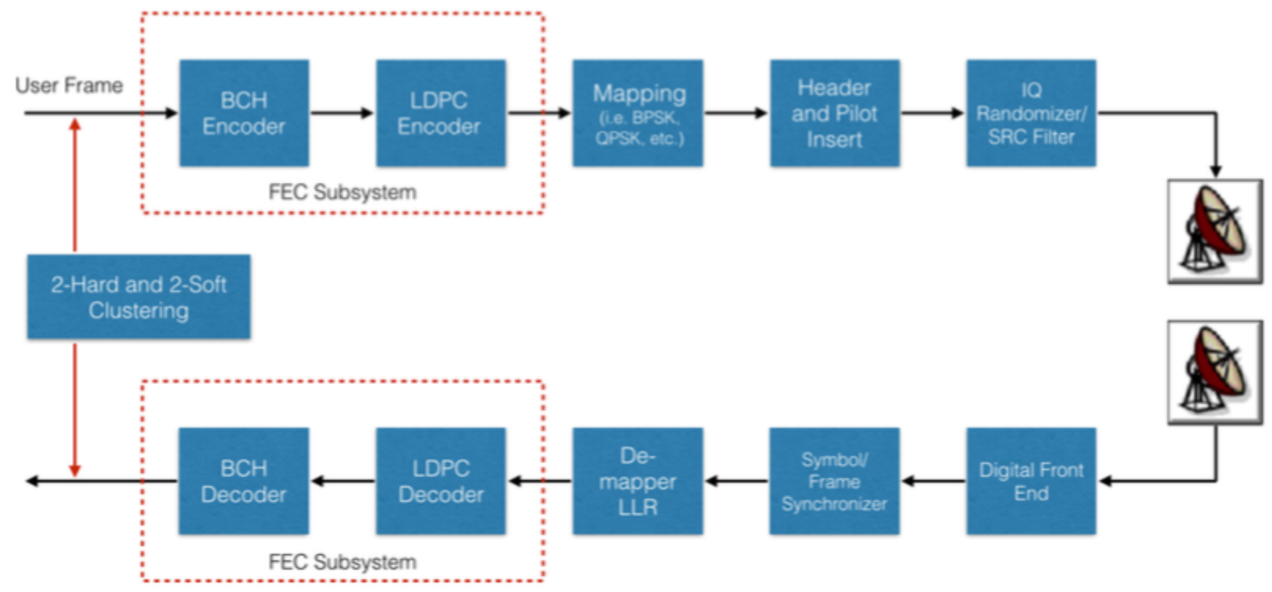


Fig. 1: Example of proposed system

2-Hard and 2-Soft Clustering

Begin with the two definitions from [5]:

Definition 1 *Let* $B = \{0, 1\}$. *We call* $\mathbf{X} \in B^{n\times k}$ *a* k*-clustering matrix if and only if each row of* $\mathbf{X}$ *has exactly one 1. The subset of* k*-clustering matrices of* $B^{n\times k}$ *will be denoted by* $\mathcal{H}_{n,k}$.

Definition 2 *We call* $\mathbf{X} \in \mathbb{R}^{n\times k}$ *a soft* k*-clustering matrix if and only if the elements of* $\mathbf{X}$ *are nonnegative and the sum of the entries at each row is equal to one. The associated subset of* $\mathbb{R}^{n\times k}$ *will be denoted by* $\Omega_{n,k}$.

2-Hard and 2-Soft Clustering

From Definition 1 and Definition 2, the following non-convex clustering programs are formed:

(a) 2-Hard Clustering:

$$\min_{\mathbf{X}\in\mathcal{H}_{n,k},\mathbf{Q}\in\mathbb{C}^{n\times k}} \|\mathbf{XQ}-\mathbf{A}\|_F \text{ s.t. } \mathbf{QQ}^T=\mathbf{I}_k, \tag{1}$$

(b) 2-Soft Clustering:

$$\min_{\mathbf{X}\in\Omega_{n,k},\mathbf{Q}\in\mathbb{C}^{n\times k}} \|\mathbf{XQ}-\mathbf{A}\|_F \text{ s.t. } \mathbf{QQ}^T=\mathbf{I}_k. \tag{2}$$

where **A** denotes the adjacency matrix of the graph

Proposed System

1. Utilize the algorithms (1)-(2) before encoding and after decoding
   - Simple cross referencing of clusters to determines the effective error regions as shown in the example system in Fig. 1
   - Noting every change (and the sensitivity of optical communications) leads to additional data that may be too difficult to process
2. Cross referencing is the only important operation linking the error correction,
   - Thus, it does not matter how the data is organized as long as it is consistent
   - Data consistency, on both ends, allows the algorithms to successfully identify the location of the error regions
3. Note that conventional clustering techniques lack precision when processing missing or incorrect data
   - 2-Hard and 2-Soft Clustering has shown promise to overcome this obstacle
4. No longer needed to repeatedly check the entire data set for errors with known error regions
   - Error correction becomes immensely more efficient
5. Structure is easily scalable due to point of implementation within the framework of FEC subsystem

