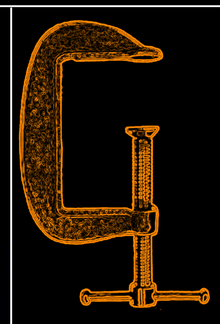
SKA-Squeeze

Non-standard
Compression Transformation

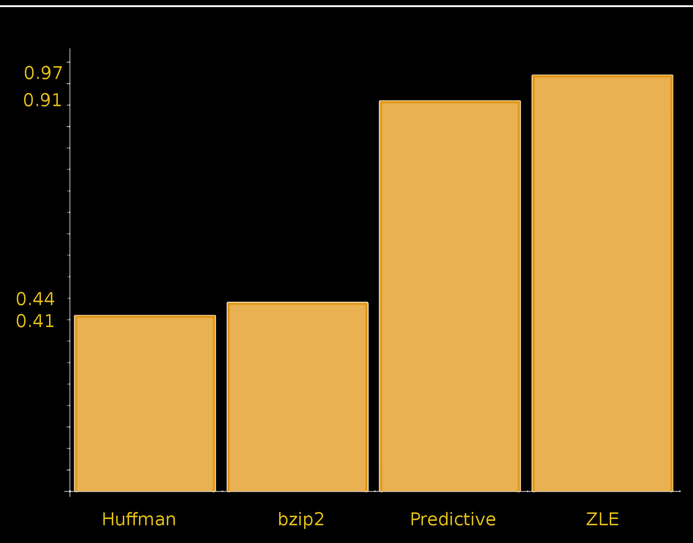
SKA AFRICA
SQUARE KILOMETRE ARRAY

Description

- Situated in South Africa and Australia
- Construction of 3000 antennae array in progress
- Produces up to 1 Petabyte of samples every 20 seconds
- Lossless compression required for network and storage
- Investigation into block-based parallel compression

