Deterministic Processing for Mission-Critical Applications

By Linley Gwennap
Principal Analyst

November 2020

The Linley Group

www.linleygroup.com

Deterministic Processing for Mission-Critical Applications

By Linley Gwennap, Principal Analyst, The Linley Group

Time-sensitive applications such as automotive and robotics require fast and consistent response times. Features such as caches and branch prediction hamper responsiveness. SiFive CPUs can disable these features to deliver deterministic responses. They also combine Linux and RTOS CPUs in the same cluster to enhance responsiveness while offering smaller die area than competitors. SiFive sponsored this white paper, but the opinions and analysis are those of the author.

Sometimes, every microsecond counts. When a car is hurtling down a highway at 100 feet per second, or when a robotic arm is welding a steering column every 7 seconds, control systems must respond in the same amount of time, every time. That means ensuring that the CPU provides a deterministic (consistent) response to interrupts and other requests. These mission-critical applications need the right CPU design to keep the software on track.

While determinism is a basic requirement, other factors are highly desirable. Particularly for cars and planes, critical factors include size, weight, and power (SWaP). In these vehicles, space is limited, weight requires fuel to move, and power requires either fuel or batteries, which add weight. Minimizing these interrelated factors adds room for passengers or valuable cargo and reduces operating (fuel) cost.

RISC-V CPUs from SiFive can help customers meet these requirements and minimize SWaP. Their open-source instruction set reduces licensing cost. Their customizable instruction set enables application-specific instructions that can provide tremendous gains in power efficiency (performance per watt). Above all, the CPUs can be configured for deterministic performance to support the needs of mission-critical applications.

SiFive Delivers Determinism

High-performance CPUs have several features that improve average performance but cause unpredictable deviations when executing critical code such as an interrupt service routine (ISR). For example, CPUs typically rely on fast cache memories that hold frequently used information. To simplify software and optimize performance, the CPU hardware decides which information to store in the cache. Thus, sometimes the ISR will be in cache and other times not; the latter case will require many extra cycles to load the necessary instructions and data.

At design time, SiFive U-Series CPUs can be configured to include tightly integrated memory (TIM) instead of or in addition to cache. While TIM responds as quickly as cache, software controls what is stored there. For optimal interrupt performance, software can store the ISR code in TIM. SiFive also provides a similar structure called data local storage (DLS) that can hold any data structures that the ISR requires. The TIM and DLS are typically small, perhaps 32KB each; for larger blocks of code and data, the company supports way locking in the larger level-two (L2) cache. This technique allows

software to lock part or all of the L2 cache so that critical code and data doesn't have to be loaded from slow DRAM.

In addition, most CPUs predict the direction of branch instructions using the accumulated history of prior branches; these predictions are usually correct, but when it guesses wrong, the CPU takes several extra cycles to revise its calculations. Thus, the execution time of an ISR will vary, as the accumulated branch history will differ depending on the code that was being executed when the interrupt occurred. This variance can be more than 1,000 CPU cycles, as Figure 1 shows. Disabling branch prediction eliminates this variance and, when combined with TIM and DLS, can produce a highly deterministic execution time. SiFive U-Series CPUs allow software to disable branch prediction, simplifying this task.

