Inter-channel Lossless Data Compression for Distributed Acoustic Sensing

Aleix Seguí[1], Arantza Ugalde[2], Sergi Ventosa[2]
[1]ETH Zürich, [2]Institut de Ciències del Mar (Barcelona)

EGU: Fibre Optic Sensing in Geosciences, Catania, 16-20 June 2024.

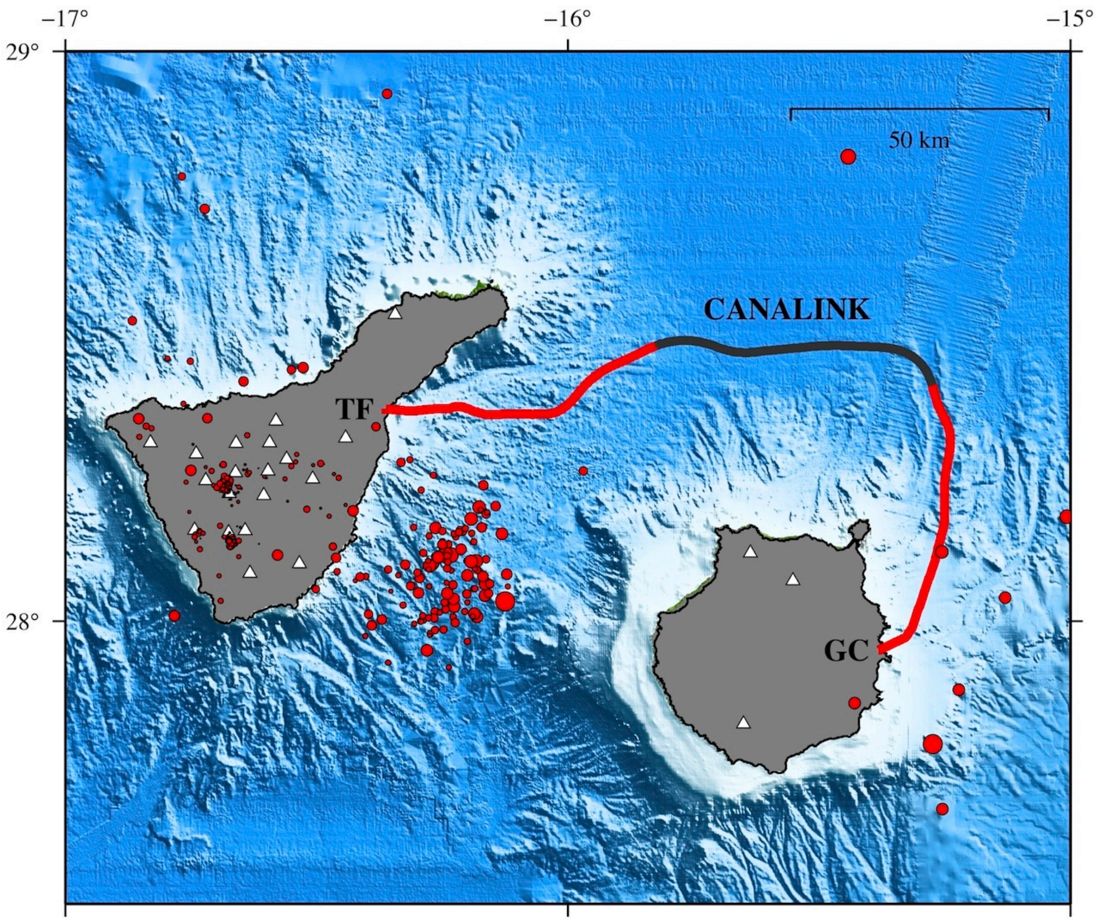
Agenda

1. Motivation and DAS data description
2. Intra and inter-channel prediction
3. Compression tools and pipeline
4. Results
5. Conclusions and Future steps

Motivation: Long-term Monitoring using Distributed Acoustic Sensing

- Distributed Acoustic Sensing (DAS) monitoring:
  - Long distance.
  - High-density.
  - Real-time.

