

M I C R O P R O C E S S O R

www.MPRonline.com

THE INSIDER'S GUIDE TO MICROPROCESSOR HARDWARE

MOORE, MOORE, AND MORE AT ISSCC

Annual Circuits Conference Shows More Than Circuits

By Peter N. Glaskowsky {3/24/03-01}

At the International Solid State Circuits Conference (ISSCC 2003), Gordon Moore reaffirmed the relevance of his namesake law, Chuck Moore of the University of Texas at Austin revealed an intriguing new processor architecture, and—oh, yes—there were papers

on many interesting circuit designs. ISSCC has grown well beyond its roots as a showcase for innovative solid-state circuits to cover architectural theory, commercial implementations, and roadmaps for related technologies.

Among the actual circuits disclosed at ISSCC 2003 were high-resolution CCD and CMOS imagers for motion-picture cameras, a Z-80 on a glass substrate, and several 90nm designs from Intel, including a 10GHz RISC core, a 5GHz FPU, and a 5GHz clock-distribution scheme with less than 10ps of skew, meant for a commercial x86 microprocessor.

Pace of Progress Predicted

Gordon Moore's ISSCC keynote simply affirmed that Moore's Law is safe for the rest of the decade, but that's a significant statement, given how far the industry will go in that time. As Moore said, "No exponential is forever, but we can delay 'forever.'"

The Intel presentations at ISSCC showed what we can expect from the company's 90nm process, coming on line later this year. The 10GHz RISC processor core Intel described as part of an experimental TCP/IP offload engine was the fastest programmable engine shown at ISSCC this year. Figure 1 shows a high-level block diagram of the core, in which the 10GHz RISC core is surrounded by a memory system and support circuits running at 312.5MHz (1/32 the core speed). The RISC core executes predecoded instructions stored in a wave-pipelined microcode ROM with 320 entries, each 80 bits in length. This ROM sustains one access per

clock, with a two-cycle latency for nonsequential accesses—when taking a branch, for example.

The core is built in Intel's dual- V_T 90nm CMOS process using semidynamic flip-flops. The dual- V_T process provides two types of transistors—one with a lower threshold voltage to permit higher switching speeds, the other with less leakage current for use where speed is not so critical. Intel says about 85% of the core transistors in this device

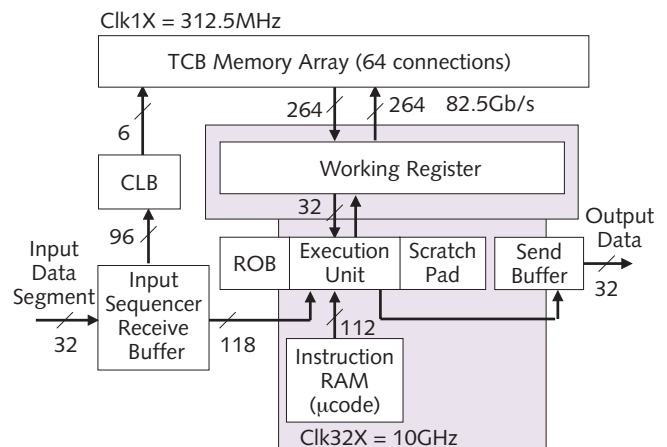


Figure 1. This TCP/IP offload engine described by Intel at ISSCC has a 10GHz RISC core surrounded by 312.5MHz support circuits. This device has been implemented in Intel's dual- V_T 90nm CMOS process.

are low- V_T devices to achieve the target speed, which is fast enough to handle TCP/IP offload on a 10G Ethernet link. The 7.9mm² core has 260K logic and 200K memory transistors and consumes 1.9W at 1.2V. Intel showed a picture of the first test chip running in the lab at about 7GHz just two weeks after it returned from the fab.

Although Intel did not describe product plans based on this device, we believe the company is evaluating TCP/IP offload engines for future network and server processors, where TCP/IP processing represents a large portion of the workload. During the ISSCC presentation, Intel said the CPUs in a dual-processor server (of unspecified configuration) can be 100% utilized for TCP/IP processing for just one Gigabit Ethernet connection, using the worst-case minimum packet size. Given that a relatively small offload engine can handle 10Gb/s traffic using 90nm process technology, it makes good sense to incorporate such an engine in future networked systems.

In a separate presentation, Intel described a 5GHz floating-point multiply-accumulate unit also built in a dual- V_T 90nm CMOS process. Because this unit is single-precision only and does not include a register file, we do not expect to see it appear as-is in future Intel processors. The unit occupies about 2mm², including the FIFOs and scan-chain logic used for input and output. Like the TCP offload engine, this core uses semidynamic flip-flops on a nominal 1.2V supply, and some 90% of the transistors in the design are low- V_T devices. Intel estimates power consumption for 1.2V, 5GHz operation is 1.2W.

