The Linley Group

MICROPROCESSOR report

Insightful Analysis of Processor Technology

SiFive Raises RISC-V Performance

Series 7 Comprises First Superscalar RISC-V CPUs

By Bob Wheeler (November 12, 2018)

Designing a CPU that scales from microcontrollers to multicore processors is difficult, but that's SiFive's approach with its 7 Series. At the Linley Fall Processor Conference, the RISC-V startup revealed its latest CPU. The dual-issue in-order design is its most complex core yet, moving into the same class as Arm's "little" Cortex-A family. SiFive will offer versions for real-time embedded processing as well as Linux applications.

At the high end, the company's new U74MC intellectual-property (IP) core builds on the U54, which already offers multicore configurations and Linux compatibility. The standard U74MC includes a double-precision floating-point unit (FPU). Up to nine of the 64-bit cores can share an L2 cache with ECC protection. For deeply embedded designs, the company introduced the 32-bit E76 and 64-bit S76, which include a single-precision FPU. They improve performance compared with the existing E31 and E51 (see *MPR 6/5/17,* "SiFive Begins Licensing Cores"). RTL for the E76, S76, and U74 is now available.

Following a $50 million funding round announced in April, SiFive has sharpened its focus on IP for embedded applications. It also disclosed a license agreement with Western Digital, a strategic investor. Although the company announced standard 7 Series cores, part of its differentiation comes from configurability. Customers can start with the specification for an off-the-shelf core and add or remove standard instruction extensions, change memory details, and edit other features. Within weeks, SiFive delivers RTL for a core that consumes only as much area and power as the customer application allows.

Rocket Separation

The new dual-issue 7 Series CPU represents a departure from SiFive's previous designs, which are based on the open-source Rocket CPU. The U54 employs a simple five-stage scalar pipeline that achieves 1.5GHz in TSMC 28nm technology (see *MPR 10/9/17,* "RISC-V U54 Runs Linux"). It implements the RV64I base ISA plus the multiply and divide (M), atomic (A), and compressed (C) extensions. Optionally, it handles single-precision (F) and double-precision (D) floating-point extensions. SiFive is developing a vector unit for future 7 Series cores, but the RISC-V vector (V) extensions remain incomplete. (The ISA specification abbreviates the combination of I, M, A, F, and D instructions as G, denoting a general-purpose scalar instruction set).

As Figure 1 shows, the 7 Series extends the pipeline to eight stages and adds multiple execution units for superscalar operation. The first issue slot performs memory operations (load/store) and simple integer operations, whereas the second slot performs any integer operation (including multiply/divide), branch resolution, and floating-point operations. SiFive added a second fetch stage and a second data-memory-access stage to enable larger L1 cache and scratchpad memories. A second decode stage handles superscalar dispatch and leaves headroom for future optimizations that combine multiple instructions in each issue slot.

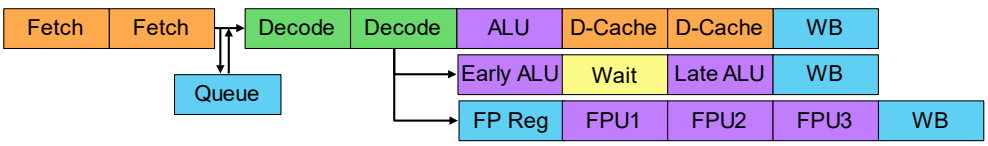

Figure 1. SiFive 7 Series pipeline. WB=writeback. Adding a second memory-access cycle in the fetch and data-cache stages enables larger TCMs without reducing clock speeds.

For More Information

RTL for the E76, S76, and U74 cores is available now. SiFive withheld 7 Series license fees. For more information, access *scs.sifive.com/core-designer/*.

Both issue slots include ALUs in stage five that handle most arithmetic instructions. Branch resolution can use the output of these ALUs immediately, resulting in a five-cycle misprediction penalty. If an ALU instruction requires the output of a pending load, however, it proceeds to stage seven, which contains a second set of ALUs. These "late" ALUs enable a zero-cycle load-to-use latency, meaning a dependent ALU instruction can issue in the cycle immediately after the instruction that loads its data. In contrast, the U5x requires a nondependent instruction between the load and use instructions. If a branch is resolved using the late ALUs, the misprediction penalty rises to seven cycles.

