The Linley Group

Tech Insights

MICROPROCESSOR report

Insightful Analysis of Processor Technology

P650 Pumps Up Performance by 50%

SiFive Upgrades RISC-V Line After Just Six Months

By Jim Turley (December 27, 2021)

Progress comes quickly in the microprocessor business, but even so, it's rare to see a company upgrade its flagship's performance by 50% in just six months. RISC-V vendor SiFive wants to establish its credentials as a viable Arm alternative, and not just on the basis of cost, but in performance and power efficiency. The CPU licensor is gaining ground on its British rival, but it still has a long way to go.

The new P650 core upgrades the P550 announced in July (see *MPR 7/5/21*, "P550 Doubles Floating-Point Performance") with an additional execution unit, larger caches, better branch prediction, a different load/store pipeline, hypervisor support, and a handful of other tweaks. Integer performance improves by 50%, according to the company's simulations. Most of that gain is through more instructions per clock (IPC), with the remainder due to a 10% higher maximum clock frequency.

The P650 targets two relatively high-end markets: automotive automation and data-center accelerators. RTL should reach preview customers in 1Q22, with a general release around midyear. That timing puts initial production silicon around the end of 2023 and P650-based equipment sometime in 2024.

SiFive compares its P650 to Arm's Cortex-A77, citing similar SPECint 2006 numbers. The RISC-V design should have a smaller die area, though, mostly because of its smaller caches. That feature was a primary charm of the P550, too: similar logic area but smaller caches yielding similar performance in less space, at least for some workloads. Semiconductors are in short supply, so a smaller die translates into more working chips per wafer, hence more product and more profit. Few CPU architects approach high-performance design that way, but it pays the bills.

Evolutionary Change From the P550

Given the close timing between the P550 and P650 announcements, SiFive clearly designed them in parallel, and the company acknowledges that the same engineering team worked on both concurrently. The P650 comprises a superset of P550 execution resources, including the same

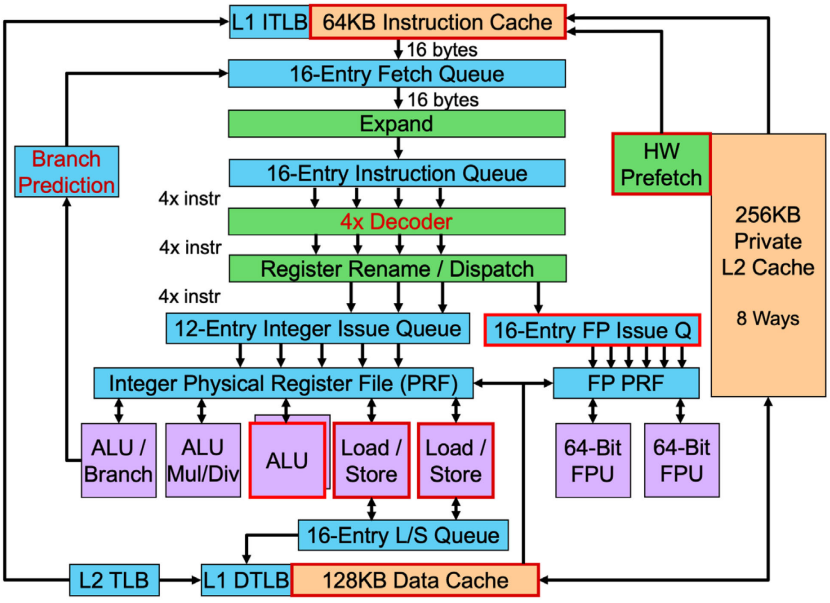

Figure 1. SiFive P650 CPU. The design includes four integer ALUs, two load/store units, two floating-point units, and larger L1 caches. Changes from the P550 are highlighted in red.

Price and Availability

Production RTL for the P650 is scheduled for general availability in 2Q22 to 3Q22. SiFive withheld pricing. For more information on the new CPU, access www.sifive.com/cores/performance-p650.

two 64-bit floating-point units, two load/store units, and three asymmetric integer ALUs. The P650 generalizes the load/store units and adds another integer ALU for a total of eight execution units to the P550's seven, as Figure 1 shows.

The integer pipeline is 10 stages long, and the load/store pipeline is 13. The P650's twin load/store units can together perform two loads, two stores, or one of each per cycle, providing more flexibility than the P550's one load plus one store.

As with its predecessor, the P650 pulls 128 bits from its instruction cache into a fetch queue and then into a 16-entry instruction queue. This decoupling helps mitigate stalls in the case of mispredicted branches. Four decoded instructions (one more than the P550) then move from the instruction queue through the instruction-decode stage, register renaming and dispatch, and then either the integer or FP issue queue, both of which are 12 entries deep. The P650 can dispatch four instructions per clock into these issue queues, up from three in the P550. SiFive says this additional issue slot and instruction decoder, combined with the extra integer ALU and the more flexible load/store units, account for the lion's share of the SPECint boost.