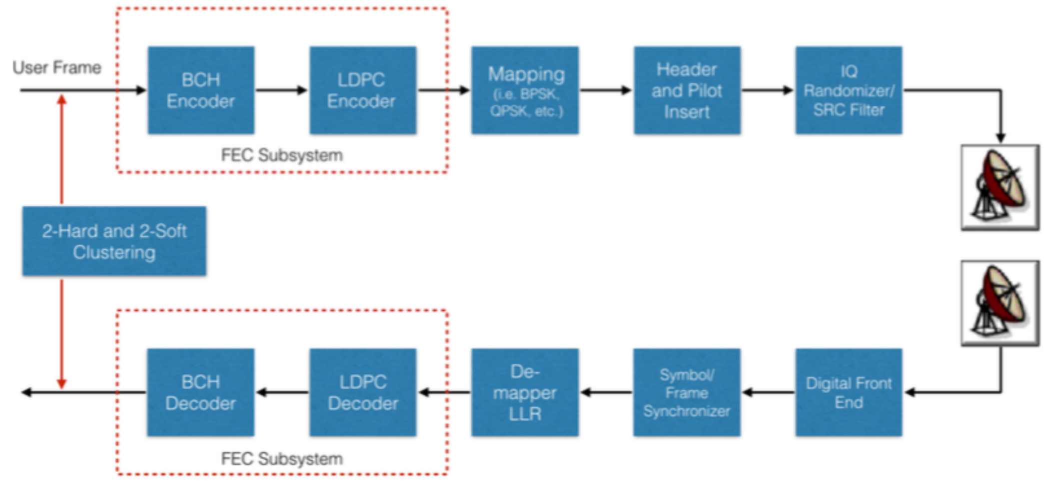


Fig. 1: Example of proposed system

1. Security Components:
   - Data Organization
     - Operations on the data will have no effect on ability of the algorithm to perform well as long as it is consistent for the encoder and decoder
   - Cluster Representation
     - The clusters determined by the algorithm have no underlying meaning other than being correct or incorrect
     - No sensitive operation is being performed or outputted that needs to be secured further

Results

MATLAB Simulation Details

1. Simulated under CCSDS suggested parameters for an optical communications set up
2. Run for a QPSK constellation with 1/2 rate LDPC
3. Fig. 2 shows the data being run through both the 2-Soft and 2-Hard Clustering algorithms before encoding
4. Fig. 3 shows the data being run through both the 2-Soft and 2-Hard Clustering algorithms after decoding
5. A refers to the adjacency matrix in each plot
6. True clusters are shown in the top right of each figure
   - Comparison of the true clusters of each figure verifies that the data being clustered is correct and gives an indication of what the algorithms should roughly output
7. 2-Soft and 2-Hard Clustering results on the bottom of each figure

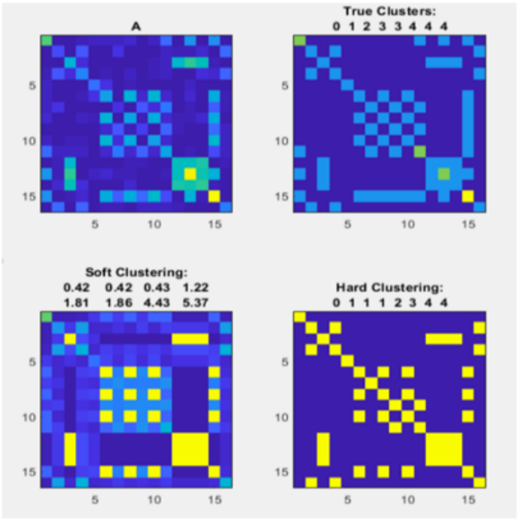


Fig. 2: Graph clustering using 2-Soft and 2-Hard Clustering formulations ($n = 8$) before encoding