The U74MC has a 64-bit register file and data path, L1 instruction and data caches protected by SECDED ECC, a physical-memory-protection (PMP) unit, and a memory-management unit (MMU), allowing it to run Linux. The MMU implements the 39-bit (SV39) version of the RISC-V virtual-memory system. The PMP protects up to eight memory regions, enforcing permissions on user-mode accesses. The core can also include a core-local interrupt controller (CLIC) to enable interrupt prioritization and preemption. To thwart side-channel attacks, system software can clear branch history when switching processes.

The 32-bit E76 and 64-bit S76 are microcontroller-class CPUs that, compared with the U74, omit the MMU but add optional tightly coupled memory and a new fast-I/O (FIO) port. SiFive configures the E7x cores with an instruction cache, instruction TCM, or both; in the lattermost case, the cache excludes the TCM. For data, customers can choose a cache or TCM. The optional FIO port bypasses the core-complex bus to provide low-latency I/O access. It has a register interface and dedicated SRAM, and it adds only two cycles of load-to-use latency to the SRAM's access time. For real-time designs, such as disk controllers, customers can use instruction TCM and disable dynamic branch prediction at boot time.

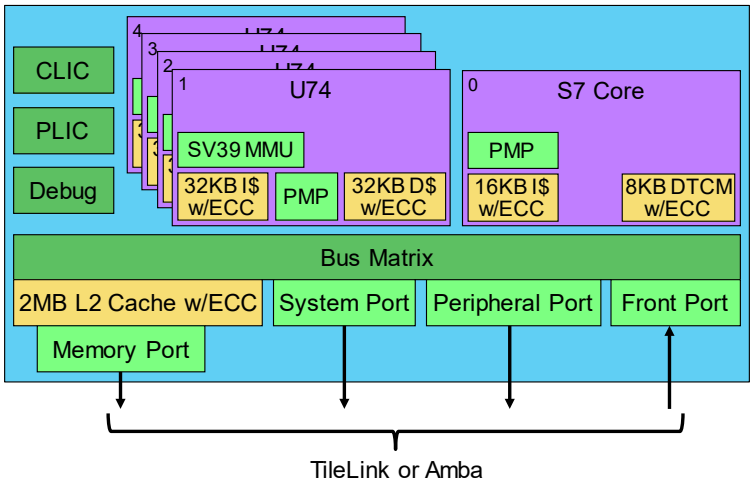

Figure 2. U74MC multicore configuration. Customers can mix and match U74, S76, and E76 CPUs in one cache-coherent core complex.

Heterogeneous-Multicore RISC-V

The U74MC has a cache-coherent bus that can connect multiple CPUs with the L2 cache and peripherals. Whereas the existing U54MC offers four U54 CPUs plus one U51, the newer U74MC handles a mix of U74 and E7x/S7x CPUs, up to a total of nine. For example, SiFive can instantiate four U74s for Linux applications plus one S76 for system management, as Figure 2 shows. An IoT processor might use one U74 for applications and an E76 as a sensor hub. In either configuration, the core complex provides a fully coherent and shared memory space, and a platform-level interrupt controller (PLIC) distributes global interrupts.

The L2 cache provides a memory port to connect an external controller. Other ports include a front port for bus masters, a system port for high-speed peripherals, and a peripheral port for low-speed peripherals. All of these ports use TileLink as their native interconnect, but SiFive configures the core with Amba (AXI/APB) bridges on customer request. TileLink is an open-source coherent interconnect with cached and noncached variants.

A simple microcontroller design can use one E76 core with TCM, FIO, and CLIC functions while omitting an L2 cache and the PLIC block. Excluding memories, such a design occupies 0.112mm^2 in TSMC's 28HPC process when using a nine-track standard-cell library. SiFive estimates this design will dissipate 20.4mW when running the Dhrystone benchmark at 400MHz (again, excluding memories). For maximum performance, a 12-track library should yield 875MHz operation in a worst-case corner, with the core occupying 0.174mm^2 and consuming 74.4mW.

To assist customers in simulating their designs, SiFive delivers a Verilog module it calls SiFive Insight. This module provides a human-readable list of all signals required to debug the RTL. The company believes Insight enables customers to debug their designs without needing to understand the core RTL.

