VOLUME 13, NUMBER 6
MAY 10, 1999

MICROPROCESSOR REPORT

THE INSIDERS' GUIDE TO MICROPROCESSOR HARDWARE

StrongArm Speed to Triple

Long Arm of Intel Stretches Pipeline, Raises Clock to 600 MHz

by Jim Turley

EMBEDDED PROCESSOR FORUM

It's been three years since StrongArm debuted, and it'll be another year before the next generation is ready, but the time seems to have been well spent. Intel's former i960 design team in Arizona has recast the one-time Digital Semiconductor protégé as an Intel part *par excellence*. The second-generation StrongArm processor (Intel avoids calling it StrongArm-2) should reach 600 MHz by early 2000 while still staying below the self-imposed 500-mW barrier. Like the original StrongArm-110, "SA-2" should lead the industry in MIPS/watt.

At last week's Embedded Processor Forum, Intel design manager Jay Heeb lifted the veil on SA-2, revealing the alluring outline of a new pipeline, a process shift, and some design tradeoffs that result in performance degradations on a per-clock basis but increase frequency drastically. He also reiterated Intel's plan to produce both standalone microprocessors (à la the SA-110) and integrated devices (like the recently announced SA-1110; see MPR 4/19/99, p. 15).

Scalar Pipeline Still Short by Intel Standards

Succinctly, Intel's three major goals were to preserve StrongArm's lead in uniscalar performance (SA-2 is still not superscalar), keep power consumption below 500 mW, and make SA-2 manufacturable on standard Intel processes. To maintain a speed lead, Intel had four years of catching up to do.

The usual course of action in such cases is to lengthen the pipeline, and this Intel did. Early ARM chips have a constipated three-stage pipeline with a heavily burdened final stage. The first StrongArm (see MPR 11/13/95, p. 16) opened up the pipeline to five stages, rectifying some congestion. Rather than disembowel this arrangement completely, Intel simply eliminated a few of the remaining pinch points.

ARM's charming (not to say peculiar) inline operand shift is an obvious critical path for anyone contemplating the ARM execution pipeline. All ARM chips must be prepared, in essence, to execute two separate operations in sequence: a shift followed by any ALU operation. These two operations must be serialized; they can't be done in parallel for the same instruction, due to the data dependency. Apart from this particular aspect, ARM has a pretty normal architecture.

To relieve this bottleneck, Intel's designers spread the shift-and-ALU compound operation across one and one-half pipeline stages. As Figure 1 shows, the inline shift is now executed in the back half of the previous clock cycle, after the register access. Thus, normal ALU operations get a full clock cycle to execute, as they would with any normal RISC design.

Pushing the inline shift back up the pipeline caused its own problems, however, crowding the register-access stage, which usually gets its own cycle. Therefore, it too was spread over one and one-half cycles.

Finally, the instruction fetch from cache was extended by a half-cycle, sharing the early half-cycle with a new branch-target buffer (BTB) access. Data cache accesses were similarly spread over one and one-half cycles. The extended cache access hampers performance, as SA-2 now incurs a two-cycle load/use penalty, one cycle longer than what most ARM chips suffer. (A peculiarity of the

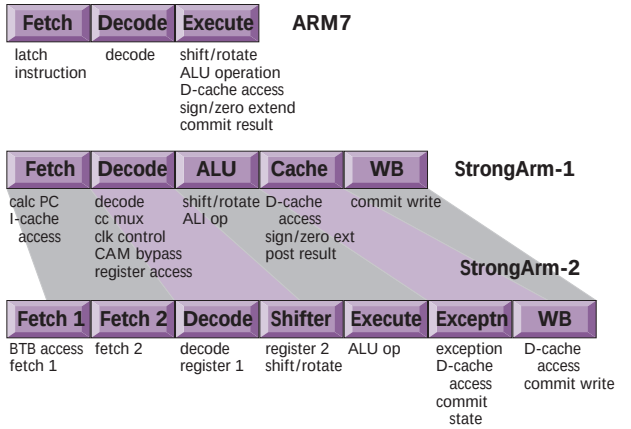

Figure 1. Intel's StrongArm-2 opens up the previous StrongArm pipeline from five stages to seven, with cache accesses, register accesses, and execution now taking 1.5 cycles each.