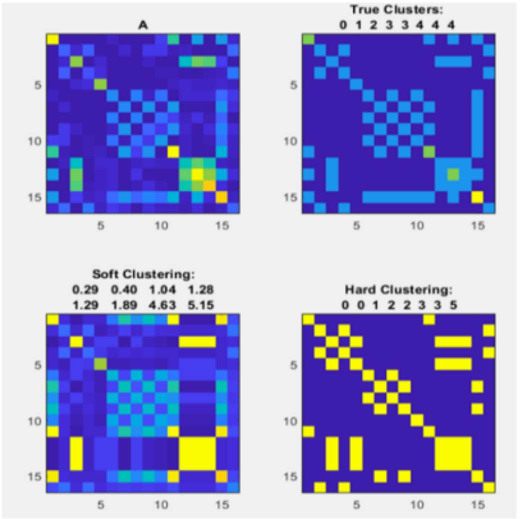


Fig. 3: Graph clustering using 2-Soft and 2-Hard Clustering formulations ($n = 8$) after decoding

Results

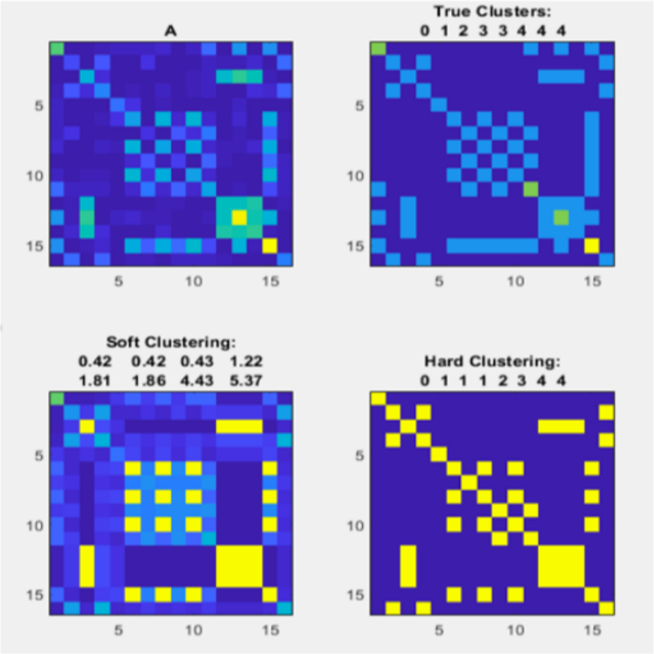


Fig. 2: Graph clustering using 2-Soft and 2-Hard Clustering formulations ($n = 8$) before encoding

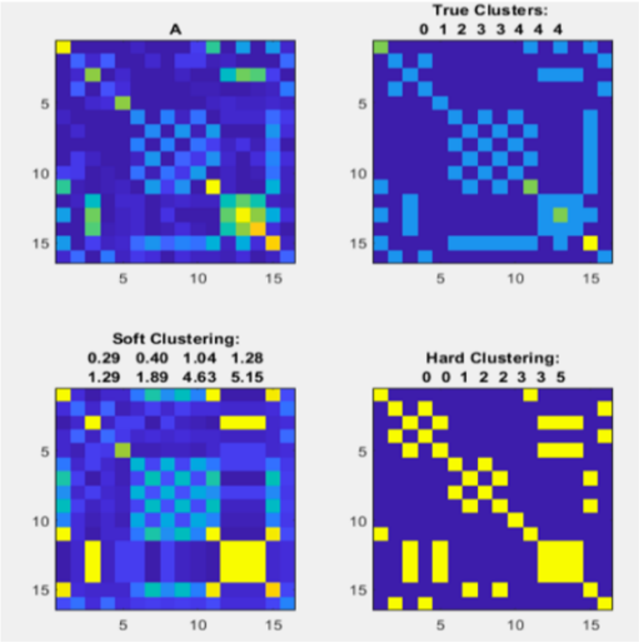


Fig. 3: Graph clustering using 2-Soft and 2-Hard Clustering formulations ($n = 8$) after decoding

1. This is one iteration of many that were simulated
2. Note that the number of clusters can be optimized for better precision
   - *k = 8* was used for aesthetic purposes