Three Little Cores

SiFive targets embedded applications in storage, augmented/virtual reality, and wireless infrastructure. The latter two markets, however, may require vector extensions for machine-learning inference and baseband processing, respectively. Until the company upgrades its cores with a vector unit, it can't meet these DSP requirements. On the other hand, storage-controller designs can take advantage

of initial 7 Series features including 64-bit real-time CPUs, fast access to custom accelerators, and coherent shared memory between real-time and application CPUs. This last feature simplifies software design and reduces latency in controllers that employ a real-time core for the protocol stack (e.g., NVMe) and an application core for high-level features such as data deduplication.

	SiFive E76	Arm Cortex-M7	SiFive U74	Arm Cortex-A55
Instruction Set	32-bit RISC-V	32-bit Arm v7-M	64-bit RISC-V	64-bit Arm v8
Max Clock Freq	1.6GHz†	1.1GHz	1.6GHz†	1.6GHz†
Max IPC	2 IPC	2 IPC	2 IPC	2 IPC
CoreMark Perf	4.9CM/MHz	5.0CM/MHz	4.9CM/MHz	4.4CM/MHz†
Die Area*	0.065mm²	0.067mm²	0.22mm²	0.65mm²†

Table 1. SiFive-versus-Arm CPU comparison. The new dual-issue 7 Series delivers integer performance on a par with that of Arm's comparable CPUs. All metrics assume TSMC 28nm HP technology. *Without memories. (Source: vendors, except †The Linley Group estimate)

As Table 1 shows, the E76 compares with Arm's Cortex-M7 in integer performance (see *MPR 10/6/14,* "Cortex-M7 Doubles Up on DSP"). Both are dual-issue MCUs that deliver about 5.0 CoreMarks per megahertz, with the Arm design having a slight edge. Cortex-M7, however, includes DSP/SIMD extensions that the E76 omits; both vendors offer optional FPUs. According to our estimates, the E76 trails the M7 in power efficiency, but we expect it will achieve a higher clock speed.

SiFive's new S76 lacks 64-bit competitors—Cortex-R8 is similar except it's a 32-bit design (see *MPR 3/14/16,* "Arm Revs Up Cortex-R8 for 5G"). The R8 has an 11-stage dual-issue pipeline with out-of-order capability, it handles DSP instructions, and it includes a Neon SIMD unit. Its CoreMark performance, however, falls short of the S76's despite its reordering capability. On the other hand, Arm supports lockstep operation, with multiple R52 CPUs running in parallel for greater protection from runtime errors.

The U74 competes directly with Cortex-A55, Arm's newest "little" 64-bit CPU, which pairs with Cortex-A75 in the Big.Little scheme that mobile processors employ (see *MPR 6/5/17,* "Cortex-A55 Improves Memory"). The U74 and A55 both have an in-order dual-issue pipeline with eight stages, but we expect SiFive's design will achieve about 11% more CoreMarks per megahertz. The U74 should also lead in power and area efficiency. On the other hand, the A55 includes an FPU that handles Neon SIMD. Whereas the SiFive FPU completes one MAC per cycle regardless of precision, the A55 can complete two double-precision or four single-precision MACs per cycle.

The SiFive and Arm offerings differ in their multicore configurations as well. The 7 Series implements a shared L2 cache, which is available to E76 and S76 CPUs as well as U74 CPUs. By contrast, Cortex-A55 has a private L2 cache and shared L3 cluster cache, whereas Cortex-M and Cortex-R CPUs don't support private L2 caches. Furthermore, Arm customers can't build a cache-coherent multicore complex that combines Cortex-A CPUs with Cortex-M and Cortex-R CPUs. SiFive says its shared-memory architecture is important in storage controllers, but we see few other embedded designs that require this architecture.

Tuning the Open-Source Model

Although the 7 Series CPU is far from revolutionary, it delivers an impressive 63% performance boost over the existing U54. More importantly, it allows SiFive to compete against dual-issue Arm CPUs including Cortex-M7 and Cortex-A55. The U74's increased performance also expands the range of Linux applications RISC-V can serve. To compete with Arm for applications that require DSP or AI processing, however, the RISC-V Foundation must finish the vector extensions so that SiFive can add a vector unit.

The startup's revised strategy, which focuses on customer configurability for deeply embedded designs, differentiates its approach from Arm's while recognizing the immaturity of the RISC-V ecosystem. Establishing a new ISA takes time, but deeply embedded systems have relatively little software to port, making them a good starting point. RISC-V is building momentum through academics, startups, and established vendors. SiFive is leading the charge among IP vendors, establishing a new high-water mark for RISC-V performance. ♦