A clock-distribution scheme we believe was designed for Intel's forthcoming Prescott processor (see *MPR 3/10/03-02*, "Spring 03 IDF Details Prescott") was described in another Intel presentation. The distribution network uses three layers—a "pre-global" clock network (PGCN) driven by the PLL that provides the source clock, a global clock grid (GCG), and local drivers. The PGCN and GCG are located inside eight horizontal clock-distribution stripes designed into the host processor's floorplan.

The inputs of selected PGCN drivers are shorted together locally to provide a measure of skew attenuation. GCG drivers are inverters of variable size, allowing drive strength to be tuned at design time. The local drive elements provide clock gating for power management and they support delay programmability, which Intel says can be used to help locate critical paths during silicon debugging.

Die-area allocation was specified at about 0.25% at the device and lower metal layers, and less than 2%, 3%, and 5% for the M5, M6, and M7 metal layers, respectively. Power consumption was given as 0.75W/GHz for the PGCN and 1.75W/GHz for the rest of the distribution network. Intel said the network must provide a bandwidth some 38% higher than the product's planned operating frequency to allow the PLL to lock up correctly. For this reason, the network is designed to scale to 6.9GHz of bandwidth to support a 5GHz processor. A micrograph included with the presentation,

which we presume represents the Prescott die, is shown as 10.2mm × 10.7mm (109mm²) in size.

Looking at Imagers

Two high-definition imaging devices for digital cinema applications were described at ISSCC 2003, a color device implemented as a CCD and a monochrome CMOS imager. The CCD came from DALSA (www.dalsa.com) and boasts an imaging array with 4,096 × 2,048 pixels, each 8.4 microns square. This array is slightly larger than one frame of 35mm motion-picture film, making it compatible with existing camera lenses. A storage array of equal size buffers the captured pixel values. This chip is 36.4mm × 36.6mm (1,332mm²) overall and is created by stitching together eight reticle images.

The DALSA imager includes color filters fabricated on top of the imaging sites. Pixels on successive rows alternate between green and red filters and green and blue filters. The color resolution of the imager is thus lower than the monochrome (luminance) resolution, as is the case in the human eye. To support a maximum rate of 60 frames per second, the CCD can be read out across 16 outputs, each transferring 40 million pixels per second, to be digitized with 14 bits of resolution per sample—a raw data rate of almost 9Gb/s.

The imager is supported by commercial digitizing and image-processing hardware that handles the necessary image-processing functions, including lossless image compression. Even with this compression, the output of the camera system is a 790MB/s datastream, or nearly 3TB per hour.

Micron's CMOS imager offers an effective resolution of 3,840 × 2,160 pixels and is much smaller, the whole chip being designed to fit within the standard 20mm-square reticle window. The Micron chip also includes integrated 10-bit ADCs, a significant benefit of its CMOS manufacture. With imaging sites about one-quarter the size of those in the DALSA chip, Micron's CMOS imager offers less dynamic range. DALSA specified its chip as having a "dark noise" equivalent to about 20 electrons per site, with a saturation level at about 80,000 electrons. For the Micron chip, the corresponding figures are 42 electrons of dark noise and a saturation level of 25,000 electrons.

The Micron chip, built in a 250nm CMOS process, also operates at 60 frames per second. The effective data rate from the Micron chip is 5.2Gb/s prior to image processing. For color recording, color filters or a multichip camera would be required.

From the Sublime to the Remarkable

Several ISSCC papers dealt with new implementations of existing microprocessors. Easily the slowest microprocessor shown this year was a Z-80 manufactured by Sharp on a glass substrate, using low-temperature techniques such as chemical vapor deposition. The Z-80's 13,000 thin-film transistors (TFTs), with 2-micron gate lengths, occupied 169mm² and operated at 3MHz at 5V.

Although Sharp's system-on-panel technology is, by Sharp's own evaluation, some 20 years behind conventional silicon LSI processing, there are potential advantages to being able to put complex logic circuits on the same glass substrate as a TFT liquid-crystal display. The incremental cost of additional logic on the substrate is very low, as is power consumption. Sharp plans to move from today's approximately 3-micron design rules to 1.5-micron devices this year and to 800nm by 2005. At the last dimension, Sharp believes an 8-bit processor would operate at about 20MHz, fast enough to enable practical products.

