A Conversation with Dan Dobberpuhl

The computer industry has always been about power. The development of the microprocessors that power computers has been a relentless search for more power, higher speed, and better performance, usually in smaller and smaller packages. But when is enough enough?

Two veteran microprocessor designers discuss chip power and the direction microprocessor design is going. Dan Dobberpuhl is responsible for the design of many high-performance microprocessors, including the PDP-11, uVax, Alpha, and StrongARM. He worked at Digital Equipment Corporation as one of five senior corporate consulting engineers, Digital's highest technical positions, directing the company's Palo Alto Design Center. After leaving Digital, Dobberpuhl founded SiByte Inc., later acquired by Broadcom. In an October 1998 article, *EE Times* named him one of "40 forces that will shape the semiconductor industry of tomorrow." He has written numerous technical papers and is coauthor of the text, *Design and Analysis of VLSI Circuits*, well known to a generation of electrical engineering students. Dobberpuhl is the named inventor on nine issued U.S. patents and has several more pending patent applications in various areas of circuit design.

David Ditzel directs our conversation with Dobberpuhl. Ditzel is vice chairman and chief technology officer of Transmeta Corporation, which he cofounded in 1995 to develop a new kind of computer—one that would learn how to improve its performance and save power as it ran, by using software embedded in the processor itself. Before founding Transmeta, Ditzel was director of SPARC Labs and chief technical officer at Sun Microsystems' microelectronics division. Ditzel came to Sun from AT&T Bell Laboratories in 1987, where he was the chief architect of the CRISP Microprocessor, AT&T's first RISC chip. His work first attracted industry-wide attention in 1980, when he coauthored "The Case for the Reduced Instruction Set Computer (RISC)."

DAVE DITZEL Dan, do you want to introduce yourself and say a couple of words about your professional background?

Have we maxed out

YET ON MICROPROCESSOR

POWER? TWO INDUSTRY

VETERANS DISCUSS

THE TRADE-OFFS.

DAN DOBBERPUHL I

recently founded a new fabless chip company called P.A. Semi. I have been in the industry for 36

years, and have been developing microprocessors since 1976. I have seen a lot of changes during that period in terms of both silicon technology and microprocessor development.

DITZEL You've got a wonderfully long history here. And our topic today is really about low power. Do you

remember the very first computer you worked on, and can you take a guess at what kind of power that computer probably took?

DOBBERPUHL Well, it's interesting Dave. The power dissipation of the early MOS chips wasn't all that high, because the frequencies were so low and the chips were small and the transistor counts were very low. But they were typically in the range of three to five watts.

DITZEL In those days you probably used a minicomputer or mainframe to run your CAD tools, and there was a big difference between the performance and size of those development machines and the microprocessor chips you were developing.

DOBBERPUHL Sure. So when we developed the LSI-11, our processor design environment was basically a PDP-10, which was the time-sharing machine of the era. That was a large multi-rack system. And at the time the performance ratio was very high between those devices and the

chips we were developing, in factors of 10 to 100. Over the next 10 years we brought that to equivalence and then basically the CMOS devices took over.

DITZEL So let's set the stage. The LSI-11 started about what year?

DOBBERPUHL The original LSI-11 started in the early 70s. It was designed jointly by engineers at Digital and Western Digital.

DITZEL And that had probably a MIPS rating. How many million instructions per second do you think that ran?

DOBBERPUHL I believe the clock rate on the original LSI-11 was about 1.3 megahertz. That was the microinstruction rate. I would guess that the machine was probably about 0.2 MIPS. The next generation, the LSI-11/23, bumped the clock rate all the way up to 3.3 megahertz.

DITZEL Another chip project that you're fairly famous for

is the DEC Alpha chip. That seemed to take a somewhat different approach in terms of trying to drive up the performance, but you drove up the power at the same time. How much did the power go up from that of the LSI-11. How hot did the Alpha chips get?

DOBBERPUHL: When we started on the first Alpha chip design, which was in about 1987 or 1988, we did an extrapolation of where we thought the industry leaders would be in terms of clock rate and performance over the two-and-a-half to three years it would take to do the development.

At the time I think that people were in the 30- to 33-megahertz clock range in 1987-88. We projected that they would not be any higher than 100 megahertz by 1991, so we set our goal as 200 megahertz in 1991. Everything else in that development was subservient to the 200 megahertz.

That was the goal. Whatever it took. If we could get the current in and the heat out, that was all that mattered. It turned out that the power dissipation of the 200-megahertz Alpha, which was in 0.75-micron technology, was about 30 watts, which was at the time fairly incredible for a CMOS chip.

DITZEL In particular, it's not just the total power, but I think because the voltage is low, the current is fairly high.

DOBBERPUHL It was very high, about 10 amps average current. I remember that the clock driver width totaled about 20 inches to get the sub-nanosecond rise times that we needed. And current pulled during switching was about 50 amps. So it was fairly spectacular at the time, because people really thought about microamps and milliamps in CMOS, and we were talking tens of amps.

