Low-Power Design, Pentium Dominate ISSCC

Rule Change Makes Conference More Commercially Relevant

by Brian Case

With 112 papers in 18 sessions, the technical program of the 41st International Solid-State Circuits Conference was the largest ever. The program was so large, in fact, that the number of parallel tracks was expanded from three to four to fit all the papers into the two-and-a-half-day program. The large program can be partly explained by the fact that ISSCC no longer requires a paper to be the first public disclosure of a chip.

The main theme of this year's ISSCC was low-power design, which was the primary or near-primary focus of 25% of the papers. In light of this fact, it was curious that the "Technology Directions: Low-Power Technology" session was assigned one of the smaller presentation rooms. Anyone arriving only a couple minutes late found the audience already overflowing several people deep outside the entrance to the conference room.

Meanwhile, just a couple of doors away, the "Video and Communication Signal Processors" session was being held in a cavernous meeting hall that was filled to only about 20% of its capacity.

To their credit, the organizers realized their mistake and swapped the two sessions after the first paper of each session was completed. After the swap, the largest hall was completely filled with seekers of low-power wisdom, a clear indication that this topic is now extremely important to IC designers.

The most impressive space crunch, however, came during the "Microprocessors" session. To accommodate the huge audience, the hotel staff expanded the room during the midsession break by moving the presenter's podium and slide-projection screen ten yards farther back into the depths of the hotel. What generated all the interest? Intel's Pentium presentation, of course.

The Lust for Low-Power Was Evident

Judging simply by the size of the audience for the low-power technology direction session, there is extremely strong worldwide interest in techniques for reducing the power consumption of integrated circuits. Of course, everyone is looking for the holy grail: greatly reduced power consumption with no sacrifice in performance. The high level of interest is due to the growth in the market for battery-operated devices—laptop computers, PDAs, cellular phones—and the trend toward energy-efficient "green" PCs.

Unfortunately, no dramatic advances or magic bullets were revealed. In his plenary session paper, Eric A. Vittoz of the Swiss Center for Electronics and Microtechnology explored ways to approach the limits of low-power design. Tellingly, his paper was essentially a summary of known techniques, some widely used, some not yet commonly implemented.

Among the techniques cited were reducing supply voltage, using the simplest possible logic realization to minimize the number of transistors that switch, suppressing clocks to inactive logic, reducing clock frequency, and removing power from inactive blocks of RAM and logic (using precharged logic/bit lines and not precharging inactive blocks). These techniques are obvious and familiar to those who have followed the development of low-power chips such as PCMCIA interfaces and microprocessors for notebook computers.

Dr. Vittoz sees the reduction of power-supply voltage as the most fertile area for improvement. In particular, as fabrication processes have shrunk, the benefits have been applied exclusively to increasing clock frequency instead of reducing power consumption. In fact, Vittoz points out, process technology has shrunk from 6.0 microns to 0.6 micron—a factor of 10—while supply voltage has dropped from 5 V to 3 V—a mere 40% reduction. His point is that power-supply voltage has been kept artificially high to conform to an industry standard. The 13-W Pentium is an example of the penalty for this simple-minded conformance.

One problem with supply voltage reduction cited by Vittoz is an attendant decrease in performance. As a solution, he suggests parallelism: splitting the task among several duplicate pieces of hardware operating at a lower voltage and correspondingly slower rate. This technique may work well for some tasks—such as image processing—that are inherently parallel, but it is next to useless for general-purpose microprocessors because of the serial nature of program code.

A better idea was presented in a paper by James Burr of Sun Microsystems and John Shott of Stanford. The performance of logic gates is severely reduced when the supply voltage approaches the transistor switching threshold voltage. Their solution is to use back biasing (applying a bias voltage to the chip substrate) to tune the transistor thresholds appropriately. Using back biasing with a 2-micron CMOS process, the authors fabricated and tested an encoder/decoder chip that functioned correctly over a supply voltage range from 5 V all the way down to an astonishingly low 200 mV.

The benefit of back biasing over standard technology is either a performance increase with the same sup-

ply voltage or a decrease in power dissipation with the same performance. For example, a ring oscillator fabricated along with the encoder/decoder operated twice as fast at 5 V—300 MHz with back biasing vs. 150 MHz for the standard process. With back biasing, 150-MHz operation was achieved at 1/10th the power of the standard process, and 20-MHz operation was achieved at 1/5000th the power of the standard process.

Still, even back biasing is nothing new. After the presentation, a chip designer from HP pointed out that HP has been using back biasing in its complex chips for years to adjust transistor thresholds for process variation, reliability, and performance. Burr and Shott have shown that the technique can be used to achieve a significant reduction in power dissipation while lessening the performance degradation of low-voltage operation.

