Reed Electronics Group

MICROPROCESSOR REPORT

www.MPRonline.com

THE INSIDER'S GUIDE TO MICROPROCESSOR HARDWARE

AMCC's Titan Core

New Power Architecture Core Uses Only 2.5W at 2.0GHz

By Tom R. Halfhill {7/23/07-01}

AMCC and Intrinsity are two companies looking for more respect. Not that they're disrespected, but they're not exactly household names, either. If their joint project succeeds, both companies will become stronger contenders in their respective markets.

Three years ago, AMCC surprised everyone by acquiring IBM's PowerPC 4xx-series standard-products line for $227 million. Included in the deal were royalty-free licenses for IBM's Power 4xx-series processor cores and an IBM design team. (See the sidebar, "AMCC Strikes a Big Deal for PowerPC" in *MPR 4/26/04-02*, "IBM Loosens Up CPU Licensing.") Overnight, AMCC became an important supplier of networking and communications processors. Soon, AMCC began introducing new chips designed around IBM's Power 405 and Power 440 processor cores. Since 2004, AMCC has grown a $55 million-a-year product line into a business that annually generates $100 million. However, AMCC still isn't as widely known as larger competitors like Broadcom and Freescale Semiconductor.

Meanwhile, Intrinsity was struggling to sell MIPS-compatible embedded processors built with Intrinsity's proprietary Fast14 logic. (See *MPR 5/27/03-03*, "Update on Intrinsity Fast Products.") When that strategy stalled, Intrinsity abandoned its fabless-semiconductor model and began offering Fast14 technology to other companies as licensable intellectual property. (See *MPR 1/10/05-02*, "Intrinsity Takes Its IP on the Road.") In addition, Intrinsity began refining its Fast14 development tools and offering custom architectural and design services. Altering the company's direction was difficult. At times, Intrinsity seemed to be hanging by a thread.

Now AMCC and Intrinsity have joined forces to create an entirely new Power Architecture processor core. Code-named Titan, the 32-bit semicustom core relies heavily on Intrinsity's Fast14 logic to reach high clock speeds (up to 2.0GHz in 90nm bulk CMOS) while consuming remarkably little power (2.5W). In addition, Titan is part of a dual-core "processor complex" that supports coherent multiprocessing. AMCC can replicate this dual-core complex to create quad-core processors and perhaps even larger multicore designs. AMCC and Intrinsity unveiled Titan at Microprocessor Forum in May, and AMCC plans to announce the first Titan-based SoCs this fall.

If Titan succeeds, it will admit AMCC and Intrinsity to an exclusive club formerly limited to Freescale, IBM, and P.A. Semi—the only other companies creating original Power Architecture designs. Titan may also win new respect for Fast14 technology, which can accelerate ARM- and MIPS-compatible processor cores, too. Indeed, Titan bears so much weight on its shoulders that perhaps AMCC should have named it for an ancient Titan god: Atlas.

High Performance in Bulk CMOS

About two years ago, AMCC reluctantly concluded that IBM's existing Power 405 and Power 440 cores lacked the processing power and multiprocessing capabilities required to fulfill AMCC's future plans. Both IBM cores date to the 1990s. Although IBM introduced the new Power 460S and Power 464 cores last year, they are much like the Power 440. (See the sidebar, "IBM's New Licensable Power Cores" in *MPR 11/27/06-01*, "Power.org's United Roadmap.") AMCC