DITZEL There were very few chips that took that much power. I remember one engineer saying, "Oh my goodness, the circuit breaker coming into my house barely handles that amount of current. And you're trying to pull that into one small silicon chip." How could you deal with those problems? Weren't there big issues in just keeping the chip cool at that point and getting the power in on the wires?

DOBBERPUHL Sure, I think actually the cooling wasn't as difficult a problem as getting the current in, because we were in wire-bond package technology and the pin counts weren't all that high. So there were two big issues. One of them was just the total current, the average 10 amps. The big problem was the peak current, which was in excess of 50 amps.

The fact that it switched in less than a nanosecond meant that we had to do something different. The big issue was basically the power supply stability—how to get

around the $V=L*di/dt$ problems. That's where we came up with the idea of using very large on-chip decoupling capacitors in order to smooth out the current waveforms going into the chip.

DITZEL Did such large amounts of current cause surprises and things you hadn't expected to be problems, which you think you were one of the first to hit?

DOBBERPUHL Fortunately, we were able to identify most of them in advance. I can't recall that we were surprised by anything subsequent to getting the chips back.

But we had to do a lot of things like the decoupling capacitance, which was integrated into the clock driver, which you could see easily with your naked eye on those chips. It went down the center of the chip. Massive transistors plus decoupling capacitors were distributed across the whole center of the chip.

So I think the decoupling capacitance and coming up with rules to manage on-chip inductance, also an issue at the 200-megahertz clock rate, were kind of a first.

DITZEL How did you build those capacitors? I actually think that's going to turn into an interesting issue here. Did you build that out of transistor material?

DOBBERPUHL Those were just ordinary NMOS devices.

DITZEL But you have a very giant transistor, so to speak, that had a lot of capacitance.

DOBBERPUHL Exactly. And that was a bit of an issue for the process guys as far as the amount of total gate capacitance or gate area that was on the chip, in terms of the yield.

But it was also true that we knew at the time that most of the defects in the gates were along the edges; so these things were made relatively square, and you got a good ratio of area to perimeter.

DITZEL This is one of the cases where reading the textbook example of how you solved the problem may be changing with time. As we go into deeper sub-micron technologies, we're finding that gate leakage, which was almost insignificant when you did those early processors, now could turn out to be a larger problem, that building a capacitor that way might generate too much leakage.

DOBBERPUHL It's a big problem. Leakage, as I'm sure we'll get into, is a big issue today. It was much less of an issue 10 years ago. These large decoupling capacitors, if you just make them out of modern thin gate oxide, are way too leaky.

DITZEL So in the future people are going to have to find new ways of building capacitors.

DOBBERPUHL There are a lot of different things you can

do, including having a thicker layer of dielectric in the capacitor. Basically the transistor is suffering from the same problems. The capacitor is an easier problem because you have a lot of area available, so you don't need the capacitance per square meter in a capacitor that you do in a transistor.

DITZEL You went from building a chip that had one of the highest amounts of power in the industry in the Alpha to next working on the StrongARM chip [http://www.arm.com/armtech/ARM_Arch?OpenDocument], which was a very, very low-power chip. That's going from, "I don't care about power, I'll use whatever it takes to get megahertz," to doing almost the opposite. How did you undergo this religious conversion?

DOBBERPUHL Well, it was interesting—not the least of which was the amount of, if you'll excuse the pun, heat we took for the power in the Alpha chip. We wanted to show that not only could we design the highest-performance chip, but we could also design a high-performance low-power chip.

We presented Alpha in 1991 at ISSCC [International Solid-States Circuit Conference], and I think at ISSCC 1992 I was impressed with the work that Bob Broderson and some of his colleagues and students were doing at Berkeley on low-power devices. Bob was preaching the religion of low voltage.

After ISSCC, I went back to Massachusetts and said to the process guys, "Let's run an experiment and see if we can just fabricate a relatively low-voltage, low-power Alpha."

So we reduced the VDD supply from 2.5V to 1.5V. As a result we got a pretty decent performance/power ratio on an Alpha chip. As I recall, the frequency went down to about 100 megahertz from 200, but the power was down to less than 5 watts from the original 30.

DITZEL A huge improvement.

DOBBERPUHL Huge improvement—of course, lower performance but much better performance per watt. That looked encouraging, and we said, "Well, what if we started from scratch and tried to design a really high-performance low-power Alpha chip?" I got excited about doing that.

In the meantime, another group of folks had left Digital, then returned, but were going to be based in Austin, Texas. Among them were Rich Witek and Jim Montanaro.

They also got very interested in low-power high-performance CMOS. But they said, "Instead of building a

low-power Alpha, let's soup up an existing low-power chip like ARM." So for awhile we had one group working on a low-power Alpha and another working on a high-performance ARM.

Well, the low-power Alpha turned out to be an interesting technical idea but not an interesting marketing and business idea, because the whole premise of Alpha was high performance. Doing anything of lesser performance didn't seem to make a whole lot of sense.

So that concept kind of died and we all concentrated on the idea of building a high-performance version of an existing low-power chip. We ended up, for various reasons, choosing the ARM architecture.