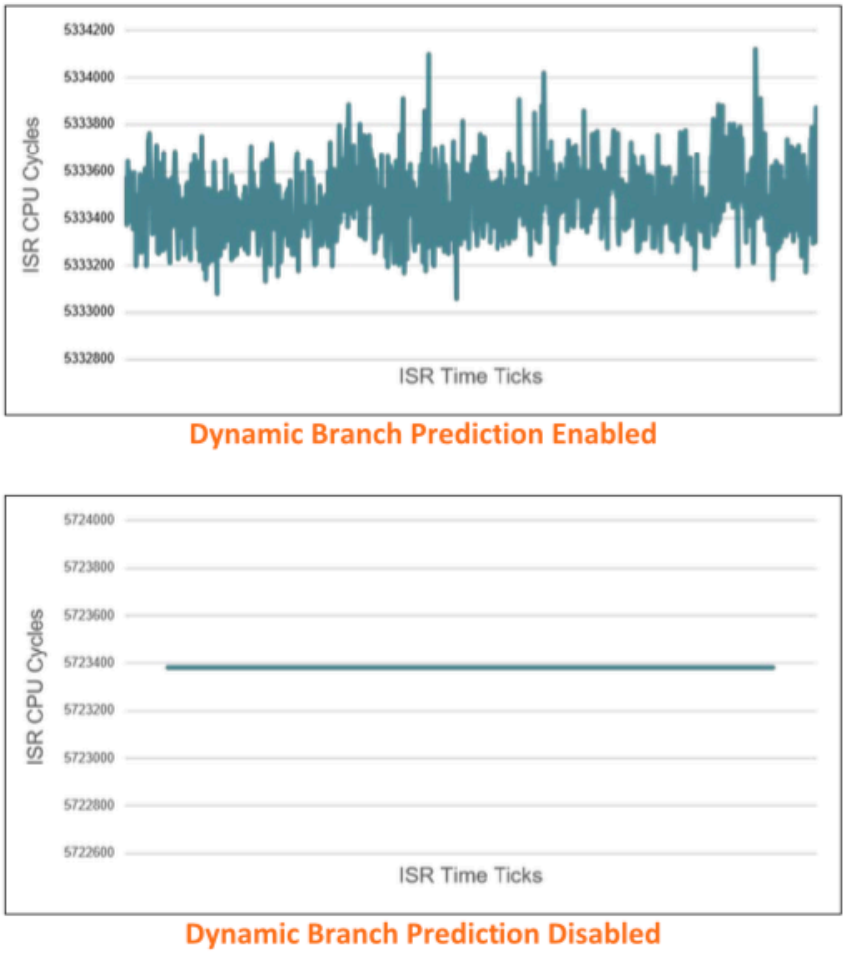

Figure 1. Deterministic interrupt processing. When a CPU employs dynamic branch prediction, execution time of the interrupt-service routine (ISR) can vary by more than 1,000 cycles, but disabling this feature enables a consistent response time. (Source: Microsemi)

Mission-critical applications also require high reliability. On-chip memory structures are subject to soft errors that can randomly flip a single bit. Although rare, such an error often causes the software to crash, something vehicle designers abhor. To avoid this problem, SiFive CPUs can be configured at design time with error-correcting codes

(ECC) on all memory structures including TIM, DLS, L1 cache, and L2 cache. The only exception is the instruction cache; since instructions are never modified, this cache uses a parity bit to detect errors and simply reloads any corrupted information. ECC adds die area but can automatically correct single-bit errors even if the data has been modified while in memory.

Another challenge for these applications is minimizing interrupt response time. A high-level operating system (HLOS) such as Linux simplifies software development, but its complexity slows interrupt processing. A real-time operating system (RTOS) responds quickly but lacks many software features of an HLOS. SiFive enables the best of both worlds by combining HLOS-capable U-Series cores and simpler S-Series cores in a single integrated subsystem that the company calls Mix+Match. To simplify data sharing, these heterogeneous cores all use the same L2 cache and the same memory map. Customers can combine different CPU types to construct a design that meets their cost and performance needs.

Meeting Customer Needs

Microsemi, now part of Microchip, has a long history of serving customers in the military-aerospace industry. In 2019, the company announced a new product combining its PolarFire FPGA technology with a quad-core processor subsystem. The new PolarFire SoC FPGA offers a SiFive Mix+Match CPU design that features four U-Series cores and one S-Series core, as Figure 2 shows. In this heterogeneous configuration, the S-Series CPU can run an RTOS for essential system functions, ensuring fast and deterministic interrupt response, while the U-Series CPUs run high-level software on Linux.

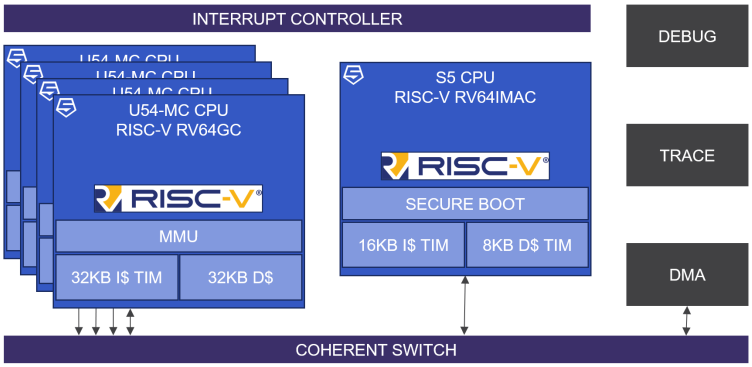

Figure 2. PolarFire SoC FPGA CPU cluster. The SoC combines an S5 CPU for real-time operation and four U54 CPUs that can run Linux and other high-level software.

Coherent Logix is a small chip vendor in Texas that has several mil-aero customers. With an eye on supporting these customers, the company chose SiFive to supply CPU cores for its next-generation HyperX network processor. Coherent evaluated several CPU suppliers but chose SiFive for its deterministic real-time processing capabilities. The chip vendor is also using SiFive Custom Instruction Extensions (SCIE) to accelerate domain-specific workloads such as multispectral imaging, software-defined radio, synthetic-aperture radio, and secure data transmission. These extensions improve performance and help Coherent optimize its solutions for SWaP.

Working with these and other customers, SiFive has gained expertise in safety standards such as DO-254. It can assist in safety audits and documentation that these standards require. To verify the reliability of its designs, the company uses extensive software test suites. In addition to the hardware-based error correction discussed above, SiFive can assist the customer in creating software-based error detection and redundancy schemes.