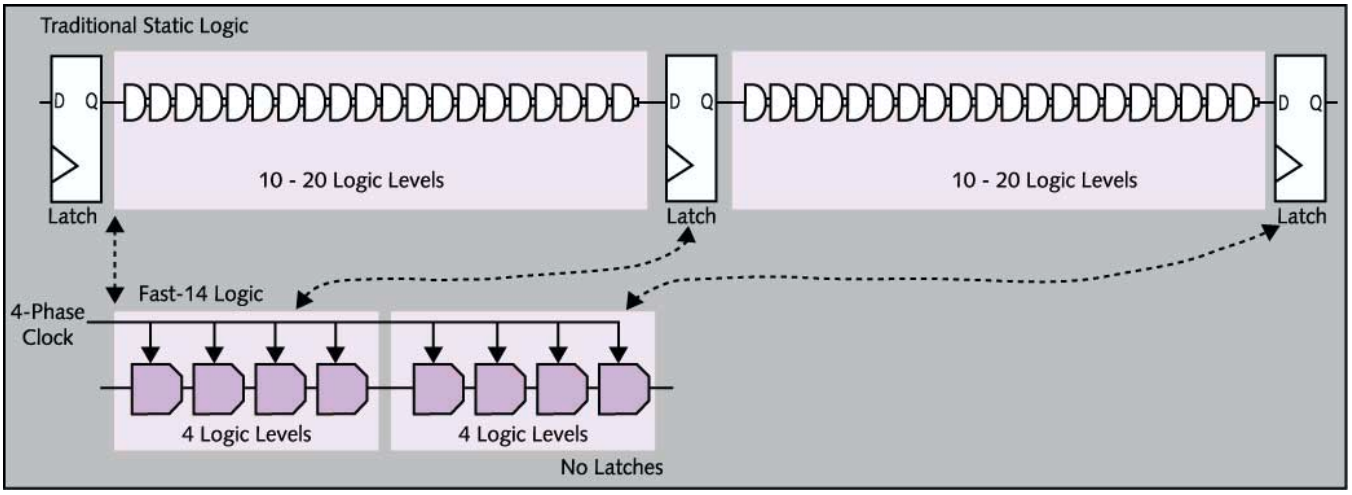

Figure 1. Titan's high-performance design is predicated almost entirely on the efficiency of Intrinsity's Fast14 1-of-N domino (NDL) logic. This latch-free logic uses four-phase overlapped clocking and optimized transistors to reach higher clock speeds while conserving power. As a result, Titan has a much shorter pipeline and significantly fewer transistors than other processors with similar features and performance.

will continue using IBM's 405, 440, and 460 cores in some products, but for its highest-performance devices, AMCC decided to create a new Power Architecture core from scratch. Although this undertaking was more expensive and risky than using IBM's ready-made cores, it was deemed necessary to keep AMCC's top-shelf products competitive.

AMCC says its customers are demanding more general-purpose performance, higher chip-level integration, and lower power consumption. Large networking customers say their software is deeply inspecting more than 90% of the data packets passing through their routers, instead of merely glancing at the headers and forwarding the packets to their destination. Among other things, deep inspection is required to scan for malware and assign priorities to packets. AMCC also perceives more demand for intelligent storage processors, wireless-infrastructure SoCs, and controllers in color laser printers. When AMCC announces the first Titan-based SoCs, they will probably be members of AMCC's G-series family for control- and data-plane network processing.

One way to deliver more general-purpose performance is to deepen the processor's pipelines and crank up the clock frequency. But deeper pipelining would increase the complexity of the core and inflate power consumption. To hold the line on power, AMCC investigated using some advanced fabrication technologies. However, those technologies are more expensive than bulk CMOS and would make the design project more difficult. (AMCC, unlike IBM and Freescale, is a fabless semiconductor company.) Instead, AMCC sought a solution that would deliver high performance and low power in the economical bulk CMOS processes available from independent foundries like TSMC.

That search led AMCC to Intrinsity. Both companies have design teams in Austin, Texas, and some of their employees know each other. After giving up its own ambitions to make chips, Intrinsity was shopping around its Fast14 technology, improved design-automation tools, and architectural design services. Intrinsity said it could help AMCC design a new Power core that would reach multigigahertz clock speeds with a short pipeline, low active power, and low leakage. Furthermore, the new core could be inexpensively fabricated in bulk CMOS at a proven process node (90nm). All it required from AMCC was a leap of faith.

Fast14 Boosts Performance, Cuts Power

AMCC made the leap. Titan's eight-stage simple-integer pipeline is a model of brevity—some other high-performance processors have pipelines two or three times as long. Yet Titan's maximum target clock frequency is 2.0GHz for a device manufactured in 90nm bulk CMOS (TSMC's 90-GT process). Titan is capable of reaching even higher frequencies, but AMCC is restraining the core by reducing its voltage to 1.0V. (The nominal Vdd in TSMC's 90-GT process is 1.2V.) At 1.0V, power falls to a mere 2.5W under typical conditions.

Excluding register files and caches, Titan has only 1.6 million transistors—positively tiny by modern standards. Less logic means less leakage, especially in deep-submicron fabrication. Titan's small size is even more impressive for a 32-bit Power processor that boasts two-way superscalar execution, instruction reordering, dynamic branch prediction, hardware support for coherent symmetric multiprocessing (SMP), a full-fledged memory-management unit (MMU), and a double-precision FPU. Clearly, this core isn't a stripped-down econobox.