- Long term archiving initiatives face serious challenges:
  - Generation of huge amount of data.
  - Need vast storage capacities.
  - Long data transfer times.

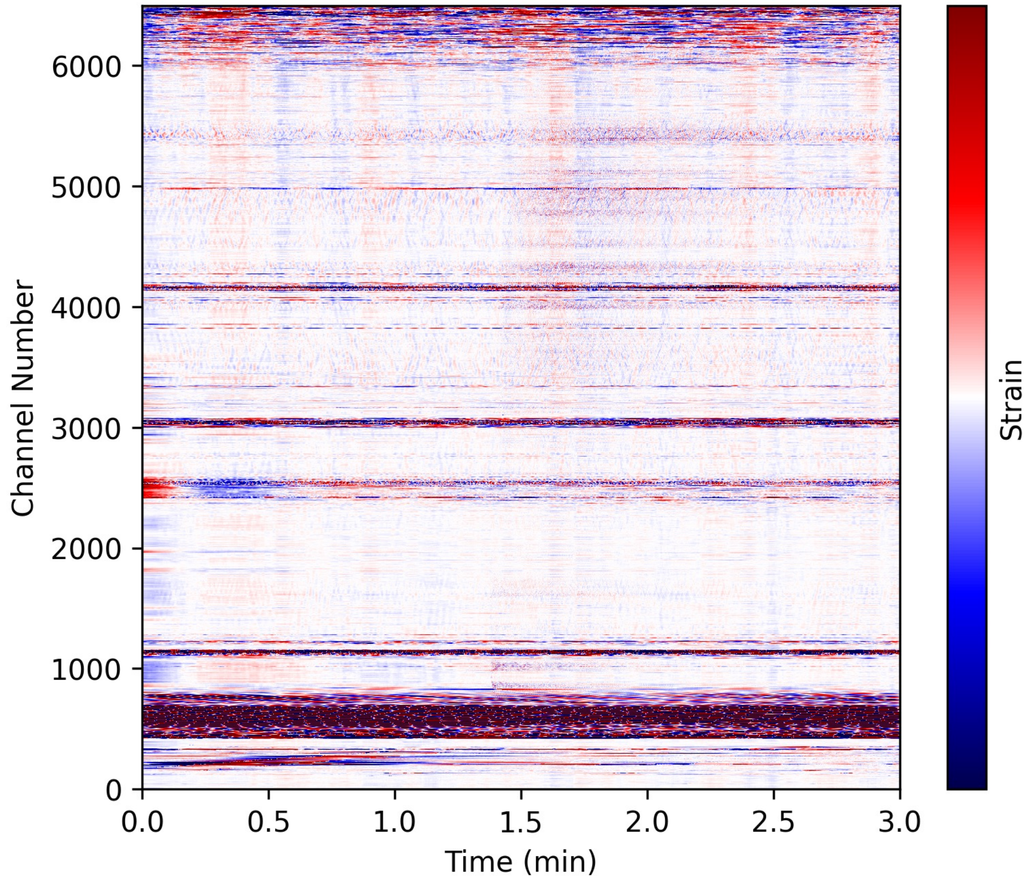


Fiber optic cable CANALINK and seismicity in Tenerife and Gran Canaria (2020).