original StrongArm-1 causes a second delay cycle for sign-extended data.) The extra cache-access phase is needed to reach the higher clock rates, but it increases the SA-2's average clocks per instruction (CPI) by about 5%, according to Intel. Assuming that this alteration allows the company to increase the frequency of the pipeline by much more than 5%—which it does—the change is a good one overall.

In some sense, SA-2 has a ten-stage, not a seven-stage, pipeline; some pipe stages are just a half-cycle in duration. The first, third, and fourth stages each perform two sequential and unrelated functions, one on the rising edge and one on the falling edge of the clock. It's likely that Intel will eventually draw these out into full-cycle pipeline stages in its next overhaul of the microarchitecture. Perhaps StrongArm-3, which is already being developed in Austin, will have this ten-stage pipeline.

The quest for faster clock frequencies claimed a second victim as well. Before SA-2, where the inline shift and the normal ALU operation were in the same clock cycle, operands were forwarded from the output of the ALU back to the input of the ALU in a single cycle, avoiding a pipeline stall when the result of one operation was dependent on the result of the immediately previous operation. Now, with these two functions handled in different clock cycles, Intel had to decide whether results should be forwarded to the shifter, causing a one-cycle delay for dependent operations, or to the ALU, avoiding the penalty but compromising compatibility.

MICHAEL MUSTACCHI

Jay Heeb of Intel describes the internal changes to the next-generation StrongArm.

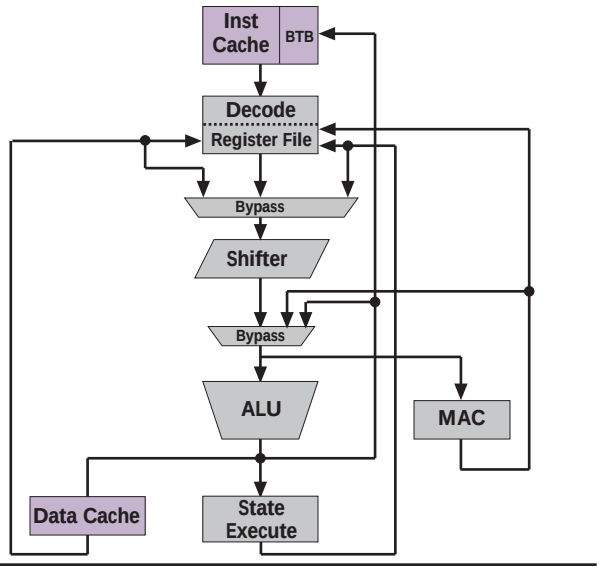

Figure 2. Because of its longer pipeline, SA-2 uses different bypassing for operand forwarding. ALU results that are not needed for a subsequent shift operation are immediately available; in-line shifts dependent on previous data now incur an extra delay cycle.

In the end, the design team did both. As Figure 2 shows, results from the ALU output are routed back to the ALU input and also to the multiplexer that feeds the input of the shifter stage. ALU operations that depend on the instruction immediately preceding them execute with no trouble. However, inline shift operations that depend on the result of the previous instruction cause a one-cycle bubble in the pipeline while the shifter waits for its input. Again, Heeb said, the overall hit was less than 5% in CPI, in exchange for a 50% gain in clock frequency.

Branch Prediction for StrongArm

Longer pipelines have their usual failings, and Intel, alas, is not immune to them. Specifically, long pipes incur long delays when branches are mispredicted. ARM—unlike MIPS, PowerPC, or the i960—has never had either static (compiler-implied) or dynamic (hardware-inferred) branch prediction. Until now.