AMCC and Intrinsity attribute Titan's magic to Fast14. Instead of traditional static logic, Fast14 uses 1-of-N domino logic with four-phase overlapped clocking and a preponderance of NMOS transistors. Intrinsity's proprietary design tools automatically optimize the size of each transistor for its load current and required switching speed. The overlapped clock phases simplify timing closure. TSMC will manufacture

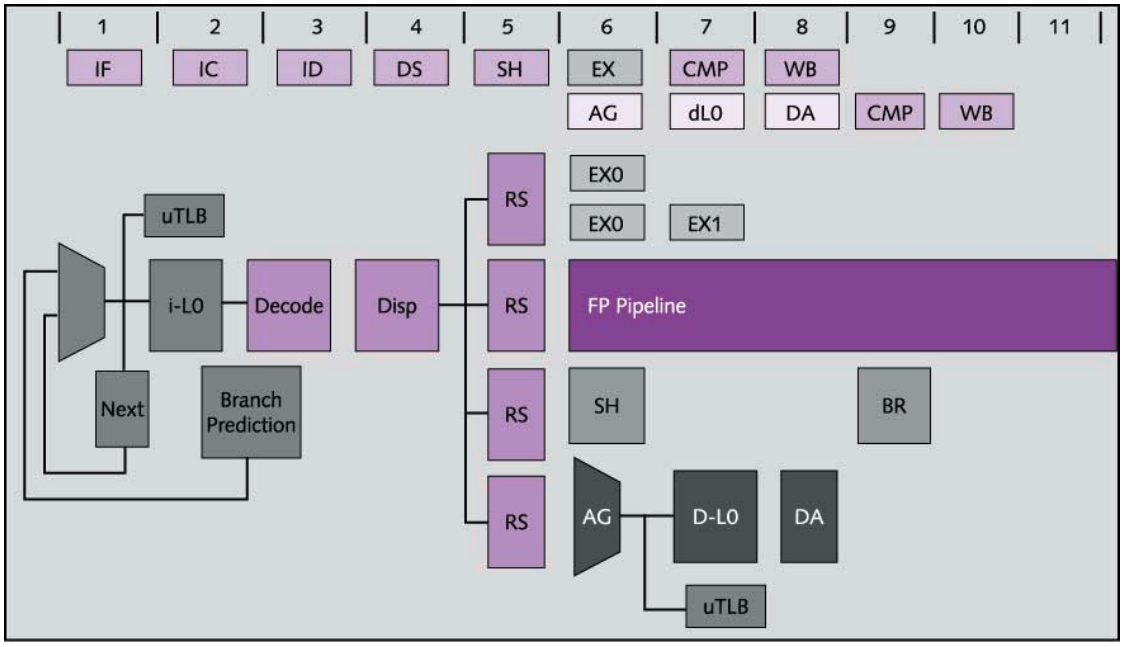

Figure 2. Titan pipeline diagram. Each of the four reservation stations preceding the execute stages can dispatch one instruction per clock cycle. Although the processor can execute only two instructions per cycle, the four stations ensure that each pipeline receives a steady flow of instructions. Note the gap between the last two stages in the branch-instruction pipeline—it allows clock gating to shut off power to the branch unit when it's idle. Titan has extensive clock gating, even to the degree of shutting off the ALU's shifter and Boolean-logic circuits when they're not needed.

the first Titan chips at 90nm, but Intrinsity says the design will scale to future 65nm, 45nm, and 32nm processes.

As Figure 1 shows, Fast14 circuits need no latches in the dynamic-logic paths, and they have only four levels of logic in places where a traditional static circuit would have 10 to 20 levels. Fast14 is so efficient that Titan's single-cycle ALU has only four gate delays in the execution path—and that includes a multiplexer for bypassing instructions. Intrinsity says Fast14 is two to three times faster than synthesized static logic. By lowering the core voltage, AMCC is trading some of that speed advantage for lower power consumption. (For a detailed explanation of Fast14, see *MPR 8/13/01-02*, "Intrinsity's Dynamic Designs.")

Pipelines Are Short But Fast

Like IBM's Power 440, 460S, and 464 processor cores, Titan is a two-way superscalar machine. It can simultaneously dispatch any two instructions after decoding, and it can issue up to four instructions from its reservation stations on every clock cycle. Each station is four entries deep. To maximize utilization of the issue slots, Titan executes integer instructions out of order and dynamically predicts branches.

Branch-prediction resources include a return-address stack and a 2K-entry branch-history table that obeys a two-bit G-share algorithm (strongly taken, weakly taken, weakly not taken, strongly not taken). The processor doesn't discover mispredicted branches until late in the pipeline, incurring a seven-stage penalty. Although that penalty is rather severe for a processor with an eight-stage pipe, it's not bad compared with other high-performance processors, because they typically have deeper pipelines.

Simple integer instructions pass through eight stages. Complex integer instructions require an additional execution stage. Loads and stores require ten stages. The FPU executes instructions in order, but integer instructions can bypass floating-point instructions and vice versa. Figure 2 illustrates the various pipelines.