The Challenge of DAS Data

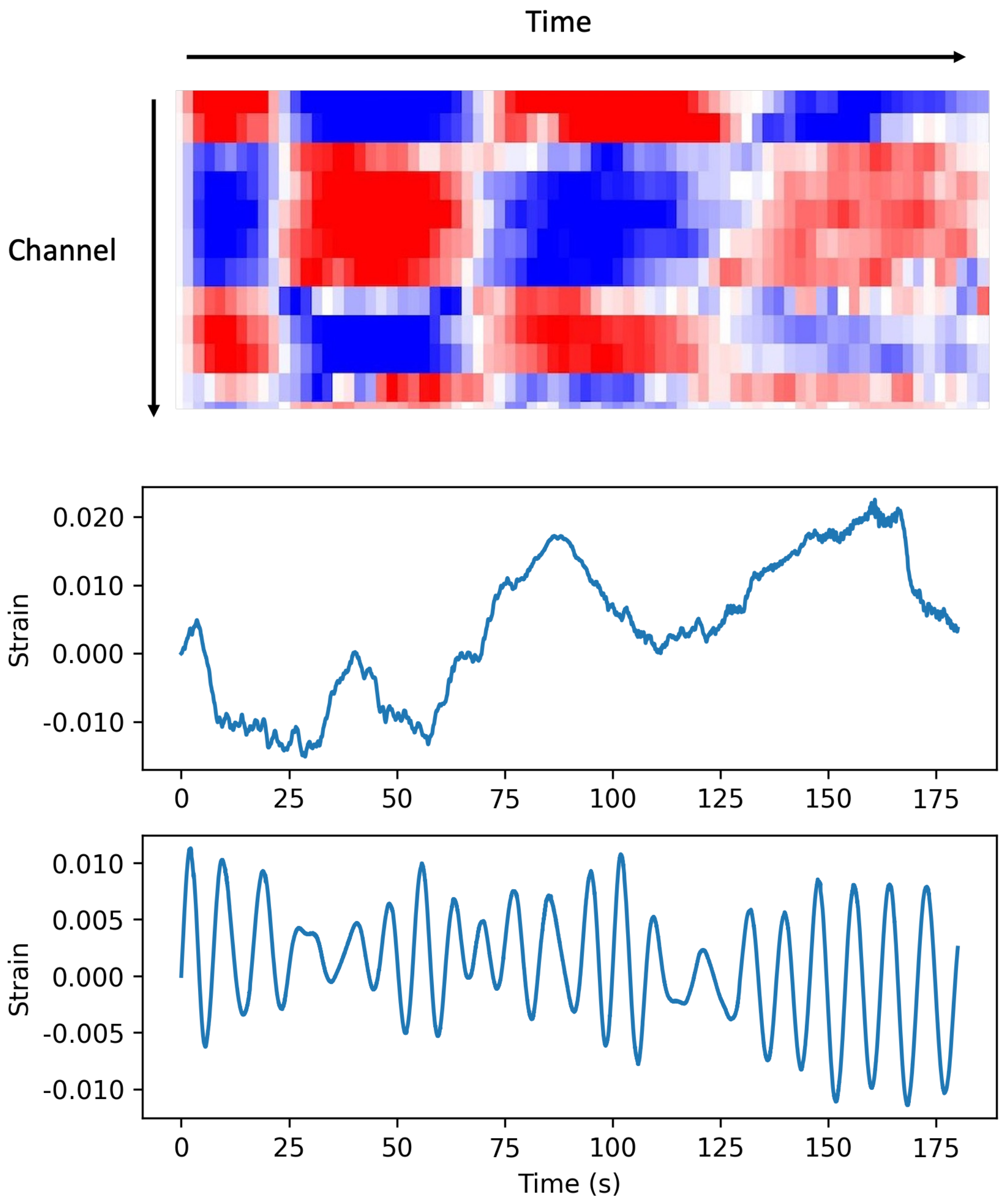
- Large sample rates on thousands of channels simultaneously.
- Hundreds of **terabytes** per year.
- Lots of noise.

- Lossless methods are explored.
- Several approaches exist.
  - Dictionary Compression (ZIP, LZMA…).
  - mSEED.
  - Audio coding.
  - Integer Coding (Bin Dong et al., 2022).
  - Inter-channel coding (Chengjun Wu, 2023).



Opportunities for compression

- Exploit redundancies:
  - **Temporal** (intra-channel)
    - Linear Predictive Coding.
  - **Spatial** (inter-channel)
    - Warping and scaling.
    - Frequency analysis.
  - **Energy** compaction
    - Transformation (e.g., DFT).
  - **Statistical** distribution
    - Entropy Coding.