Despite the tremendous interest in low-power techniques, there seemed, in general, to be more heat than light. It is perhaps disappointing that the overriding message seems to be, in Dr. Vittoz's words, that "...limiting the power has its price." The industry must be more flexible in accepting lower operating voltages for complex chips, and the relentless pursuit of maximum performance must be tempered by a consideration for power dissipation at all levels: logic block, chip, and system.

Intel Demos 150-MHz Pentium System

The microprocessor session was one of the best offered at an ISSCC in recent memory. Certainly, the highlight came early with Joseph Schutz's paper on the new version of Pentium known as P54C *(see **080301.PDF**)*. The presentation fitted into the low-power theme by stressing the new chip's ability to reduce power consumption using gated clocks.

Figure 1 shows a graph of power vs. time for the chip running real applications. The solid line is for the chip without the power-saving design changes; the dotted line shows power consumption for the chip with the enhancements. In general, the enhancements result in power savings, but, curiously, when the Indeo application is running, there is a short time when power consumption for the enhanced chip is significantly greater. This blip in power consumption remains unexplained.

Perhaps the most interesting part of the presentation came at the end: the demonstration of a 150-MHz Pentium system. Schutz had the system connected to a high-resolution video projector so the audience could witness the performance. Indeed, spreadsheet operations, Photoshop image manipulations, and 3D renderings seemed to fly by at an impressive pace.

Intel has talked about high-speed x86 processors at ISSCC before; in particular, a 100-MHz 486 was the subject of a paper in 1991. Intel's record of delivering commercial versions of these technology demonstrations in a timely manner, however, is not sterling: the 100-MHz 486DX4 is only now being introduced.

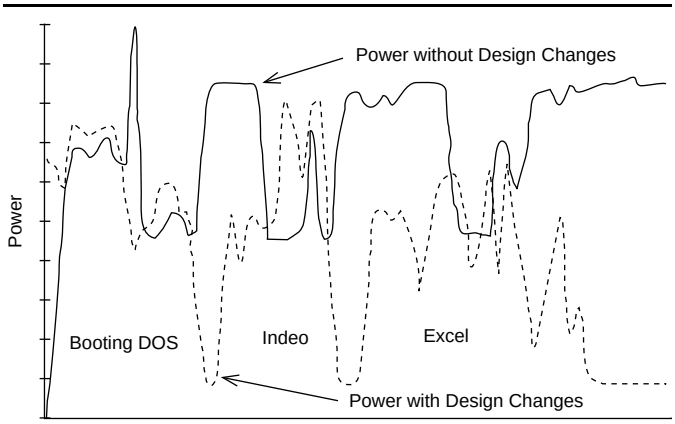

Figure 1. With the addition of gated clocks and other design changes, Intel's new Pentium consumes much less power.

In a break from this tradition of vaporware, the 100-MHz Pentium detailed in this year's paper was officially announced within a few weeks. The paper claims that the chip operates at up to 133 MHz at room temperature, even though a 150-MHz version was demonstrated. Perhaps Intel is being conservative with its claims this time.

There *Were* Other Microprocessors

The other papers in the microprocessor session were also interesting. Toshiba and Silicon Graphics presented information about the integer chip of their two-chip, high-performance MIPS implementation called TFP *(see **071102.PDF**)*, while MIPS Technologies (now a subsidiary of Silicon Graphics) presented a paper on the R4200 *(see **070701.PDF**)*.

The latter chip uses just about every trick in the book to achieve low power consumption. Of course, the operating voltage of 3.3 V contributes significantly. Other incremental improvements resulted from a simple, five-stage pipeline that implements both integer and floating-point operations, clever logic design techniques, and the write-back cache that reduces use of power-hungry I/O pads.

IBM detailed its 33/66-MHz 486SLC2 processor with a 386-style bus and a 16K, four-way set-associative cache. It is implemented in a 0.8-micron CMOS process tailored for 3.3-V operation. The process has three levels of metal and a single level of polysilicon. The resulting chip measures 9.0 ×7.7 mm^2 and dissipates 1.8 W.

HP continues to follow its own unique design philosophy for processor chips. Its 140-MHz PA-7200 *(see **080302.PDF**)* continues to use off-chip primary caches instead of on-chip caches. The off-chip caches are cycled at the processor clock rate, which requires "industry-standard" 6-ns SRAMs (if there is such a thing) to meet the 7.1-ns cycle time.

The first sign that HP is moving toward the use of

more traditional on-chip primary caches is that this processor implements a small, 2K "prefetch-miss cache." HP is holding out as long as possible before building a processor chip with on-chip primary caches, and with good reason. Because it designs its own processors and the systems in which they are used, it makes sense to optimize the overall design to exploit large, bulk SRAMs.