Smaller enhancements are in the branch-prediction logic, which the company changed in unspecified ways. The P650 also has a bigger reorder buffer (ROB), more internal (i.e., invisible to software) integer and FP registers, and better data prefetching.

Cache sizes are selectable: 32KB or 64KB for the instruction cache and a 128MB option for the L1 data cache. L2 caches range from 256KB to 2MB, and the shared L3 cache can be up to 16MB. As with the P550, L1 and L2 caches are private per core, whereas all cores in a cluster share the L3.

Coherent Core Cluster Rises to 16

Clustering has replaced massively multicore designs as the preferred route to higher performance. SiFive has now added 16-core complexes to its repertoire, with up to four groups of quad P650 cores operating in a single coherent structure. Its P550 predecessor supported eight-CPU complexes (two groups of four), so the new design should appeal to designers hoping to boost their RISC-V performance density and complexity. All cores connect via SiFive's proprietary bus. The company does, however, expose an Amba Ace–compatible interface outside of the core complex so designers can create their own coprocessors or accelerators that function as part of the coherent RISC-V cluster. Less ambitious designers can use the Amba AXI or Chi interfaces for peripherals that don't need the cache coherence or the bandwidth of the front-side bus.

Also new to the P650 CPU—and to the RISC-V ecosystem—is hypervisor support, which follows the RISC-V Foundation's approval earlier this month of the H extension. SiFive CTO Lee doubles as Chairman of the RISC-V Foundation's Technical Steering Committee, and the P650 was developed alongside of the emerging specification. Hypervisors are important both for multitenant servers, where operating systems can be isolated and containerized, and for safety-critical deployments, where application code is quarantined or sandboxed to increase reliability.

Performance Improves but Still Lags Arm

SiFive claims that its new P650 delivers 40% better IPC than P550, on the basis of SPECint2006, and that another 10% comes from the faster clock frequency enabled by a newer silicon process, for 50% greater overall integer performance. Both numbers are based on maximum-cache simulations, not silicon, as P650 chips are more than a year away. The company projects at least 11 SPECint per gigahertz, which closely matches Arm's Cortex-A77, as Table 1 shows. At a projected clock rate of 2.75GHz in TSMC 5nm silicon, the P650 at full speed should achieve more than 30 SPECint2006.

That's a fine number, but it uses an aging benchmark to compare a two-year-old CPU with a competitor that's still two years from production. SPEC2006 has a small code footprint that favors SiFive's small-cache philosophy, whereas SPEC2017 is more representative of large embedded and server workloads—ostensibly the P650's target market. The comparison also ignores clock-frequency differences. Cortex-A77 can achieve 3.0GHz in a 7nm process (see *MPR 5/27/19*, "Cortex-A77 Improves IPC"), a 10% bump despite the older technology.

SiFive has been silent about the P650's floating-point performance and its SIMD prowess. Just as it did with the P550, the company has declined to publish SPECfp numbers and to make comparisons with Arm or other CPUs. On the surface, the P650 would seem to stand toe-to-toe with the A77. Both have two 64-bit SIMD/FP units fed by a four-issue pipeline, but that similarity can conceal wide variability in FP latency and throughput. Cortex-A77 can do two 128-bit SIMD operations per cycle, double the P650's rate. The lack of performance estimates, guesses, projections, or simulations indicates the SiFve design would fare worse in the comparison.

A more recent competitor is Cortex-A710, Arm's first iteration of its latest architecture, Arm v9 (see *MPR 5/31/21*, "Arm v9 Yields Three New CPUs"). RTL for that architecture began shipping more than a year ago, and A710-based chips already appear in MediaTek's Dimensity

9000 and Qualcomm's Snapdragon 8 Gen1. That timing still puts the A710 two years ahead of the P650. Apart from being earlier, the A710 outguns the P650, boasting two dedicated branch units, an extra load/store unit, two branch predictors, an L0 cache, a larger ROB, the Scalable Vector Extension (SVE2), and other enhancements.

Cortex-A510 is the A710's little sibling and closer to the P650's feature list. It has just three decoders and three integer ALUs but equivalent load/store capability. On the minus side, the A510 is an in-order machine, but it includes SVE2. Designers can group A510s in an 8-CPU cluster, either standalone or as the little helpers to a gang of A710s; the P650, on the other hand, has 16-CPU capability. The A510 also offers a merged dual-core design that shares some hardware resources to save area and power (at the cost of performance)—an option the P650 omits. Like the A710, the A510's announcement came last summer, with chips already here.

