Real-time lossless and lossy compression of MBES water column

Jordi Portell[1,2]
David Amblàs[3,4]
Garrett Mitchell[5]
Matías Morales[6]

[1] Dept. Física Quàntica i Astrofísica, Institut de Ciències del Cosmos (ICCUB), Univ. Barcelona (IEEC-UB), 08028 Barcelona, Spain
[2] DAPCOM Data Services S.L., Parc UPC-PMT, 08860 Castelldefels (Barcelona), Spain
[3] Scott Polar Research Institute, University of Cambridge, CB21ER, UK
[4] CRG Marine Geosciences, Dept. of Earth and Ocean Dynamics, University of Barcelona, 08028 Barcelona, Spain
[5] Fugro USA Marine Inc., Houston, Texas, US 77081
[6] Kongsberg Maritime, 03570 Villajoyosa, Spain

Contact:
jportell@fqa.ub.edu
damblas@ub.edu

Context

Multibeam echosounders allow collecting **water column acoustic images**
→ Mapping of gas seeps, imaging of fish and marine mammals, oceanographic studies, location of mid-water targets, determination of sunken structures…

Vast amounts of data to be acquired, stored and transferred
→ Water column data acquisition **often avoided** or only temporarily activated.

Data compression mitigates this problem, but very few studies available
→ Develop a **tailored data compression** solution for this kind of data.

Kongsberg MBES water column datagrams not very different from science data format of *Gaia* space mission of the European Space Agency.

FAPEC is a **high-performance versatile data compressor** based on an entropy coding method similar to that being used onboard Gaia.

We present an adaptation of **FAPEC to the MBES water column datagram file format**.

Artistic impression of the Gaia space observatory mapping the Milky Way
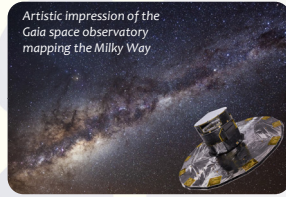
Features of Kongsberg MBES water column data

Binary files with time stamps, instrumental and sensors information, navigation, attitude… and **raw samples**.
Non-uniform 2D array: pings and samples per beam.

Beams per ping: depending on model, swath aperture angle, resolution and scene (depth).
Samples per beam: strong variation with beam angle; also with scene throughout the beams.
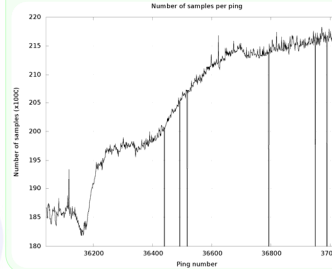

Samples per ping vs. depth: shallow waters mean shorter pings and more samples.

Representation of backscatter raw samples from Kongsberg EM302

Samples per beam vs. angle (4 pings shown). Depth increases left, ping time increases down. Vertical dimension relates to beam angle. Focus of parabola indicates location of echosounder.
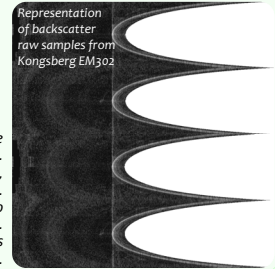
FAPEC tailoring

Non-uniform array of data → Typical **image compressors cannot be directly applied**.
FAPEC has 2 stages: **pre-processing** + entropy coding → pre-proc can be tailored.
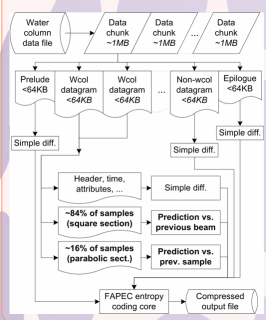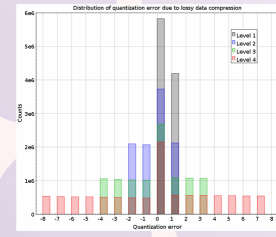

Focus on raw samples, properly handling different sections of the array. **Prediction-based differential coding** used.
File is split in chunks: enforce **robustness** and multi-threading.

Lossless by default.
Lossy option. Rounding quantizer for better quality.
Fully operational and reliable solution.
Software implementation (ANSI C).

~84% of samples (square section): **Prediction vs. previous beam**
~16% of samples (parabolic sect.): **Prediction vs. prev. sample**

Test setup and lossless data compression results

EM2040 (Kongsberg): Barcelona harbour, 2015 (shallow water, structural elements, shoaling fish), ~4.3 MB/s
EM302 (Fugro): Gulf of Mexico, 2015 (~1000 m, actively emitting hydrocarbon fluids, "GC600"), ~0.1 MB/s
EM710 (CRG UB): Outer continental shelf of south-eastern Iberian peninsula, ~1 MB/s

FAPEC 18.0 Beta vs. some 'standard' lossless compressors. Linux laptop (Core i7 2.8GHz), single-core tests. Also ARM tests.