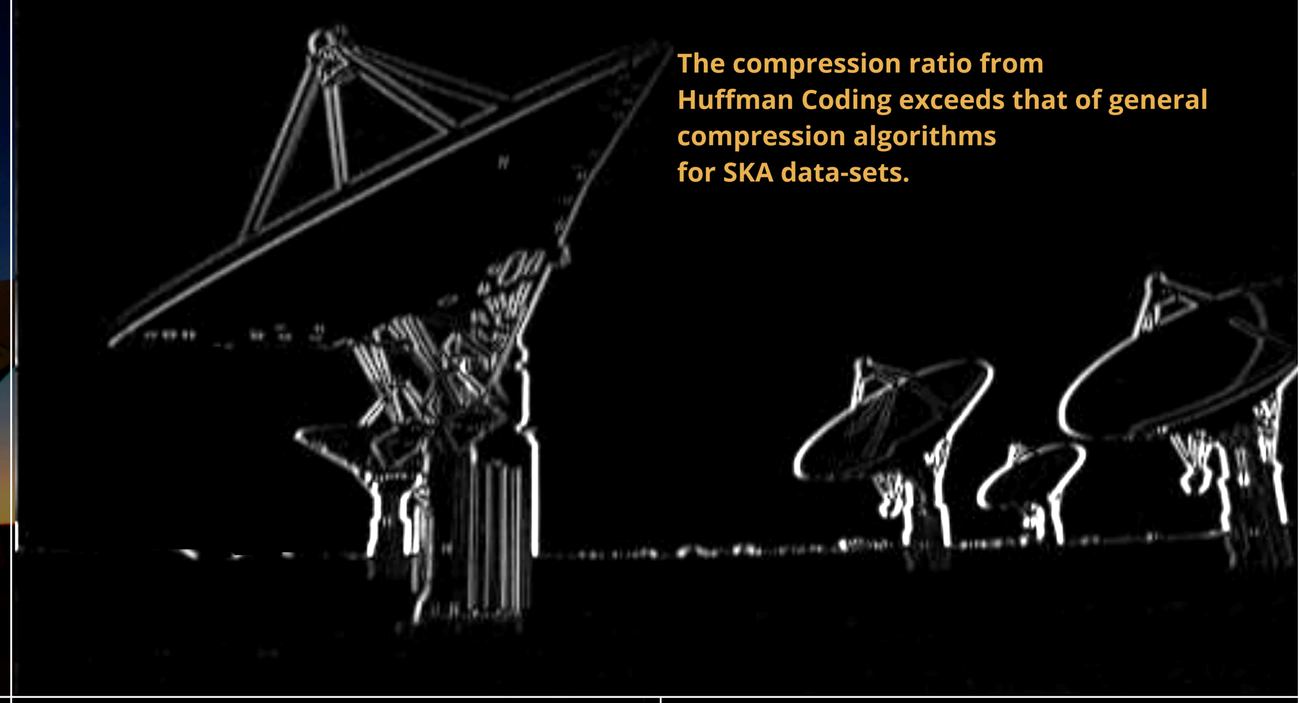


Compression Ratio



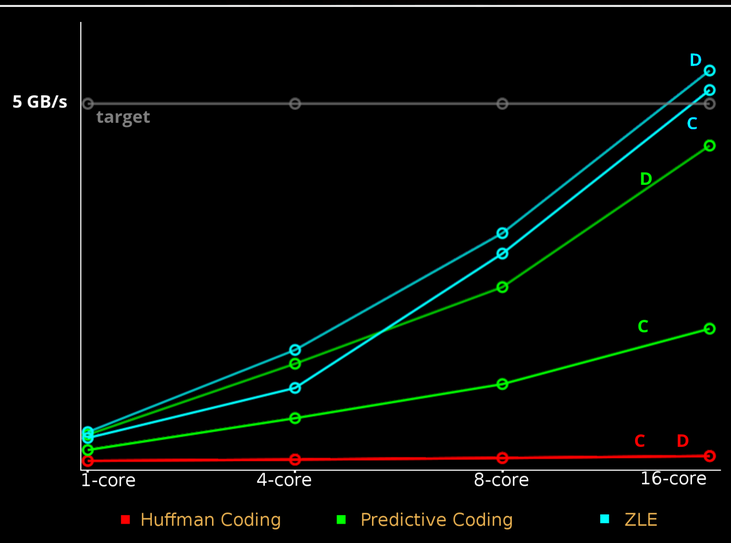
Aims

1. To achieve sustained compression throughputs of 5 GB/s on multicore CPUs and GPUs.
2. Obtaining effective compression ratios comparable to those of standard compression utilities.

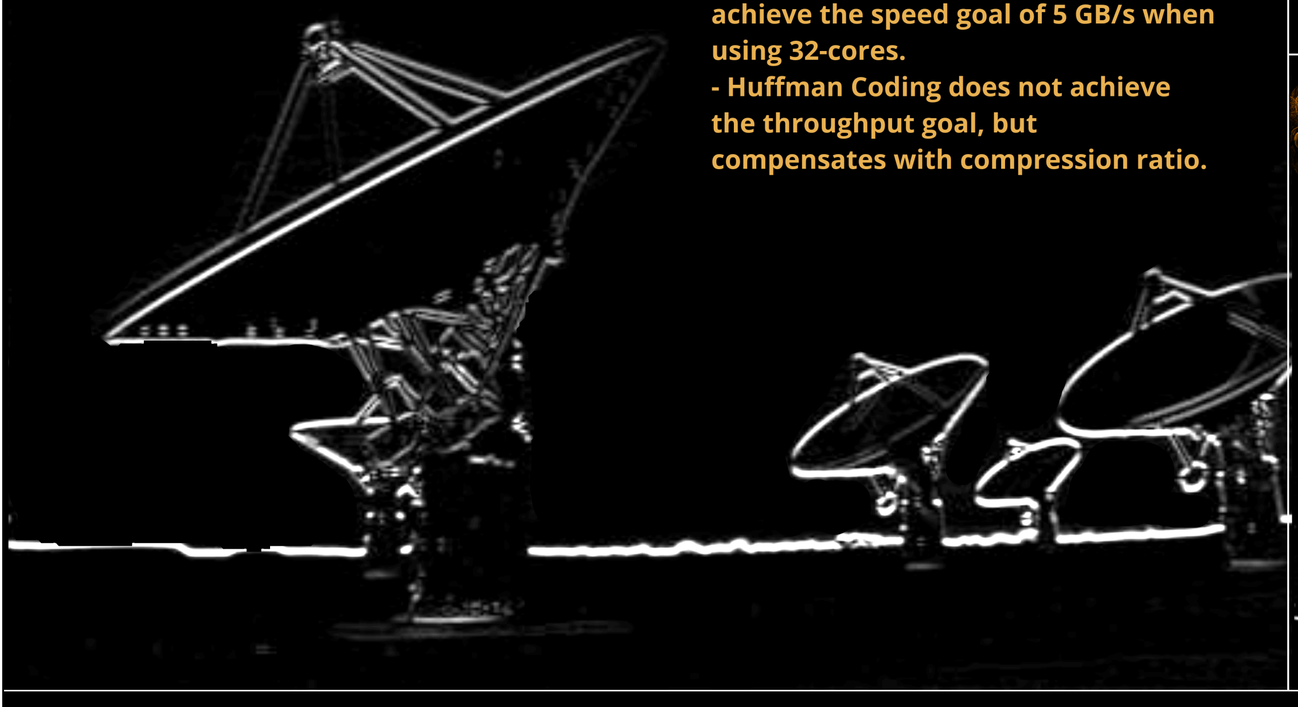


The compression ratio from Huffman Coding exceeds that of general compression algorithms for SKA data-sets.