SA-2 alleviates some of its long-pipe woes with a combination of dynamic branch prediction and a branch-target buffer (BTB). The 128-entry BTB is simply a direct-mapped cache that holds the target address of change-of-flow instructions. If the current instruction hits in the BTB and the dynamic predictor predicts that the branch will be taken, the resulting target address is sent to the fetch logic.

The SA-2 uses the familiar two-bit saturating-counter prediction algorithm, but with a twist. Normally, such an algorithm shifts only between adjacent states (from weakly taken to strongly taken, for example), as Figure 3 shows. A correct prediction advances one state; an incorrect prediction weakens it by one state. A common enough algorithm, but Intel swerved to avoid the pedestrian and took a different path.

When SA-2 mispredicts a branch, it forces the prediction all the way to the strongly opposite state, rather than passing through the intermediate state first. Thus, if the predicted state was "weakly taken," and the prediction proves false, the state immediately changes to the opposite extreme, "strongly not taken."

This little tweak yielded up less than a 1% improvement in the performance of simulated benchmarks. The difference may be nugatory, but it seems to have caught the design team's fancy. Branch accuracy improves by just 2–3%, and that only on "a few concocted cases," according to Heeb. He would not speculate on whether the modified algorithm would be applied to other Intel processors, or, indeed, if it would even be applicable.

Intel's observed improvement in branch-prediction accuracy may have been just an artifact of the specific tests it chose to simulate. It could also be peculiar to the ARM

instruction set or compilers and not transferrable to other architectures. ARM supports predicated execution for all instructions, so short forward branches typically don't appear in optimized ARM code. That leaves mostly long branches and loop terminators, which behave differently.

New Instructions Remain Mysterious

There has been persistent speculation about how Intel might change StrongArm's instruction set. The company's ARM license grants it far more freedom than most ARM licensees enjoy, obviously including microarchitectural changes but also extending to instruction-set enhancements. During his presentation, Heeb betrayed no clues about any such ISA alterations.

Nevertheless, we suspect that Intel has extended the instruction set of SA-2 in various ways. The most obvious (and fashionable) choice would be to add pseudo-DSP instructions. ARM already has a MAC instruction, which is all many low-end DSP algorithms need. More ambitious functions, such as those in AltiVec (see MPR 5/11/98, p. 1) or MDMX (see MPR 10/28/96, p. 17), would go a long way toward moving SA-2 firmly into the ranks of CPU/DSP hybrid processors. Integer SIMD instructions, such as those in MMX or VIS, would be a good start, and bit- and byte-swapping instructions, such as those in AltiVec, would make SA-2 far more useful in networking, territory that is clearly drawn on Intel's map for expansion (see MPR 5/10/99, p. 5).

Whatever the specifics, Intel seems hesitant to divulge the extent (or indeed, the existence) of StrongArm's DSP extensions until sometime later this year.

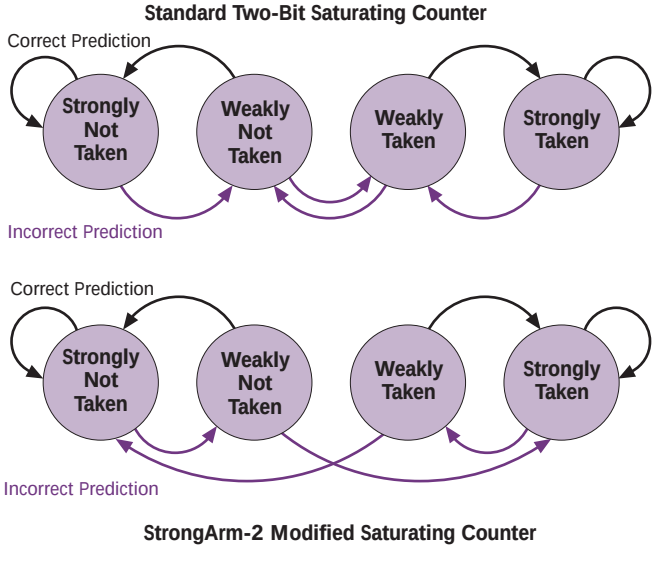

Figure 3. StrongArm-2 modifies the normal two-bit saturating prediction state (shown at top) by saturating to the strongly opposite state after two incorrect predictions.

