Lossless and Lossy **Compression** for Distributed Acoustic Sensing Using Inter-channel Predictions

Aleix Seguí[1], Arantza Ugalde[2], Sergi Ventosa[2], Josep Ramon Morros[3]
[1] ETH Zürich; [2]Institut de Ciències del Mar; [3]Universitat Politècnica de Catalunya

Motivation

- 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.

Recording at **~1 MB/s**

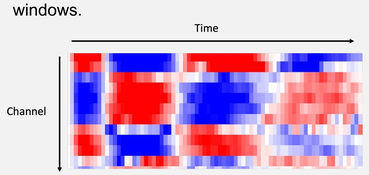
Storing **> 50 GB/day**!

Compression Opportunities

We exploit redundancies in 2D signals.

- **Temporal** (intra-channel)
  - Linear Predictive Coding.
- **Spatial** (inter-channel)
  - Frequency analysis.
- **Statistical** distribution
  - Entropy Coding.

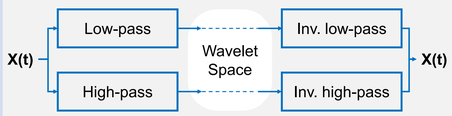
Data is sampled in non-overlapping windows.

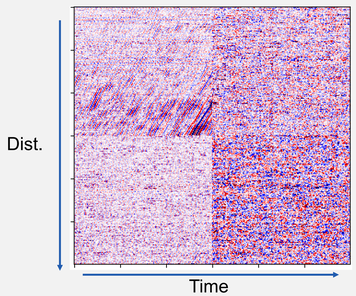


Compression Process and Techniques

2D Wavelet Filters

- Apply LGT 5/3 Wavelet decomposition in both dimensions
- Separates signals from high-frequency noise.

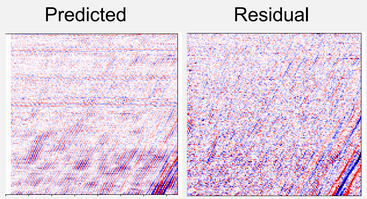




Linear Predictive Coding

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

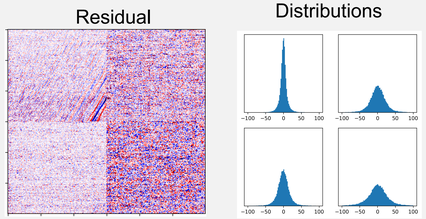
$$c_i(t) = \sum_{k=1}^{K} a_k c_i(t-k)$$



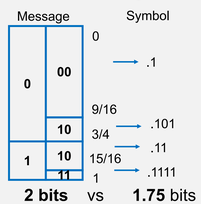
- Applied only along the time dimension in the low-frequency region.
- The coefficients are quantized and transmitted in the bitstream.
- LPC accurately models low frequencies.

Context-Adaptive Arithmetic Coding

The **entropy** (H) expresses the limit on how much a source channel can be losslessly encoded.

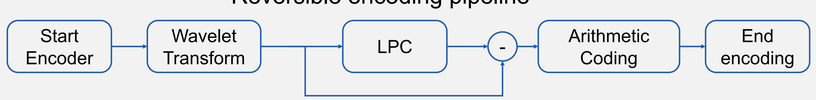


- A different model is needed for every region to model the entropy. We use a conditional probability mode, $p(x_n) = p(x_n|x_{n-1}, x_{n-2}, \dots)$, conditioned to past samples.
- Arithmetic coding assigns different **number of bits** to different messages according to the probability.



- Decompose numbers in **sign** (+ o -), base range (**0-16**), high range (**>16**)

Reversible encoding pipeline



Compression Results

Data			H5 GZIP	H5TurboPfor	JPEG2000	Ours
Interrogator	Freq.	Tested Channels				
Aragon Photonics	50 Hz	3000	1.22	1.14	2.31	**2.53**
ASN OptoDAS	100 Hz	1000	1.38	1.10	2.81	2.80

Compression factor (size is divided by xN)

Sign up for release!

Lossy Mode

- High-frequency noise is responsible for most of the file size.
- Integer quantization is application-dependant and quality metrics should be studied.
- Lossy compression has great potential, with **compression factors of more than x5**.

Conclusions and Future Work

- Compression is an effective solution for storage of DAS data.
- LPC and Wavelets effectively achieve decorrelated signals.
- An efficient implementation is necessary and will be released soon.
- Lossy compression should be considered for some applications.

Acknowledgements

- This work was partly funded by the Spanish Ministry of Science, Innovation and Universities, grant PID2020-117142GB-I00 [DeeLight project].
- Funding was provided by the European Union NextGenerationEU/PRTR Program projects PSI (ref. PLEC2021-007875) and TREMORS (ref. CPP2021-008869).

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