CCIT Report #893 January 2016

Switch Codes: Codes for Fully Parallel Reconstruction

Zhiying Wang, *Member, IEEE*, Han Mao Kiah, *Member, IEEE*, Yuval Cassuto, *Senior Member, IEEE*, and Jehoshua Bruck, *Fellow, IEEE*

Abstract

Network switches and routers scale in rate by distributing the packet read/write operations across multiple memory banks. Rate scaling is achieved so long that sufficiently many packets can be written and read in parallel. However, due to the non-determinism of the read process, parallel pending read requests may contend on memory banks, and thus significantly lower the switching rate. In this paper we provide a constructive study of codes that guarantee fully parallel data reconstruction without contention. We call these codes "switch codes", and construct three optimal switch-code families with different parameters. All the constructions use only simple XOR-based encoding and decoding operations, an important advantage when operated in ultra-high speeds. Switch codes achieve their good performance by spanning simultaneous disjoint local-decoding sets for all their information symbols. Switch codes may be regarded as an extreme version of the previously studied batch codes, where the switch version requires parallel reconstruction of all the information symbols.

Index Terms

Distributed-storage codes, network switches, batch codes, combinatorial designs.

I. Introduction

Consider a shared memory system required to serve write and read requests at a certain rate. In the write path, k fixed-size packets arrive each time unit, and need to be stored in the memory system. In the read path, each time unit the memory system needs to output a requested set of k previously written packets. To meet these requirements, the system uses n banks of physical memory, where each memory bank works at a rate of one packet write and one packet read each time unit. The design objective of the system is to minimize the number of banks n that are needed to fulfill the above mentioned read/write specifications. Figure 1 gives a pictorial description of such a memory system.

The main application for such a memory system is within *network switches* (and similarly routers), wherein the memory system is used as a *switching fabric* writing packets upon their inbound arrival, and later reading them for their outbound transmission. Two features of the abstract system model are especially fitting for switching applications: 1) the symmetry between read and write rates – each at k packets per time unit, and 2) flexibility to choose the k read packets from the currently stored packets. The first feature is required for flow conservation in the switch, and the second provides flexibility to accommodate priorities, congestion, blocking, and other factors affecting the packet read schedule.

The main challenge faced by the switch memory system is *contention* between the requested packets on the bandwidth of the memory banks. Simply put: if a bank is used to output one of its packets in a time unit, then it cannot output another packet in the same time unit. For example, consider the simple case of $k = 2$ packets used with $n = 2$ banks. This scenario is depicted in Figure 2. Packets are marked by letters progressing lexicographically with arrival time. Packets arrived in the same time unit are called a *generation*. Each generation contains $k = 2$ packets and are stored in the $n = 2$ banks. So for the write path $n = 2$ banks are sufficient. However, it is clear that in the read path the system of Figure 2 does not work, because requests like (A, E) or (D, F) cannot be served at a single time unit. From the example of Figure 2 it is clear that supporting arbitrary packet requests in the read path

Zhiying Wang is with the Center for Science of Information, Stanford University, Stanford CA USA (email: zhiyingw@stanford.edu).

Han Mao Kiah is with the School of Physical and Mathematical Sciences, Nanyang Technological University, Singapore (email: hmkiah@ntu.edu.sg).

Yuval Cassuto is with the Department of Electrical Engineering, Technion – Israel Institute of Technology, Haifa Israel (email: ycassuto@ee.technion.ac.il).

Jehoshua Bruck is with the Department of Electrical Engineering, California Institute of Technology, Pasadena CA USA (email: bruck@caltech.edu).

This work was done when Han Mao Kiah was a postdoctoral research associate at the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign. The work of Y. Cassuto was supported in part by the European Union Marie Curie CIG grant, by the Israel Science Foundation joint ISF-UGC program, and by the Israeli Ministry of Science and Technology. The authors wish to thank Omer Shaked for his contributions to the development of the switch-code model. Part of the results in the paper were presented at the 2013 and 2015 IEEE International Symposia on Information Theory held in Istanbul and Hong Kong, respectively.