Throughput

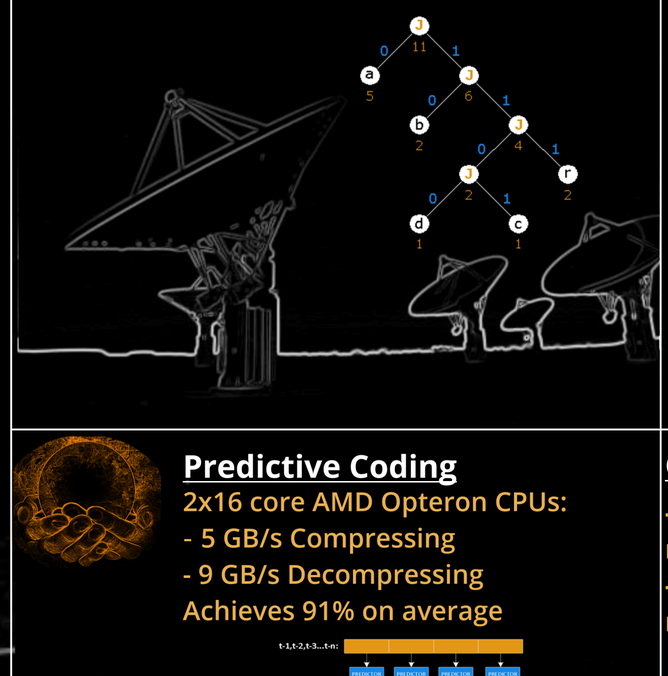


- Both Predictive Coding and ZLE achieve the speed goal of 5 GB/s when using 32-cores.
- Huffman Coding does not achieve the throughput goal, but compensates with compression ratio.



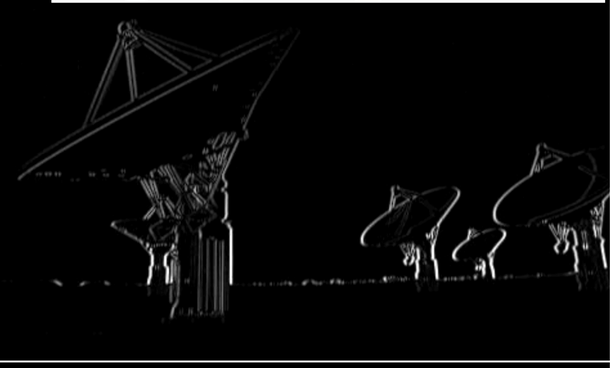
Huffman Coding

- 79 MB/s throughput (16 cores on an Intel Xeon E5 and a Tesla K20X)
- 41.18% Compression Ratio



Zero-Length Coding

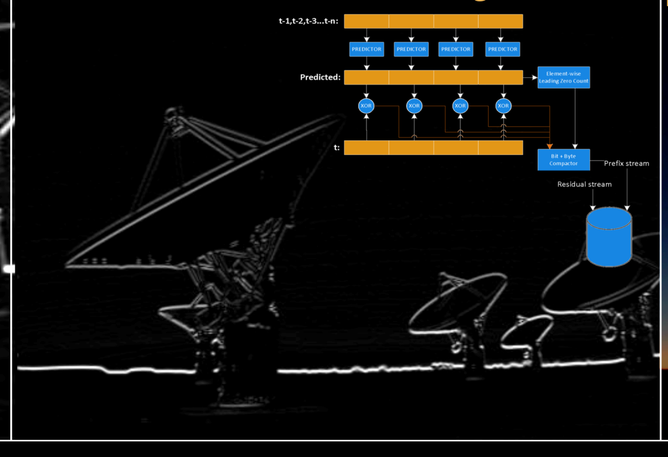
- 5+ GiB/s compressing and decompressing (Xeon E5)
- 96-97% Compression ratio



Predictive Coding

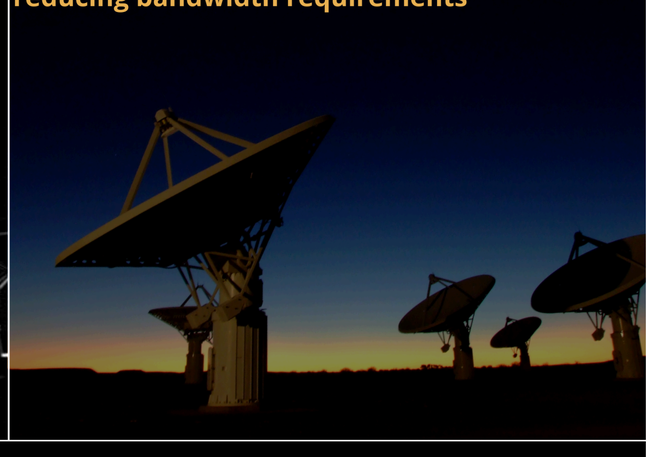
2x16 core AMD Opteron CPUs:
- 5 GB/s Compressing
- 9 GB/s Decompressing

Achieves 91% on average



Conclusion

- Use Huffman Coding for long-term storage requirements
- Use Predictive Coder / Zero-Length Coder for reducing bandwidth requirements



Department of Computer Science

University of Cape Town
Private Bag X3
Rondebosch 7701
Email: dept@cs.uct.ac.za
Tel: 021 650 2663

Supervisors

A/Prof. James Gain
A/Prof. Patrick Marais
Jason Manley (external)

Team

Benjamin Hugo
Brandon Talbot
Heinrich Strauss

University of Cape Town • iYunivesithi yaseKapa • Universiteit van Kaapstad
SPES BONA