Another silicon-on-insulator implementation, the port of the 180nm Alpha EV7 to 130nm SOI, was presented by Hewlett-Packard. The shrink was described as taking the chip's die size from 397mm² to 251mm²; maximum clock speed from 1.25GHz to 1.45GHz; core supply voltage from 1.65V to 1.2V; and maximum power consumption from 155W down to 100W (estimated).

Cisco described an improved implementation of the Toaster3 network processor (see *MPR 10/16/02-02*, "Toaster3 Pops Up at MPF") to create the NT3. The semicustom NT3 is nearly the same as the fully synthesized T3, but a few minor improvements, such as doubling the instruction store on each core to 4K, were made to the chip itself, and a new package was adopted.

The primary benefit of the semicustom implementation is improved clock speed. We estimated the T3's core frequency at about 160MHz. At ISSCC, Cisco said the NT3 runs more than 2.5 times faster, with the fastest parts yielding more than 600MHz. At MPF, Cisco showed typical T3 implementations, with up to four chips cascaded in series for some applications. With its much higher core frequency, fewer NT3s could be used to achieve the same performance. Power consumption is slightly higher, quoted at 20W for 500MHz operation compared with 14W for the T3. Die size also increased, from 259mm² to 349mm². Cisco reported the NT3 was operational in January 2002.

Architecture Also Addressed

Two pure architecture papers added some variety to the transistor-intensive ISSCC program this year. One, from Princeton University, with contributions by Agere and IBM, covered a cache-optimization technique called *timekeeping*. Timekeeping involves keeping track of the time intervals associated with cache operations like allocation, access, and eviction. Each cache line has its own timer.

Potential advantages associated with cache-line timekeeping include a reduction in cache power and increased efficiencies for victim caches, cache prefetching, and least-recently-used (LRU) algorithms for set-associative caches. Power savings can be achieved by using cache-access patterns to predict when the contents of a cache line are not likely to be requested again. If the timekeeping logic sees a regular pattern of frequent accesses followed by a long interval with no access, it may be beneficial to assume the cache

line will never be accessed again before it is evicted. The cache line can be powered down at this time, until it is needed again. With an 8,000-cycle threshold, timekeeping can reduce cache leakage current by about 75%.

Victim caches are smaller fully associative caches used to store lines of data evicted from the primary cache because of conflict misses. Timekeeping can help identify the cache lines that would benefit most from transfer to the victim cache. If a line subject to a conflict eviction has been accessed infrequently, it may be better not to allocate space for it in the victim cache. According to simulations, timekeeping can reduce victim-cache traffic by 87% while improving overall system performance.

Cache prefetch algorithms often rely on access patterns such as "A followed by B implies C." Such a pattern could trigger the prefetch of item C, eliminating its read-miss penalty, but prefetching C prematurely could force B to be refetched. Timekeeping can be used to predict the duration of B's tenure in the cache as well as contribute to more complex and accurate pattern analysis. Simulated results show timekeeping-based prefetch to improve application-level performance by 11% over prefetch methods without timekeeping.

Timekeeping also has obvious implications for LRU line-replacement algorithms, since it can measure the actual time since each way was accessed in a multiway cache. What's perhaps most interesting about timekeeping techniques is that they work even when the timing information is very coarse. The ISSCC presentation suggests that the timers associated with each cache line need have only two to five bits of precision and can be incremented as infrequently as every 2,000 core clock cycles. The overhead for timekeeping support in die area and power consumption will therefore be fairly low, permitting the technique to deliver overall benefits for future chips.

Perhaps the most important paper at ISSCC, in its potential for long-term influence on the computer industry, came from the University of Texas at Austin. The presenter was Chuck Moore, formerly chief engineer on IBM's Power4 processor and now a research fellow at UT-Austin. Moore began his presentation with a game called "Spot the ALU" that challenged attendees to identify the integer and floating-point ALUs on a Power4 die photo. According to Moore, the difficulty of this game reflects the fact that today's processors allocate the vast majority of their die area to support circuits that do not directly implement application algorithms.