Summary of methods for lossless DAS Compression

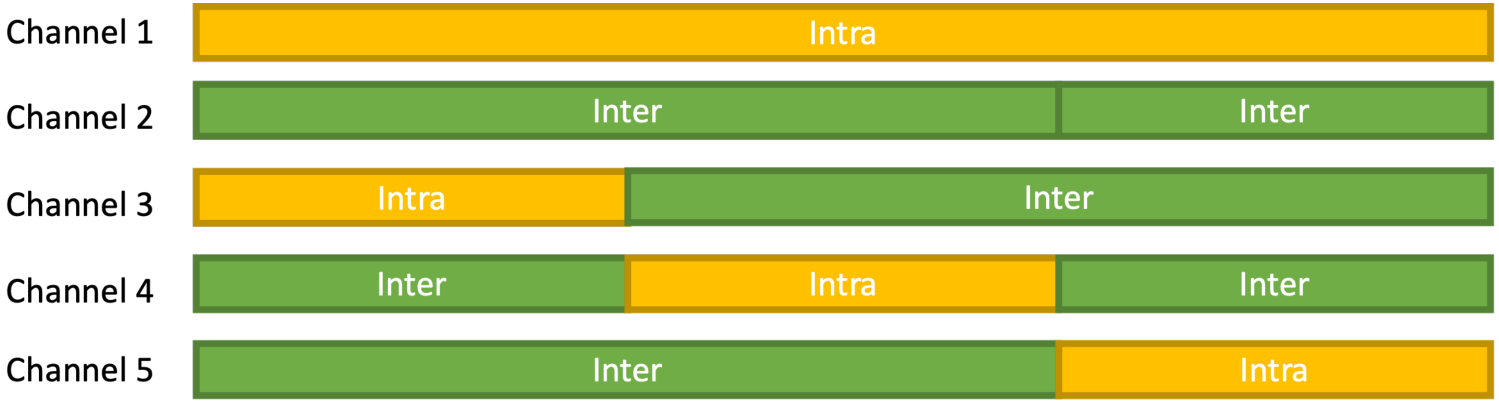
Method	mSEED	Dictionary compression (ZIP, GZIP, LZMA…)	Integer Coding (H5TurboPFor, Bin Dong et al., 2022)[1]	Inter-channel prediction (Chengjun Wu et al., 2023)[2]	Ours
Summary of techniques	• Differential encoding • Steim algorithm (consecutive differences)	• LZ77 (Lempel-Ziv 1977) algorithm • Huffman Coding	• Integer compression using TurboPFor • Optimized for speed (real-time)	• Linear Predictive Coding • Inter-channel scaling • Golomb-rice coding • Arithmetic Coding	• Linear Predictive Coding • Inter-channel scaling and shifting • Discrete Cosine Transform • Adaptive Arithmetic Coding • Exponential-Golomb
Compression	~1.25-1.5x	~1.25-1.5x	~2x	~2-3x	~3x

[1] ***Real-time and post-hoc compression for data from Distributed Acoustic Sensing***, Computers & Geosciences, Volume 166, 2022, 105181, ISSN 0098-3004, https://doi.org/10.1016/j.cageo.2022.105181

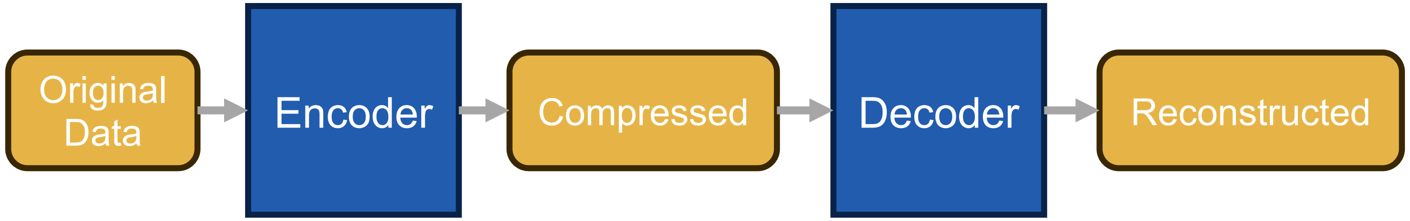
[2] ***A Lossless Data Compression Method for Distributed Acoustic Sensors.*** Optical Fiber Sensors. Optica Publishing Group, 2023.