Outperforming Arm

SiFive's CPU designs offer several advantages over Arm's Cortex CPUs. SiFive's U74, for example, competes against Cortex-A55. Both can run an HLOS, and both have an in-order dual-issue pipeline with eight stages, but we estimate the RISC-V design achieves about 15% better performance on the CoreMark benchmark at the same clock speed, as Table 1 shows. Using an "as fast as possible" implementation (in which up to 15% of the transistors are speedy ULVT type), SiFive quotes the U74 at a typical frequency of 2.3GHz in TSMC 7nm technology, similar to Cortex-A55 under those conditions. Yet the U74 has a 24% smaller die area, which helps reduce chip manufacturing cost.

	SiFive U74	Arm Cortex-A55	Arm Cortex-R82
Instruction Set	64-bit RISC-V	64-bit Arm v8	64-bit Arm v8-R
Pipeline Stages	8 stages	8 stages	8 stages
Max Instructions/Cycle	2 IPC	2 IPC	3 IPC
CoreMarks per Clock	5.1CM/MHz	4.4CM/MHz†	5.8CM/MHz
Max Clock Frequency	2.3GHz	2.6GHz	2.5GHz
Maximum CoreMarks	11.7CM	11.4CM	14.5CM
Die Area w/Memory*	0.19mm^2	0.25mm^2†	0.31mm^2

Table 1. 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 7nm (7FF) technology. *32KB L1s and 128KB L2. (Source: vendors, except †The Linley Group estimate)

Part of the die area difference is that Cortex-A55 typically includes a Neon SIMD unit that can perform four single-precision floating-point MAC operations per cycle, versus just one for the U74. SiFive customers that require SIMD capability can instead choose a CPU such as the VIU7 that implements the RISC-V Vector (RVV) extension, or they can define their own custom instructions using SCIE, but either of these choices will add die area relative to the basic design in Table 1. Customers that don't need SIMD can remove the Neon unit, which is an optional component in Cortex-A55.

Each Cortex-A55 has a private L2 cache; in a multicore cluster, these CPUs share an L3 cache. This approach improves performance on complex applications but adds considerable die area for the L3 cache controller and memory. SiFive clusters instead share the L2 cache, simplifying the design. Although Arm supports heterogeneous clusters combining different A-Series CPUs (e.g., A75 and A55), the company doesn't allow customers to combine these CPUs with its real-time M-Series (e.g., M4) cores in the same cluster; thus, customers must determine how to connect the different clusters. SiFive enables a single heterogeneous cluster with both real-time and HLOS cores and provides an integrated debugger for this type of design.

Both cores offer similar security, trace, and debug features. Cortex-A55, however, doesn't support deterministic features such as cache locking, TIM (which Arm calls TCM), or disabling branch prediction. Instead, Arm recommends its M-Series or R-Series cores for real-time operation. For example, the new Cortex-R82 performs similarly to Cortex-A55 at 5.8 Coremarks/MHz and has a similar peak clock speed. The R82 supports separate TCM for instructions and data, ECC on all internal memories, and Neon SIMD. It can be configured to run either RTOS or HLOS software. But it has a larger die area than Cortex-A55 or the U74, as Table 1 shows.

The Right Stuff

Mission-critical applications depend on deterministic response times to respond quickly and appropriately to external events. General-purpose CPUs contain many features that optimize average performance but can impede worst-case performance. Branch prediction and caches are all about playing the averages and making things easy for the software developers, but when the hardware guesses wrong, response time can degrade unexpectedly. CPUs must be able to turn off or work around these features to create a consistent and predictable response.

Determinism isn't enough for mission-critical applications. These systems need high reliability, including error correction on all internal memories. They must meet rigorous safety standards that require documentation and testing of all components. Customers also want to optimize their designs to meet critical parameters such as size, weight, and power. As the primary determinant of performance and power, the CPU can have a significant effect on these system parameters.

SiFive's highly configurable RISC-V CPUs have the right stuff to meet these needs. Its CPUs offer TIM and DLS memories to hold critical code and data, and they allow software to disable branch prediction and lock data in cache for fully deterministic interrupt handling. The S-Series cores are well suited to an RTOS with rapid interrupt response, and customers can mix and match these cores with Linux-compatible U-Series cores. To optimize SWaP, designers can implement custom instructions to accelerate specific workloads, greatly reducing the power needed to meet a specific performance target. These cores offer similar performance and smaller die area than similar Arm Cortex CPUs. These advantages make SiFive CPUs a strong choice for mission-critical applications.

Linley Gwennap is principal analyst at The Linley Group and editor-in-chief of Microprocessor Report. *The Linley Group offers the most-comprehensive analysis of microprocessors and SoC design. We analyze not only the business strategy but also the internal technology. Our in-depth articles cover topics including embedded processors, mobile processors, server processors, AI accelerators, IoT processors, processor-IP cores, and Ethernet chips. For more information, see our website at www.linleygroup.com.*