

Lossless Data Compression Via Error Correction

Sergio Verdú

Dept. Electrical Engineering, Princeton University,
Princeton, New Jersey 08540, USA
verdu@princeton.edu

Abstract. This plenary talk gives an overview of recent joint work with G. Caire and S. Shamai on the use of linear error correcting codes for lossless data compression, joint source/channel coding and interactive data exchange.

1 Summary

Over the last five decades, significant inventions have led to data compression and data transmission systems whose efficiency approaches Shannon's fundamental limits [1]. Error-correcting codes now exist (i.e. sparse-graph linear codes) that can achieve performance close to channel capacity with complexity and delay that are tolerable for many applications. Similarly, lossless data compression

algorithms exist (most notably the Lempel-Ziv algorithm) that can provably achieve the entropy rate of a wide class of sources with very low complexity. Curiously, although Shannon's development of the theories of fundamental limits for data compression and transmission shared very strong commonalities, there has been essentially no intercourse between the respective constructive theories throughout their long histories.

While Shannon's separation principle establishes no loss in asymptotic performance when compression and transmission are performed separately, it has long been expected (but not fully realized) that, in the nonasymptotic regime, gains may accrue by joint design. Furthermore, in systems such as packet-oriented wireless high data rate systems, it is sometimes cumbersome to design systems based on the separation principle.

Lossless data compression algorithms find numerous applications in information technology, such as packing utilities (e.g. gzip), modem standards, fax standards, back-end of lossy compression algorithms (e.g. JPEG and MPEG), and compression of headers of TCP/IP packets in wireless networks.

Indeed, the field of lossless data compression has achieved a state of maturity, with algorithms that admit fast (linear-complexity) implementations and achieve asymptotically the fundamental information theoretic limits.

The availability of linear codes (such as the low-density parity check codes) that allow for very efficient encoding/decoding algorithms while operating near the Shannon limit makes their application in data compression competitive with state-of-the-art methods while not suffering from some of their shortcomings.

J.R. Correa, A. Hevia, and M. Kiwi (Eds.): LATIN 2006, LNCS 3887, pp. 26–27, 2006.

© Springer-Verlag Berlin Heidelberg 2006