Encoder-decoder design and partition into segments

- 1-D signals are split into **segments**.
  - **Intra** type.
  - **Inter** type.
- Every segment has its own set of parameters.



Example of segment partitioning and types.



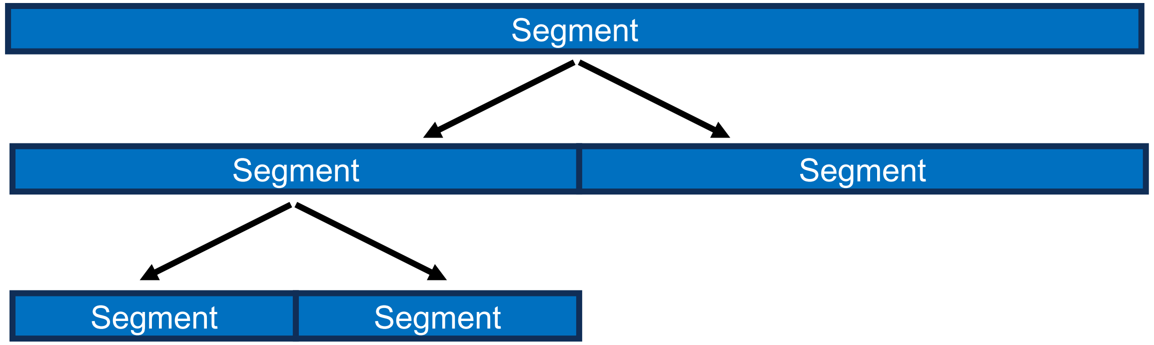
Encoder-decoder design.

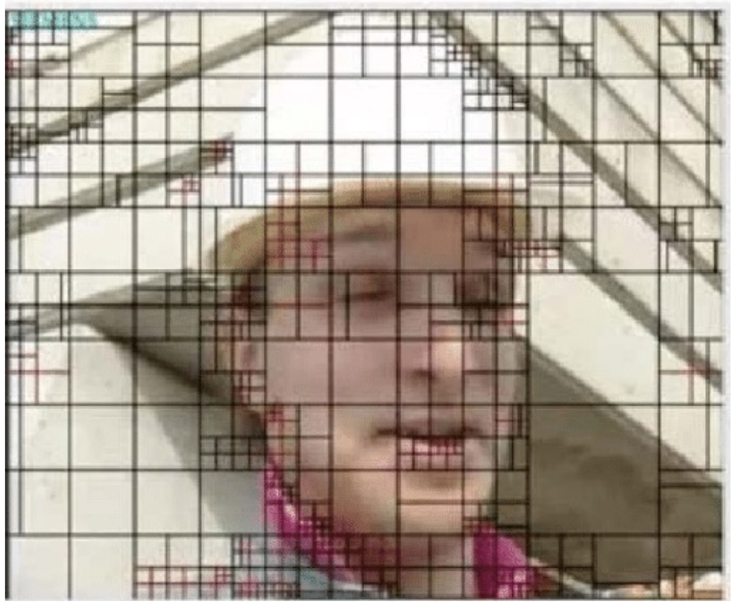
- The encoder chooses:
  - Signal partition into segments.
  - Segment parameters.
- The decoder reconstructs the signal.

Recursive partition search.

Encoder algorithm: recursive partitioning

- Inspiration from **image coding**.
- **Recursive search** of segment length with optimal parameters.
- Adapt to the **signal features**, more resolution in high-frequency areas.