IBM had a second paper describing the PowerPC 603 chip *(see **071402.PDF**)*. Operating at 3.3 V, the chip will dissipate an estimated 3 W at 80 MHz. Power dissipation is reduced both automatically—by disabling clocks to unused logic blocks—and under software control. Software can select one of three power-saving modes: Doze, Nap, and Sleep.

The final microprocessor paper was from NEC and described a 500-MHz RISC processor with a proprietary architecture. The chip is implemented in a 0.4-micron, three-level-metal CMOS process with an eight-stage pipeline. With 200,000 transistors integrated on a 7.9 × 8.8 mm^2 die, this processor is essentially a demonstration that a very high speed processor is possible, but it is not expected to sell commercially.

Work Continues on Large RAMs

Three 256-Mbit DRAMs were described in papers from Mitsubishi, Matsushita, and Oki (only one 256-Mbit DRAM was presented last year). Given that 4M DRAMs are still the staple of our computer-memory diet, it seems unlikely that 256M chips will have a near-term impact, but eventually they will be commonplace. Die size is probably one reason we will have to wait a while: even though all are implemented in 0.25-micron(!) processes, the Mitsubishi chip is 13.3 ×22.8 mm^2, the Matsushita chip is a wafer-scale-like 16.4 ×25.1 mm^2, and the Oki chip measures 13.2 ×25.9 mm^2. These chips are hardly inexpensive to produce. The good news is that they are fast. The Mitsubishi chip has a 34-ns access time, the Matsushita chip has 100-MHz serial I/O ports, and the Oki chip has integrated cache and cache tags (33 8-Kbit data blocks with one address tag per block).

Intel detailed its 16-Mbit flash memory chip, implemented in a 0.6-micron CMOS process. This process, with two metal layers and two polysilicon layers, is not the same process used for the P54C Pentium. The chip measures a manageable 11.6 ×10.7 mm^2. To maximize its applicability in all types of systems, this flash memory can operate from either a 3.3-V or 5-V supply (but a 12-V supply is still required for programming).

The chip uses complex algorithms for controlling write and erase operations. These operations are actually implemented with an on-chip, 9-MIPS processor. The processor, called the write state machine (WSM), has a 16-bit ALU, a three-port 16 ×16 register file, a 2K × 19-bit instruction memory, and a program counter unit. The program counter supports one level of interrupt and four levels of call/return.

This embedded microcontroller could allow Intel to customize flash memories for different markets and differentiate its products from those of the competition. This could be one of the highest-volume embedded RISC processors ever. And you thought Intel was only going to make money selling 486 and Pentium processors.

NEC described a 220-MHz pipelined 16-Mbit SRAM implemented in a 0.4-micron BiCMOS process. The die size is 22.1 × 11.7 mm^2. This SRAM is designed not as a general-purpose RAM part, but as a building block for a four-way set-associative cache. NEC sees a need for high-speed, synchronous second-level caches in engineering workstations.

TI Entering Projection TV Business?

Texas Instruments described its digital micromirror chip, which consists of an array of 864 ×578 individually controllable mirrors. The mirrors are fabricated on 17-micron centers, which results in a chip size of 12.0 ×16.8 mm^2. Each rotates on its axle through a field of ±10 degrees and electromagnetically transits between the two mechanical stop positions in 20 µs.

In TI's application, an 800-W xenon arc lamp was reflected off the mirror array and focused through a lens to create a 5-foot-diagonal image on a unity-gain projection screen. The resulting image had a brightness of 25 footlamberts and a contrast ratio of better than 100:1.

Continuous-tone images are created by rapidly alternating a mirror's position to create the illusion of varying shades of gray. Full-color images can be produced either by converging the images from three mirror chips and color filters or by rotating a set of color filters in front of the light source and time-multiplexing the mirrors of one chip. According to TI, either method can be used to produce 24-bit color images in VGA resolution at 30 frames per second.

ISSCC Becoming More Relevant

To be sure, much of the content of ISSCC papers is still aimed at circuit designers and device physicists, but the types of chips and the rules for inclusion in the program are making the conference more relevant to the business side of the semiconductor industry. Other papers presented at the conference concerned MPEG-2 video codec chips, modem chips, wireless communication chips, A/Ds and D/As, and disk-drive electronics. These are circuits that can impact products within a year or two. The program contains a healthy share of traditional "blue-sky" topics such as neural networks, nanoelectronics, and superconductivity, but the program committee seems to be creating an increasingly attractive balance between chips and ideas with future possibilities and those destined to impact products in the near term. ♦

