

Context-independent Codes for Off-chip Interconnects

Kartik Mohanram and Scott Rixner
Rice University, Houston, TX

Modern embedded networking, video, and image processing systems are typically implemented as systems-on-a-chip (SoC) in order to reduce manufacturing costs and overall power and energy consumption. By integrating all of the peripheral functionality directly onto the same chip with the core microprocessor, both chip manufacturing and system integration costs can be lowered dramatically. In addition to cost, managing power and energy is a first order constraint that drives the design of embedded systems based on SoCs. However, most modern SoC-based embedded systems require significant external memory that cannot be reasonably embedded into a single core. In such systems, the processor-memory interconnect can consume as much or more power than the core itself. Since the memory system is a major contributor to the overall power dissipation in SoC-based embedded systems, and since these systems will continue to require external memory for the foreseeable future, it is essential to develop advanced memory controller architectures to reduce the power dissipation of external memories and the interconnect between the embedded SoC core and that memory.

Encoding data that is stored in memory can minimize the power consumed by the processor-memory interconnect. Dynamic power is consumed by the interconnect drivers when there are bit transitions. To minimize this power, double-ended, context-dependent codes such as the bus-invert code have previously been proposed. Double-ended codes encode data at the transmitter and decode it at the receiver. For a processor-memory interconnect, this implies that the DRAM also needs to participate in such codes. Context-dependent codes use the value last transmitted on the interconnect as well as the current data in order to encode the data to minimize transitions. For example, bus-invert coding either transmits the data value unchanged or its inverse, depending on which value minimizes transitions on the interconnect. If the DRAM were modified to support such coding, bus-invert coding could reduce transitions on the interconnect by 22%.

In contrast, single-ended, context-independent codes are much simpler to implement, as they do not require modifications to the DRAM and are stateless. This paper introduces the concept of a frequency-based single-ended, context-independent code for interconnect power reduction. The simplest frequency-based code simply remaps the input space based upon the measured or expected frequency of occurrence of each data value. Despite the fact that such a code is context-independent and so does not account for possible switching on the interconnect, it is able to reduce the transitions on the interconnect by 28% on average. This simple code results in a larger power decrease on the interconnect than context-dependent bus-invert codes that are explicitly designed to minimize switching activity. Furthermore, frequency-based coding can also be used to augment limited-weight codes (LWCs). A limited-weight code maps the input data to a wider codeword in which the number of bits that are set is restricted. The proposed frequency-based assignment of codewords using a LWC can reduce transitions on the interconnect by an average of 30% over the uncoded case.

Frequency-based context-independent codes reduce interconnect power consumption without requiring the use of specialized DRAM. Despite the fact that such codes do not seek to minimize transitions on the interconnect, they incur far fewer transitions than bus-invert coding, which is considered the best practical context-dependent code. Frequency-based codes are effective because frequently occurring values usually follow either themselves or other frequently occurring values on the interconnect. So, if the most frequently occurring values are all mapped to codewords that are close (Hamming distance-based) to each other, then switching activity can be minimized. In this manner frequency-based codes simply, but effectively, reduce dynamic power consumption on interconnects to commodity DRAM.