High Efficiency Video Coding (HEVC)[1]

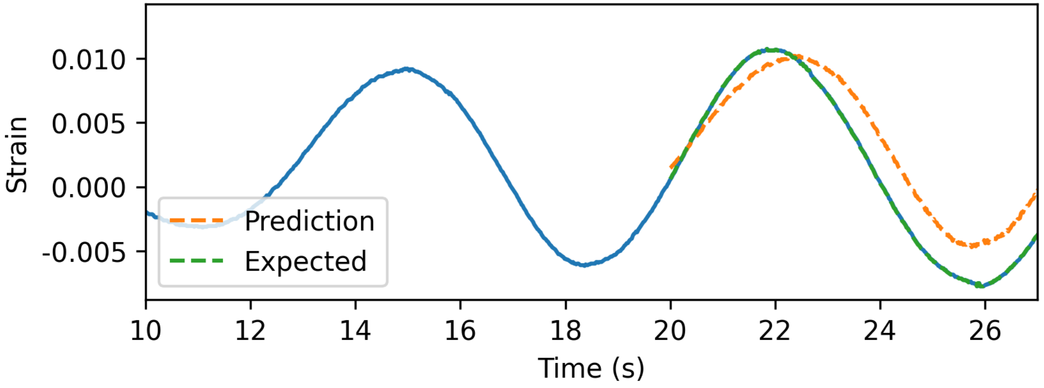
[1] Overview of the High Efficiency Video Coding (HEVC) Standard, https://doi.org/10.1109/TCSVT.2012.2221191

Prediction coding: **intra** and **inter**-channel methods

Intra-channel prediction

- Approximate samples by a linear combination of **past** samples (LPC).

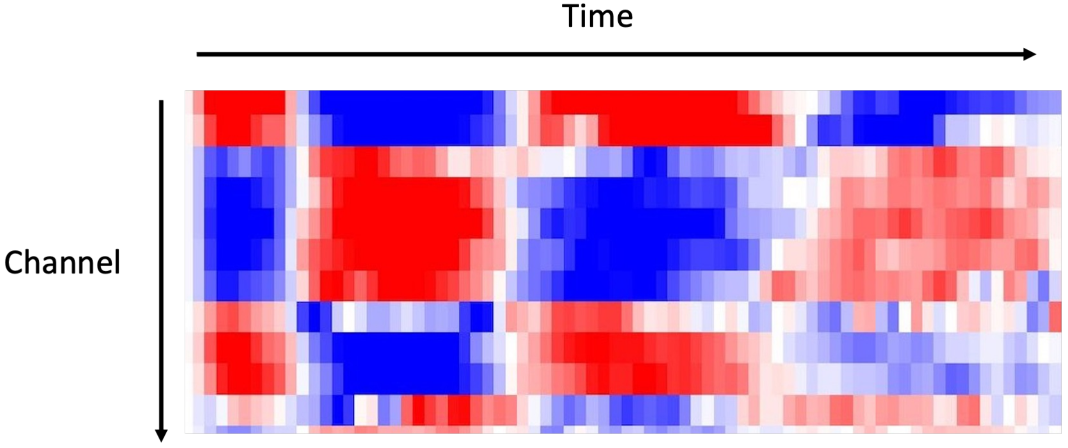
$$c_i(t) = \sum_{k=1}^{K} a_k c_i(t-k)$$