Intel Tweaks 0.18-Micron P858 Process for SA-2

Some have argued that Digital's real accomplishment with the SA-110 was not in the microarchitecture of the chip but in the company's exotic semiconductor fabrication process. Any chip built on the fabled Hudson line, it was argued, couldn't help but run fast.

With Digital's (now Intel's) 0.35-micron fab well into middle age, it's time to move ahead. More important, the Hudson fab is not compatible with Intel's others—so much of the point of SA-2 was to make it manufacturable on Intel processes. And what a process the team has chosen.

The first clutch of SA-2 processors will be built in a new 0.18-micron CMOS process. This unnamed process is closely related to Intel's P858 (see MPR 1/25/99, p. 22), the six-layer-metal process that will first be used for the misleadingly named Coppermine, due 3Q99. Misleading, because P858 uses aluminum, not copper, interconnections. Like P858, SA-2's process technology will cling to aluminum metallization, although some of the details may change from P858. The process itself is still in development. Indeed, Intel has not yet decided which fab (or fabs) will be outfitted with the required equipment.

So StrongArm will skip a generation, moving from 0.35 micron directly to 0.18 micron, passing over the 0.25-micron generation entirely. That will put SA-2 among the few announced embedded processors using 0.18-micron processing.

The core supply voltage for the SA-2 will be variable, between 0.75 V and 1.3 V; the I/O pad ring requires a separate power supply, as do all existing StrongArm chips. At the lowest voltage, the core should run at a respectable 150 MHz while drawing a mere 40 mW (including the core, caches, cache logic, and MMU, but no bus interfaces). At 1.3 V, Intel expects the SA-2 to hit a remarkable 600 MHz while still holding core power dissipation below 500 mW. There may be more. The closely related P858 process is rated for 1.5-V operation, and the extra 200 mV may give cocky clock-speed cowboys a 100-MHz bonanza.

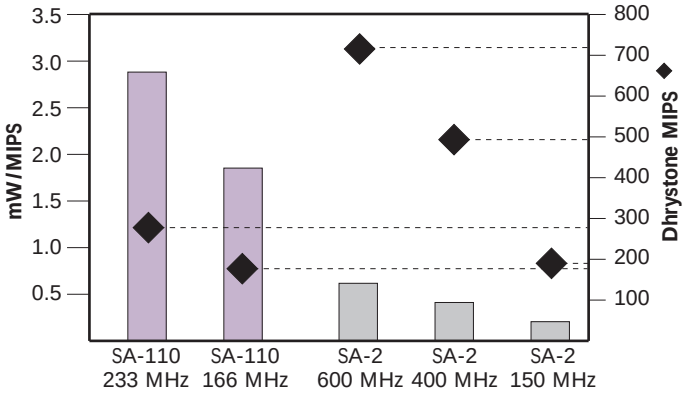

Figure 4. Although it shows just core power (with caches and MMU), this chart indicates that SA-2's ratio of performance to power will be even more exceptional than its predecessors'.

Price & Availability

Intel expects to begin sampling the first of its second-generation StrongArm processors in 1Q00. Features and pricing have not been announced.

For more information, contact Intel (Santa Clara) at *http://developer.intel.com/design/strong/index.htm*.

The SA-2 includes many of the same circuit-design tricks that Digital employed in designing the SA-1: demand clocking, gated clock trees, and automatic shutdown of unused logic. One trick that Intel did not reuse, however, was SA-1's habit of stopping its clock on a cache miss. Because the SA-2 has hit-under-miss cache handling, the core can, and will, continue running after a data-cache miss. This change improves performance, at a small cost in theoretical minimum power consumption.