SiFive is withholding the P650's die area, and it offers no power estimates. Nevertheless, the company hopes its new CPU will, like the P550, provide integer performance similar to Cortex-A77 in the same process node and clock frequency but with less power and die area. If that expectation holds true, the P650 could be a viable alternative for designers who want solid integer capability but don't need the greatest FP or SIMD performance, and who don't require Arm software compatibility.

Cars and Servers: Two Varied Markets

As the new flagship CPU atop SiFive's "Performance" brand, the P650 naturally targets servers. But the company doesn't see it competing directly with data-center application processors, such as those based on Cortex-X2, and it isn't targeting x86 processors; SiFive is ambitious but not delusional. Intel and AMD dominate the server space for good reason. Even Arm, after more than 10 years of direct frontal assault along with multiple creative and well-funded licensees, holds barely 1% of that market. RISC-V processors have no chance of displacing either of those alternatives. At least, not as the main application processor.

Instead, SiFive sees its opening in the networking- and storage-accelerator part of the server rack. Chips based on the P650 need merely play nice with Xeon and Epyc rather than trying to replace them. RISC-V's low cost relative to the incumbents, combined with its instruction-set flexibility, makes it a solid platform for special-purpose computing. And its small software base is less of a hindrance when the application is proprietary.

Like the P550, the P650 supports Linux, so the canonical LAMP software stack (Linux, Apache, MySQL, and PHP) is a short hop away. But SiFive has discovered that most of its server customers develop their own software rather than running third-party code. Thus, they're more willing to cross-compile their code, making the transition to RISC-V from x86 or Arm more tolerable.

Software independence is even more relevant in the automotive market, where SiFive has some traction. Automotive-SoC designers frequently write their own code, so they're largely CPU agnostic. The P650's new hypervisor support, as well as its ability to serve in 16-CPU clusters, could make it popular with ADAS developers that aren't already using Mobileye chips. The RISC-V design's presumably smaller die area (compared with Cortex-A77, at least) will increase yield, in turn decreasing cost.

RISC-V Offers the Alternative Alternative

Unquestionably, RISC-V's popularity has boomed in a short time, becoming the de facto alternative to Arm and x86—a position MIPS, SPARC, and others once held. Arm has hordes of licensees working in its favor; x86 has decades of accumulated software and performance superiority. But both are expensive and inflexible. Designers who want to employ a cheaper alternative or create custom ISA extensions must look elsewhere. RISC-V has proven to be a reliable foundation on which to differentiate silicon without resorting to the moonshot expense of in-house CPU development.

For many, SiFive CPUs are good enough in that they are in the same performance league as previous-generation Arm CPUs, but with lower upfront licensing fees and cheaper royalties. Naturally, SiFive executives bristle at the suggestion that they're simply peddling knockoff processors. The P550 is similar to Cortex-A75 at a little over half the die area, and the P650 promises to offer the same benefits. Perhaps channeling *Top Gear's* Jeremy Clarkson, SiFive CTO Yunsup Lee said, "Some say…RISC-V is all about reduced cost, but it's not. There's actual value that

	SiFive P550	SiFive P650	Arm Cortex-A77	Arm Cortex-A710
Instruction Set	RISC-V	RISC-V	Arm v8	Arm v9
Architecture Size	64 bits	64 bits	64 bits	64 bits
Pipeline Length	13 stages	13 stages	12 stages	12 stages
Max Decode	3 instr	4 instr	3 instr	4 instr
Load/Store Units	2x load/store	2x load/store	2x load/store	3x load/store
FP/SIMD Units	2x 64-bit	2x 64-bit	2x 128-bit	2x 128-bit
Reorder Window	64 instr	64 instr	~128 instr	~128 instr
L1 Cache Size	32KB/32KB	128KB/128KB	64KB/64KB	64KB/64KB
Max Clock Speed*	2.6GHz	2.6GHz†	3.0GHz	3.0GHz
SPECint2006†	23	30	34	40
Die Area*	0.38mm^2	Undisclosed	1.52mm^2†	1.77mm^2†
RTL Availability	3Q21	3Q22	2Q19	3Q21

Table 1. RISC-V versus Arm CPUs. The new P650 improves in several ways relative to its P550 predecessor, but it still trails Cortex-A77 and Cortex-A710. *In 7nm TSMC technology with L2 cache. (Source: vendors, except †The Linley Group estimate)

we bring, in performance per watt and performance per square millimeter."

The P650 adds "big-iron" hardware features such as hypervisor support, 16-way clustering, an accessible coherent interface for accelerators, and an integer-performance boost to what was already a capable CPU alternative. SiFive has admittedly lagged Arm technology by about 4 years, but it now believes that gap has shrunk to 2–3 years. At that rate, we may eventually see a real two-horse race for server-adjacent processors. ♦