**FAPEC offers the best compression ratios. Also speeds: >60MB/s in laptop (8MB/s in ARMv7 800MHz)**
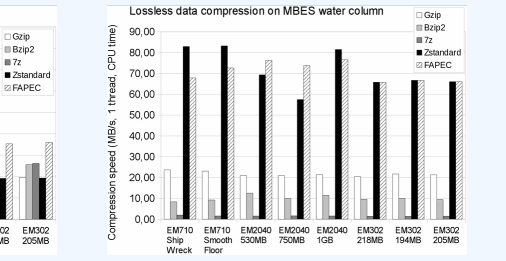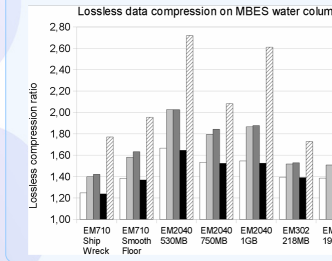
Lossy data compression results

Lossy FAPEC compression quantizes raw samples from original 256 grey shades to 128 (*level 1*), 64 (*level 2*)… up to just 8 shades (*level 5*). **Full spatial resolution is kept**.
We obtain **ratios over 2.1** with **nearly perfect quality** (PSNR > 50 dB). **Water column files smaller than bathymetry files** can be reached with the lowest quality (still good, PSNR > 30 dB).
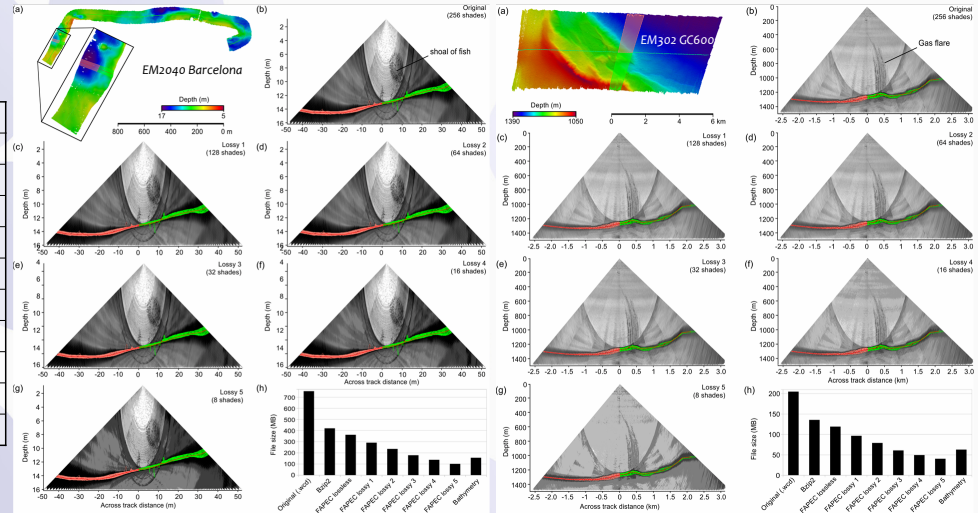**Features in the water column** can clearly be seen even with lowest-quality levels.
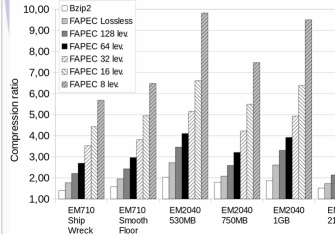
Data volume generated during **1 hour** of continuous **water column acquisition**:

	EM302	EM710	EM2040
Ping rate:	Low	Medium	High
Raw output file (.wcd):	306MB	4.2GB	14.9GB
Compressed with gzip:	220MB	3.0GB	9.6GB
Lossless FAPEC:	177MB	2.1GB	5.7GB
Lossy-3 FAPEC (32 shades):	90MB	1.1GB	3.0GB
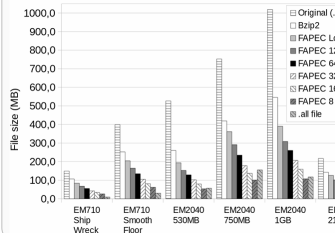
Use cases

Offshore:
- On-the-fly lossless compression and error-resilient backup.
- Continuous water column acquisition with low-quality lossy compression (similar size to bathymetry, precious extra info).
- Automatic detection of large-scale features in water column scene from variations in chunk ratios (see figure).
- Real-time compression in environments with limited resources, such as AUV or USV.
- Remote diagnostics and diagnostics (data transfers through expensive links).

Kongsberg GeoSwath 4R USV

Onshore / Lab / Data centre:
- Compression and archival of existing data files (better resiliency to eventual file corruption).
- Processing or visualization directly from compressed files (on-the-fly decompression). Lower I/O requirements.

Conclusions

- MBES water column data: valuable information, but prohibitive data cost.
- Existing data compressors: low ratios and slow compression speeds.
- FAPEC data compressor adapted to Kongsberg water column datagrams.
- Lossless and lossy compression available (different quality levels).
- Chunk-based operation: Error-resilient, ratios monitoring, multi-thread.
- Fully operational software: Linux, Windows and Mac. AES encryption available.
- Best lossless ratios at very high speeds. Higher ping rate leads to higher ratios.
- Offshore real-time compression feasible.
- Excellent lossy ratios still with good image quality. Scene features clearly kept.
- Wealth of applications, offshore and onshore.
- Further improvements being investigated: better ratios, features detection.

FAPEC is a real and reliable solution to a real problem.
Free evaluation licenses can be requested at www.dapcom.es or to the authors.

Acknowledgements