High Performance Computing Datapaths in FPGAs

Martin Langhammer,
Altera UK Ltd,
mlangham@altera.com

Holmers Farm Way
High Wycombe, Bucks
HP12 4XF
+44 1494 602 310

This abstract will examine the ability of FPGAs to effectively support HPC datapaths. Often the most difficult functions to implement in programmable logic are IEEE754 floating point operators, especially when multiple operators are combined to form large expressions. A simple way of calculating the maximum possible floating point capability of a given FPGA will be shown. A new datapath compiler tool will be described that will make it possible for FPGAs to consistently realize this ideal level of performance. A 90nm Stratix ® II FPGA can support in excess of 20 GFLOPs double precision, and 45 GFLOPs single precision; in a 65nm Stratix ® III device this more than doubles to 45 and 100 GFLOPs, respectively. Combining the high floating point performance of the devices with their exceptional flexibility will give them their greatest advantage in HPC applications – peak performance equals sustained performance for almost all algorithms.

FPGA Capabilities

The maximum floating point capability of an FPGA is largely limited by the number of dedicated hard multipliers in the device, which are required for the mantissa multiplication of a floating point multiplier. Although large multipliers, in the range of 24x24 bits to 54x54 bits can be implemented in soft logic, the large number of resources, placement constraints, and long carry chains in the adder trees affect system performance, and the amount of logic available for other floating point functions, such as adders. The Altera Stratix ® II DSP blocks are designed with four independent 18x18 multipliers, which can be combined into one 36x36 multiplier wholly contained in the block. Two blocks – one 36x36 multiplier, and four 18x18 multipliers – can construct a 54x54 multiplier with some external soft logic. The Stratix ® III DSP blocks also contain 54x54 mode, whereby most of the adder logic required to assemble the independent 18x18 multipliers is also contained in the DSP Block; a single soft logic adder is required to support a 54x54bit multiplier.

Many DSP algorithms, fixed or floating point, have a 1:1 ratio of multipliers to adders. For double precision, a multiplier and adder core operator pair require approximately 2000 4-input LUTs, in addition to the hard multiplier. In practice, an FPGA with this ratio would be impossible to route, and have no free resources to for interfaces, application wrappers, or routing overhead. Assuming a 50% device utilization by the datapath, a ratio of 4000 4-input LUTs per operator pair, or 500 4-LUTs per 18x18 multiplier could be routed.

FPGAs are designed for a diverse range applications, and many fixed point DSP applications require a higher multiplier to logic ratios than this, which means that the higher general purpose multiplier density can be wasted. A datapath compiler tool was developed to fully use a much higher multiplier density.

Floating Point Compiler

In a group of floating point operations, most inputs to a node will be denormalized, and most exits from a node will be normalized. By changing the representation of a floating point value, and allowing local overflow and underflow at any node, allows relaxation of the normalization, as long as the relative scale between any of the values in the datapath are maintained to the point where precision is not lost. Although local overflow and underflow are allowed in the datapath, global overflow and zero conditions are still tracked through the datapath, and considered in the type cast back to IEEE754 on the output of the datapath.

Normalization then occurs on a cluster level, where a cluster is a group of identical or similar operations. For most operations, positive wordgrowth is deterministic, and negative wordgrowth is predictable, often allowing the denormalization and normalization pair to be simplified. This can be illustrated by the single precision multiplier, where two 24 bit mantissa precision are multiplied by a 36 bit multiplier. The range of the possibly unnormalized input values can be estimated from the datapath structure feeding the multiplier node, and therefore only a coarse denormalization will be needed, to ensure that the mantissa multiplication does not overflow to the point of error, or underflow to the point of precision being lost. For double precision, a more conventional normalization is required before a multiplier, but the relaxed normalization is still applied with a cluster of identical operations.

Because the compiler needs to optimize the datapath at the expression or even group of expressions level, C language entry is used to describe the expressions. Simply stated, the compiler fuses together the entire datapath.

A Stratix ® II 2S180 device contains 384 18x18 multipliers, and 144K ALUT and register combinations, with a maximum of 48 double precision multipliers. With the compiler, a design with a 1:1 multiplier to adder ratio would contain 96 operators, giving 19.2 GFLOPs at a system speed of 200 MHz with only half the logic resources used. Other designs, like a Radix 4 FFT, with a high adder ratio could increase this by up to 50%.

Summary

Current FPGA contain enough multiplier resources to support HPC applications. The large logic requirement for floating point operations can be reduced using a datapath compilation tools, which will generate datapaths with up to 50 GFLOPs double precision and 100 GFLOPs single precision in current devices, with enough resources remaining for application implementation, and fitting.