It's impossible to tell yet what the overall power consumption of a complete chip will be; Intel is nearly a year away from first silicon, and an external bus interface running at a higher voltage will probably consume more power than the core logic. But these low core estimates are certainly encouraging. A hypothetical SA-210 chip could lead the industry in performance per watt, just as its predecessor still does after three long years.

Cache Can Convert to SRAM

Chips this fast live and die by their caches, and Intel has enlarged StrongArm's already engorged units by 16K apiece. The SA-2 core has dual 32K caches plus a 2K "minicache" for data. This minicache made its first appearance on the SA-1100 (see MPR 9/15/97, p. 1) and is included (or, more accurately, will be included) on the SA-1500 (see MPR 12/8/97, p. 12). It acts as a separate cache for transient data; mapped into a different memory space, the minicache alleviates some thrashing of the main data cache.

The 32K data cache can be partitioned, at the user's option, into cache plus scratchpad RAM. Nearly all of the cache (up to approximately 30K) can be allocated as RAM for tables, critical data, and so forth. The remainder of the cache is 32-way set-associative and nonblocking. Even without converting the cache to RAM, users can lock individual lines of either cache, a nice feature that's becoming more common among embedded processors.

Sadly, Heeb's crew did not remedy one of StrongArm's most distasteful features: its inability to flush its own caches. Software must still evacuate stale data from the cache with a programmed loop, although there are some minor enhancements that Heeb did not elaborate upon.

Evidence of Miracles in the Desert

If Dhrystone is any indication (and it isn't), SA-2 should churn out two to three times as many MIPS as the SA-1 generation, as Figure 4 shows. At 2–3× the clock speed, that's not especially remarkable. What is remarkable is that SA-2 can achieve 2–3× the clock speed at all. At an expected speed of 600 MHz, SA-2 will be neck and neck with Coppermine, Intel's other 0.18-micron processor, early in 2000.

The contrasts between SA-2 and the i960, Intel's other 32-bit embedded family, could not be more stark. The i960 survived on cast-off semiconductor processes two to three generations old; the i960CA is still built in a 1.0-micron process (perhaps by little old ladies with X-Acto knives). The original i960MX was so large its corners had to be rounded to fit within the reticle.

Yet the same development team guilty of the i960 has found honest work and brought forth the SA-2. Overturning their convictions, they've gone from oldest process to newest, from highest power to lowest, and from middling performance to potential industry leader.

This may all come as quite a shock to Intel's embedded division in Chandler. Unused to competing on technical merit, this group can now expect to see new doors flung open for them, doors first pried ajar after the acquisition of StrongArm. PDAs, cellular telephones, Web slates, palmtop PCs, and other handheld devices, are now all within Intel's purview.

Intel certainly has the ingredients for success: flash memory, core-logic expertise, LCD and 3D controllers, wireless and network interfaces, and more fabrication capacity than most companies would know what to do with. All it needs now is a winning recipe, something that usually comes only with more time in the kitchen.

Intel's plan to produce highly integrated devices is also highly uncharacteristic. Intel has refined the concept of producing only general-purpose processors to something approaching religion. The greater the integration, the narrower the market, a path Intel has historically shunned. The newly conjured Alchemy (see MPR 4/19/99, p. 5) is headed down exactly this same path, and the two will certainly meet. The sparks between these two—the wizened master and the upstart apprentice—will be both illuminating and spectacular. Hitachi, too, is preparing to emerge from the shadows and claim a piece of the high-performance/low-power terrain. These chips are also about one year away. Early 2000, then, is shaping up to be a watershed time for this product category.

It will be nearly a year before Intel produces first silicon of SA-2, and at least a few additional months before general sampling begins. That's an unusually long warning period for Intel, and could have a chilling (Osborne-like) effect on SA-1 sales. More important, it will cloud competitors' brows and influence design decisions made from now until 1Q00.

Until that time, SA-2 will cast a long shadow over low-power, high-performance competitors. When the clouds break and the chip arrives next year, it will usher in a new millennium for Intel's embedded strategy. M