Inter-channel prediction

- **Scale** and time **shift** a **consecutive** channel

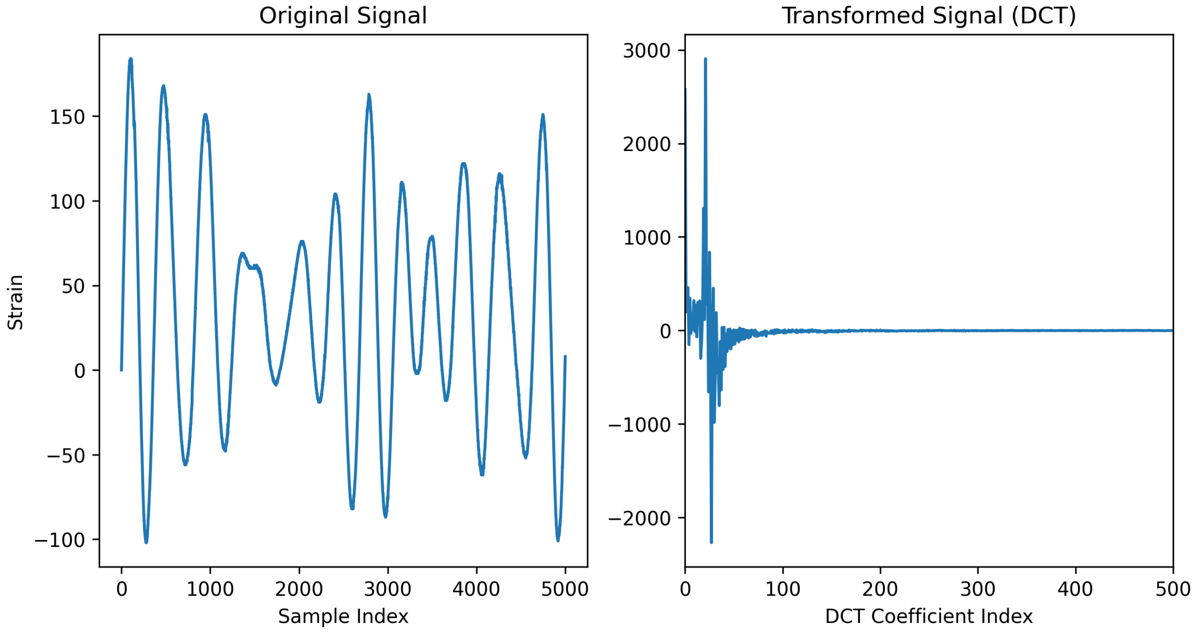
$$c_i(t) = a\, c_{i-1}(t-s)$$



DAS example data, filtered to >3Hz.

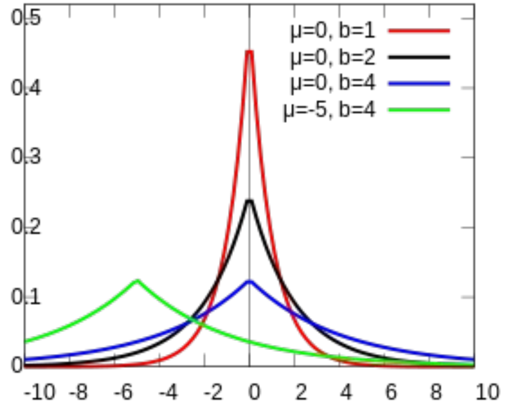
Transform and Entropy Coding

- The residuals are linearly transfomed for **energy compaction** using the Discrete Cosine Transform.



- The **entropy** (H) expresses the limit on how much a source channel can be losslessly encoded.
  - Shannon's source coding theorem: possibility to encode using $\boldsymbol{H+\epsilon}$ **bits** per symbol.

- The residuals of transformed signals typically exhibit a **Laplace distribution**.