Moore is part of a team at UT-Austin and IBM's Austin Research Lab working on the TRIPS (Tera-op Reliable Intellegently adaptive Processing System) project, which was initiated by Professors Doug Burger and Steve Keckler and is now co-led by them. The team is developing a new family of grid processor architectures (GPA) designed to permit dramatically higher ALU utilization on all types of code. Early simulations show a potential for average instructions-per-cycle values as high as 11 across SPEC2000 and Mediabench

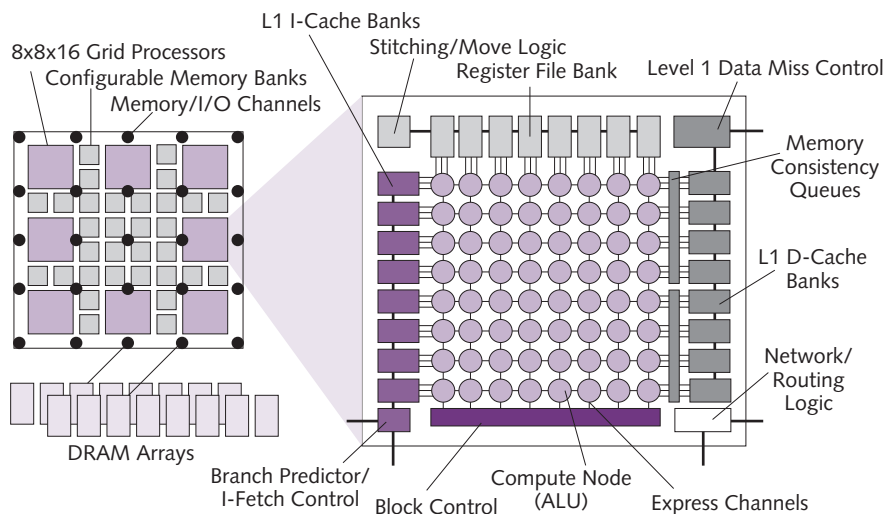


Figure 2. A future grid processor architecture (GPA) device would include several processor cores, each with an array of ALUs, integrated on a single chip with memory banks and off-chip memory interfaces.

benchmarks. In a direct comparison of a simulated 8×8 -element GPA against the Alpha 21264, the GPA design achieved IPCs from 1.1 to 14 times higher.

Figure 2 shows a GPA implementation consisting of multiple 8×8 -element cores surrounded by on-chip memories. Each processing element is a complete ALU with operand and instruction buffers. Instructions specify the destination of results, which become operands for subsequent operations. The compiler defines hyperblocks—code segments with one entry point and potentially multiple exit points—and maps them onto the processing array so that data flows from one ALU to another, as needed. At runtime, the architecture allows multiple blocks to execute concurrently.

Hyperblocks tend to make relatively few references to data elements in the architectural registers, so these references can be scheduled in parallel with code execution. According to the ISSCC presentation, the TRIPS compiler creates hyperblocks with between 14 and 119 instructions (an average of 47) and typically fewer than 10 input and 10 output register references.

The presentation claims four specific advantages to the GPA architecture that allow performance to scale better than in conventional processor architectures. First, the inherent partitioning of ALU, cache, and register files permits greater parallelism in using these resources without the need for complex and costly multiported circuits. Second,

all communication delays in the system are visible to the compiler, so that it can optimize data and instruction flow. Third, the architectural state of the processor is managed on hyperblock boundaries rather than on instruction boundaries, eliminating much of the overhead and need for centralized structures associated with instruction execution. Finally, the instruction buffers in each ALU act as distributed reservation stations, allowing the processor to schedule instruction execution across a window that includes potentially hundreds of thousands of instructions.

The combination of the two-dimensional grid of ALUs and the instruction buffers in each ALU forms a three-dimensional instruction-scheduling window, which offers opportunities for greater performance and efficiency. Loop

unrolling, a common technique for optimizing code on conventional microprocessors, could be used to create larger blocks to be mapped onto the array. Multiple independent threads could be mapped onto the array for simultaneous execution. It may even be possible to schedule the execution of a single code block in parallel with the speculative execution of subsequent blocks to increase parallelism.

We believe the GPA architecture is more promising than other array-processor architectures we've seen over the years (see *MPR 2/18/03-05*, "Extremely High Performance") because it offers a better way for compilers to find and exploit the parallelism in ordinary source code. Realizing the potential of the GPA architecture will require a tremendous research effort, and it will take years to know how this approach compares with more familiar architectures. The TRIPS team plans to tape out its first chip in 4Q04 with four 4×4 -element cores and on-chip L2 caches. Moore estimated this prototype chip will be about 350mm^2 in size, run at about 300MHz in a commodity ASIC process, and have about 1,000 I/Os. Moore says that although the operating frequency of this prototype will be limited, owing to design-resource constraints, the grid-processor architecture is being designed to support frequencies comparable with those of other high-performance processors.

We look forward to seeing further results from the TRIPS project and from the other ISSCC presenters. ♦

To subscribe to Microprocessor Report, phone 480.609.4551 or visit www.MDRonline.com