Titan's cache subsystem is a little unorthodox. For the most part, Intrinsity is using TSMC's standard SRAM cells, which are denser but slower than custom bit cells. However, Intrinsity added some custom cells for critical elements, such as the word-line drivers and sense amps. Consequently, accessing the L1 cache has a latency of three clock cycles (1.5ns at 2.0GHz). That latency includes sequential accesses to the tag RAM (stored in content-addressable memory) and data movements into or out of the L1 data RAM. These sequential accesses take a little longer than parallel accesses would, but they reduce peak power by limiting the draw on the RAM arrays. To minimize the performance impact of the three-cycle L1 cache latency, Intrinsity took the unusual step of adding L0 instruction and data caches.

That's right—while server processors are gaining L3 and even L4 caches, Titan is extending the cache hierarchy in the opposite direction. Normally, a processor's register file is effectively an "L0" cache. Of course, Titan has all the standard registers defined by the Power Architecture. But Titan also has separate 4KB instruction and data caches between the processor core and the 32KB L1 instruction and data

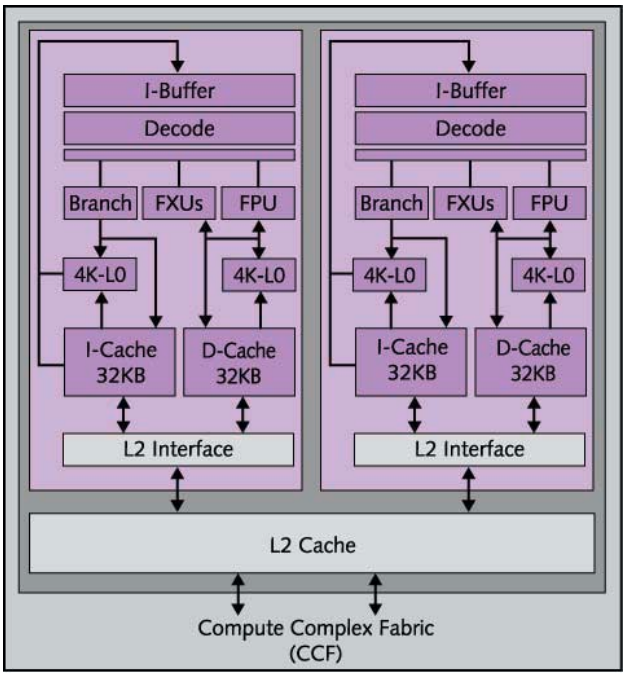

Figure 3. AMCC Titan block diagram. The default configuration joins two identical Titan cores to a shared L2 cache and high-speed on-chip interconnect fabric. Single- and quad-core configurations are also possible. Titan supports cache-coherent SMP in hardware.

caches. The processor can access these small L0 caches in one clock cycle (500 picoseconds at 2.0GHz). Interposing these small, fast caches between the processor core and L1 caches helps to accelerate tight loops, because the processor can access the L0 caches in one cycle instead of waiting three cycles for the L1 caches.

To further improve cache performance, the L1 caches are 64-way set-associative, and they support two instruction-fetch misses and four data-load misses. The L1 caches also support line locking, parity checking, and the MESI coherency protocol. Titan automatically manages the cache subsystem, so despite the unusual hierarchy, caching remains transparent to software.

Loads and stores that miss the L0 and L1 caches ripple through to the on-chip L2 cache. In first-generation Titan devices, the L2 cache can be as large as 2MB. The L2 cache interface is 128 bits wide and runs at half the core speed, providing 16GB/s of bandwidth to a 2.0GHz Titan core. L2 latency is seven cycles. The L2 cache is eight-way set-associative and, like the L1 cache, supports line locking and MESI. In addition, the L2 cache has ECC protection—vital for critical applications, such as storage controllers. It also supports snooping, which saves the processor from needlessly accessing main memory after a recent operation stores results in the cache. Under software control, programmers can define part of the L2 cache as globally accessible memory, creating an on-chip scratchpad.

Hardware Support for Multiprocessing

As mentioned earlier, Titan is designed to function as a dual-core "processor complex." In this configuration, shown in Figure 3, two identical copies of the core share the L2 cache and support coherent SMP. Although this configuration is the default, Titan is suitable for single-core implementations, and the processor complex can be doubled for a quad-core implementation. We expect AMCC to introduce devices in all these configurations eventually, as well as larger multicore designs. (AMCC anticipates a ten-year lifespan for Titan.)