More on Entropy Coding: Arithmetic Coding and Exponential-Golomb

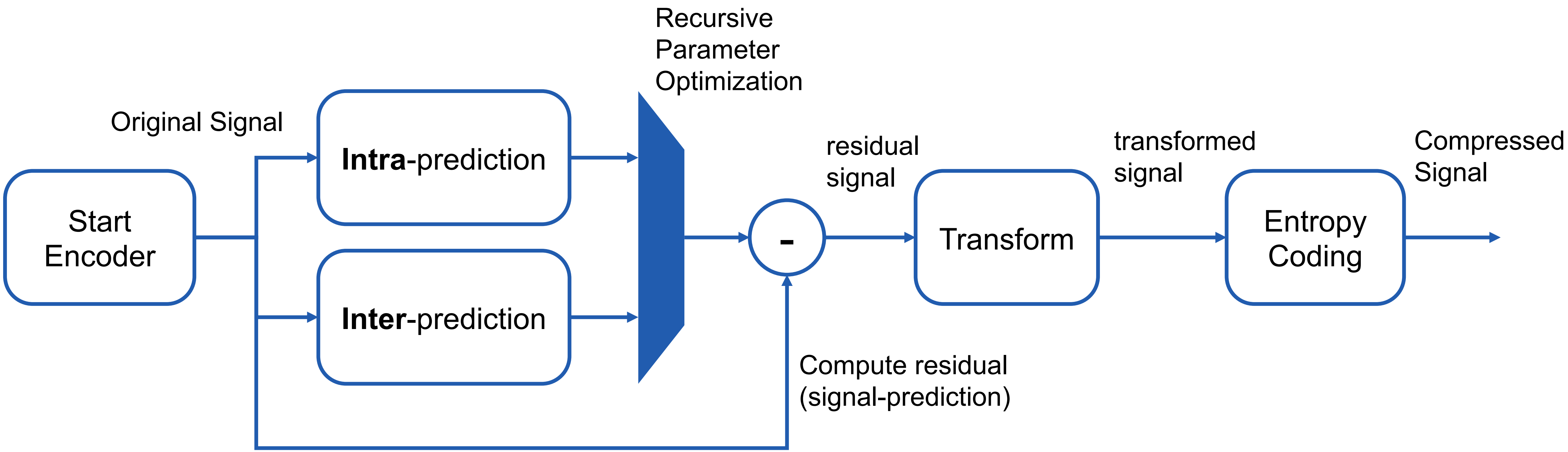
- A good entropy coder is the key for modern compression.
- We combine two methods:
  - **Arithmetic coding** for bits up to a threshold (fixed number of bits).
  - **Exponential-Golomb** for the bits above the threshold (handles outliers, any number of bits).

E-G	Bits (Arithmetic)	Symbol	
	10110	: 22	
	11011	: 27	
11	10011	: 115	Outlier

- Arithmetic coding assigns **shorter sequences** to more **probable symbols**.
  - Adaptive: update variance as new data is read.

- Exponential-Golomb represent values in **any range**, efficiently encoding large values.

Detailed Encoding Pipeline



Results

- Results tested on cables with different characteristics.
- Metric: compressed size ratio (Compressed/Original)

Dataset	Instrument	Original Format	Channels	Samp [Hz]	Channel Spacing [m]	Compressed size
Candas2	Aragon Photonics	H5	6496	50	10	**32%**
BALALINK	Aragon Photonics	H5	5984	250	10	**27%**
ETH Istanbul[1]	Silixa iDAS	ProdML	500	100	10	**60%**
NORSAR[1]	ASN OptoDAS	H5	1225	100	20	**56%**

[1] ***The Global DAS Month of February 2023.*** Seismological Research Letters (2024) 95 (3): 1569–1577. https://doi.org/10.1785/0220230180

Conclusions

- DAS Data can be losslessly compressed up to good ratios.
- Video coding techniques can be adapted to DAS data.
- Inter-channel prediction leverages correlation between channels for compression.
- An encoder was effectively implemented.

Future Steps

- Frequency domain inter-channel compression (wavelet transforms).
- Lossy compression (up to x20 smaller file sizes! [1])
- Faster implementation for real-time applications.

[1] F. Yu *et al.*, ***Data Compression Solution for Φ-OTDR: Encoding the Extracted Spectrum***, in *IEEE Sensors Journal*, https://doi.org/10.1109/JSEN.2024.3407752

Aleix Seguí
asegui@ethz.ch

Arantza Ugalde
a.ugalde@icm.csic.es

Sergi Ventosa
sventosa@icm.csic.es

References

- Bin Dong, Alex Popescu, Verónica Rodríguez Tribaldos, Suren Byna, Jonathan Ajo-Franklin, Kesheng Wu, **Real-time and post-hoc compression for data from Distributed Acoustic Sensing**, Computers & Geosciences, Volume 166, 2022, 105181, ISSN 0098-3004, https://doi.org/10.1016/j.cageo.2022.105181.
- Wu, Chengjun, et al. "**A Lossless Data Compression Method for Distributed Acoustic Sensors.**" Optical Fiber Sensors. Optica Publishing Group, 2023.
- ***The Global DAS Month of February 2023.*** Seismological Research Letters (2024) 95 (3): 1569–1577. https://doi.org/10.1785/0220230180