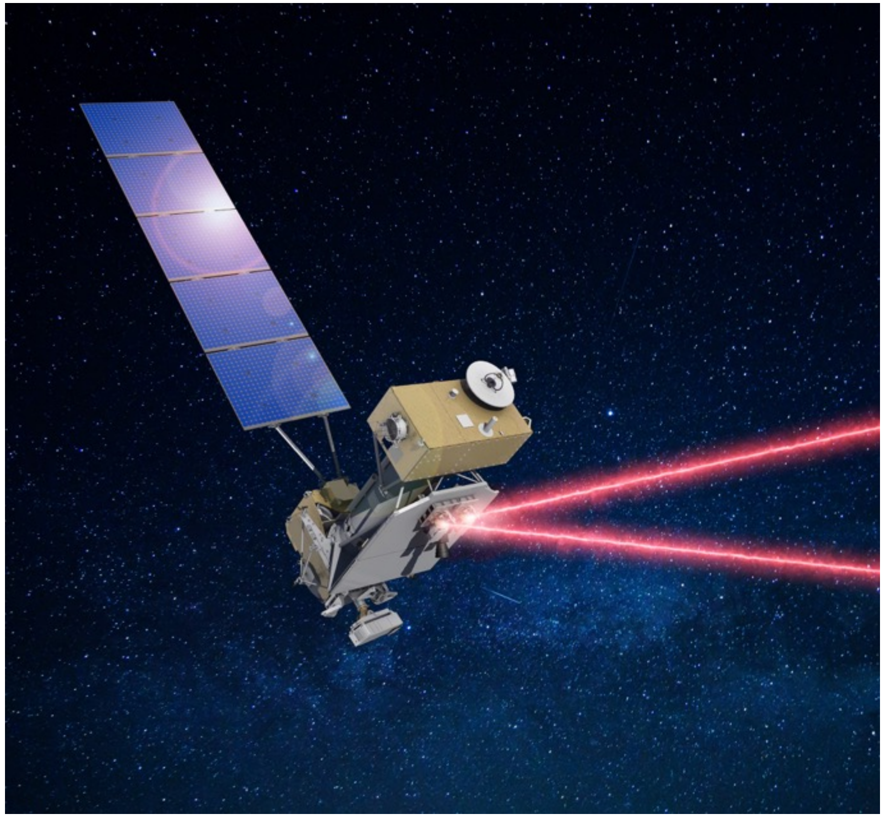
1. Larger data sets were primarily tested
   - Sizes close 1GB due to MATLAB's data processing limit
   - We expect similar results with even larger data sets

Discussion

1. True clusters show whether the algorithms can correctly cluster the error laden data
2. For Fig. 2 and Fig. 3, the true clusters are identical, so
   - Data is being outputted correctly
   - Wrong or disorganized data is not being clustered
3. 2-Soft Clustering portion depends on more detail than can be provided with just $k = 8$ clusters
   - Increasing the number of clusters would resolve this
   - Still, 2-Soft Clustering is accurate for five out of the eight clusters.
4. 2-Hard Clustering indicates directly that four out of the eight clusters were perfect, while the others had errors
   - The hard clusters clearly shows which parts of the data have errors based on the mismatched clusters
5. System knows which areas of the data are incorrect and which areas are correct
   - It can then run its error correction scheme only on the areas of error
   - No need to re-run the clusters that were correct

Future Work

1. System is more efficient and can be self-automated in the future
    - No longer needs oversight to determine how and when to correct errors
    - Can continuously use the 2-Soft and 2-Hard Clustering algorithms until minimal error clusters are detected during cross referencing
2. Note that although the general framework in this paper will work for any subsystem
    - However, efficient real-time performance will heavily depend on the specific FEC codes and parameters being implemented
3. Future work will involve further testing the proposed framework parameters
    - This includes detailed parameters of the 2-Hard and 2-Soft Clustering algorithms and FEC subsystems
    - Assists in best determining errors in real time for space communications



THANK YOU!

References

[1] K. S. Andrews, D. Divsalar, S. Dolinar, J. Hamkins, C. R. Jones and F. Pollara, "The Development of Turbo and LDPC Codes for Deep-Space Applications," in *Proceedings of the IEEE*, vol. 95, no. 11, pp. 2142-2156, Nov. 2007.

[2] F. Khan, Q. Fan, C. Lu, and A. Lau, "An Optical Communication's Perspective on Machine Learning and Its Applications," *J. Lightwave Technol*. 37, 493-516 (2019).

[3] Zibar, D, Piels, M, Jones, RT & Schaeffer, CG 2016, "Machine learning techniques in optical communication" *Journal of Lightwave Technology*, vol. 34, no. 6, pp. 1442-1452

[4] A. A. Jiang, "Machine Learning and Algorithmic Techniques for Error Correction," *2018 Information Theory and Applications Workshop (ITA)*, San Diego, CA, 2018, pp. 1-9.

[5] N. Naimipour and M. Soltanalian, "Efficient Non-Convex Graph Clustering for Big Data," *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP*), Calgary, AB, 2018, pp. 2896-2900.