Cache coherency is managed in hardware. The L1 caches in each core are aware of each other, and Titan enforces a non-inclusive store policy on the L1 and L2 caches. As part of this policy, the L2 controller filters out most of the redundant snoop traffic, improving cache utilization. Two new instructions (TLBIE and TLBSYNC) synchronize page-table entries among the cores. These instructions comply with the Power 2.04 instruction-set architecture (ISA) specification.

Two things notably missing from Titan are a 64-bit ISA and AltiVec extensions for vector math. AMCC says it omitted both features after careful deliberation. For AMCC's target applications—mostly networking and communications—a 64-bit architecture was deemed unnecessary, because the additional complexity might inflate power consumption beyond the incremental increase in performance. AMCC doesn't think memory capacity will be an issue, because Titan has 36-bit physical addressing. Of course, some of AMCC's competitors have reached different conclusions. For instance, Broadcom and Cavium Networks use 64-bit MIPS-compatible cores in their high-performance networking and communications processors. On the other hand, all of IBM's licensable Power cores are 32 bits, and all of Freescale's networking and communications chips have 32-bit Power cores, too.

AltiVec extensions are primarily for media processing, although they are useful for any applications performing vector math on integer datatypes. They are heavyweight extensions, because they add 32 128-bit registers in addition to 162 instructions. (See *MPR 5/11/98-01*, "AltiVec Vectorizes PowerPC.") By omitting AltiVec, AMCC is telegraphing its intention to remain fairly narrowly focused on networking and communications. AltiVec might be useful to some AMCC customers—printer vendors come to mind—but AMCC didn't want to burden its other customers with extensions they might never use.

In view of these considerations, it's a little surprising that Titan has a double-precision FPU. And the FPU is a standard feature, not an optional coprocessor, as with IBM's Power 405, 440, and 460 cores. Yet an FPU is redundant for packet processing. Perhaps in this case, AMCC is throwing a bone to its printer customers.

Chip Announcements Coming Soon

AMCC plans to announce the first Titan-based chips this fall. At that time, the company will disclose more information

Price & Availability

AMCC's Titan processor core, designed by Intrinsity, will debut in some AMCC networking processors, scheduled to be announced this fall. Titan is not licensable as intellectual property.

For more information visit:

- *http://investor.amcc.com/releasedetail.cfm?ReleaseID=244596*

about the on-chip interconnect fabric that will link Titan cores to external memory controllers and other integrated components. IBM's royalty-free CoreConnect technology would seem to be the obvious choice for a Power Architecture core that's similar to IBM's Power cores. However, AMCC is hinting that something different, or at least something improved, is in the works. One consideration is memory coherency for multicore SMP. AMCC says the on-chip interconnect will run at 400–600MHz, providing 12.8–19.2GB/s of bandwidth to memory.

By creating a semicustom processor core using Fast14 logic, AMCC and Intrinsity are giving Titan an inherent advantage over fully synthesizable processor cores. AMCC has no intention of licensing Titan to other companies, so the designers didn't have to make compromises for the sake of portability. Titan is more like Freescale's custom E-series Power cores, Cavium's custom cnMIPS64 cores, or P.A. Semi's custom PA6T core. (See *MPR 10/25/05-01*, "P.A. Semi: New Blood for Power.")

By lowering the core voltage below nominal, AMCC is deliberately sacrificing some of Fast14's clock-frequency headroom for reduced power consumption. According to AMCC's estimates, a dual-core Titan design will use only 5W for the processors, about 4.5W for system logic, about 4.5W for I/O controllers, and about 1W for other things. Total power for a highly integrated dual-core chip would be 15W—yet it could hit 2.0GHz, delivering up to 8,000 Dhrystone mips (4,000 mips per core). Cavium recently announced a dual-core Octeon SSP storage processor that consumes only 8W, but it must run at 600MHz to achieve that power level. (See *MPR 7/16/07-01*, "Cavium Stalks Storage.")

Closer comparisons must wait until AMCC unveils the first Titan-based chips in a few months. Integration is as important as raw processor performance, because a device that needs several auxiliary chips to perform its duties merely shifts the power burden elsewhere on the board.

Titan is an important step for AMCC. The investment required to design an entirely new Power processor core is significant, and AMCC is gambling the high end of its growing product line on Titan's success. Adopting Intrinsity's Fast14 technology is another gamble, though perhaps less risky than it seems. Intrinsity proved Fast14's worthiness by introducing some high-performance MIPS-compatible processors in years past. Mainly, it was Intrinsity's underfunded fabless-semi business model that failed. If Titan succeeds, it will further validate Intrinsity's new strategy and make AMCC a more serious contender.