Architecture and Algorithm Optimization for DSP Transforms with Large Data Size


Berkin Akin


Peter Milder


F. Franchetti


James Hoe

Linear transforms, the discrete Fourier transform (DFT) for example, are crucial kernel algorithms extensively used in several digital signal processing (DSP), scientific computing and related applications. Some of these applications including synthetic aperture radar (SAR) processing or medical imaging have large data sizes that often exceed feasible on-chip storage resources. These types of applications, whose data do not fit on-chip memory, require multiple off-chip data transfers making the problem memory bounded. Off-chip memory bandwidth becomes the bottleneck in the memory bounded problems. Memory access patterns play an important role in off-chip memory bandwidth utilization, significantly affecting the overall performance of the system. Therefore it is critical to manage off-chip bandwidth and on-chip storage effectively to get high performance for DSP transforms with large data size.

Most DSP transform algorithms do not naturally have suitable off-chip memory access patterns or they do not utilize hardware resources efficiently. However, by carefully performing "hardware aware" manipulation of high-level representations of the algorithms, we are able to balance across various resource constraints. Given a set of hardware platform parameters, we aim to produce a balanced hardware implementation (in FPGA or ASIC) with the goal of maximizing performance by saturating off-chip memory bandwidth at the lowest hardware cost possible.

Our proposed architecture, which is shown in figure 1, can handle: (1) off-chip data transfers by effectively utilizing memory bandwidth, (2) on/off-chip permutations, (3) overlapping the latency of memory access and computation by double buffering and (4) single or multi-dimensional data abstractions for multi-dimensional transforms. Although our proposed architecture can support various types of DSP transforms, we picked DFT as a target application. Figure 2 shows the performance estimates for 2D DFT on Virtex5 and StratixIV FPGAs according to our initial experiment and simulation results.

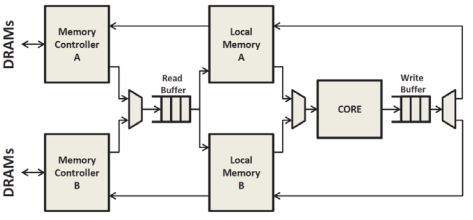

Fig. 1: Proposed architecture.

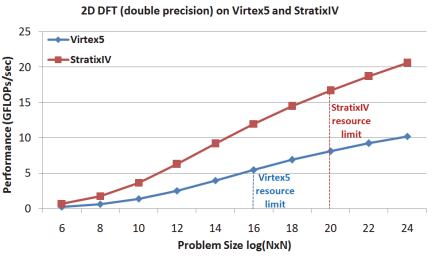

Fig. 2: Performance vs. problem size.