We set about building what became StrongARM, using, to a large degree, the same circuit techniques that we had used on Alpha, but with a power budget in mind instead of just frequency *uber alles*.

DITZEL Now you went from a 30-watt part in Alpha. Where did the first StrongARM chip come out when you were first able to measure it?

DOBBERPUHL It was 300 milliwatts at 160 megahertz.

DITZEL So you took off a factor of 100? That's pretty good in terms of technology scaling. One thing that may surprise people is that you didn't have to drastically change your circuit design technique or other issues. What did you have to do differently to get that factor of 100?

DOBBERPUHL We did use pretty much the same design techniques. What changed were the constraints and

parameters. For one thing we had to dramatically reduce the transistor count, which is a major factor in power dissipation. So it was a much simpler device—much, much simpler. Our estimate was that a factor of 3x was achieved simply due to the reduction in transistor count.

DITZEL For example, I don't think it had a floating point unit.

DOBBERPUHL That's correct. And the integer data path was only 32 bits versus 64; it was single-issue instead of superscalar, etc.

DITZEL People marvel sometimes at the small handheld devices. But in some sense the capabilities have changed a bit, but I assume those features we once had in the 50-watt chips will start to come back in the low-power chips over time.

DOBBERPUHL For certain, they will. And that's really the challenge we're facing today, in that the technology has advanced to the point where it's very easy to integrate lots of functionality. What's difficult is to manage the power concurrently with all that functionality and performance capability.

It used to be hard to design a 200-megahertz chip, just from the point of view of getting the transistors to go fast enough. Now the issue is they'll go plenty fast, but how do you keep it from burning up?

DITZEL Maybe for the readers here, you could talk just a bit about where power in the chip comes from. What makes the chip get hot, and what are the fundamental

components in making the power go up or down?

DOBBERPUHL The power is dissipated mostly in the transistors, either as they switch or as they just sit there and leak.

You can calculate the dynamic power dissipation with the formula $P = CV^2f$, where V is the power supply, C is the capacitance that is being switched, and f is the switching rate. There are some additional factors, but fundamentally the dynamic power is given by that formula.

DITZEL That capacitance is really related to the number of transistors, so you reduce the number of transistors and it reduces the power.

DOBBERPUHL Absolutely. The problem is that we want the functionality associated with large transistor counts and we want the performance associated with high frequencies. So we have to lower the voltage to manage the power. But as you lower the voltage, the transistor performance is reduced. As you reduce voltage, that can have a pretty significant impact on performance. The way to compensate for that is to reduce the threshold voltage (V_t) of the transistors.

DITZEL That sounds simple enough. Why not just make it very low?

DOBBERPUHL Well, we do, but there are problems with that. The leakage current of a MOS transistor is proportional to inverse exponential V_t . So as V_t goes down, the leakage current goes exponentially higher. Also, V_t is a function of temperature. So you have another effect that

you have to allow for.

DITZEL Particularly in mobile devices where they're encapsulated, since they can get hot on the inside, which makes the power situation worse.

DOBBERPUHL That's right. Even though the power dissipation of the mobile device is necessarily low to conserve battery life, typically there's not much that you can do to cool the device. You don't want to put a fan in there; in fact, you don't even want to put in a heat sink. A lot of times the case becomes the heat sink. So the internal temperatures can be high. Basically this lowering of V_t is a double-edged sword. You get the improvement of performance but you get an increase in static leakage power.

It used to be that the only people who worried about static power were the watch guys, because the watches had to last for six months or a year or more on their batteries. And that leakage current was significant. But for most applications it didn't matter. Now we've gotten to the point where leakage current can be in the same order of magnitude as the dynamic current of the device.

DITZEL Can you talk a little bit more about this leakage thing? What is leakage to the first order? Because we didn't hear about this a few years ago, power was just CV^2f . But now we've got leakage coming in as a new factor. What's a simple way to picture leakage for somebody who is not a transistor engineer?

DOBBERPUHL A simple way to think about the threshold voltage is as a barrier holding back electrons, like a dam holding back water. Electrons have various statistically distributed energies, like waves on a body of water. As the barrier or dam is lowered, a higher percentage of waves will spill over the dam. The same thing applies to electrons.

And the other significant factor, which we mentioned

before, is gate leakage. There the problem is that the gates are so thin that the electrons will tunnel through from the channel to the gate material and actually create a current.

DITZEL So transistors are now getting so small—maybe they're on the order of 14 angstroms in 90-nanometer

technology, and you might want to go to 12 angstroms. I've heard that just that 2-angstrom difference might cause a 10x increase in gate leakage. Is there anything people can do about gate leakage in future devices?

DOBBERPUHL Well, something has to be done, because proper scaling requires both vertical and lateral reduction. And we certainly want to keep scaling according to Moore's law. Obviously, people are working on the problem. One of the ways to attack the problem is to use a higher-density dielectric for the gate material, which would allow you to have a bigger physical dimension between the channel and the gate.

I think there are other techniques that are under development. It's a problem that's being addressed. But I think the leakage problem is a fundamental physics problem that, at the moment, has no real complete solutions. It's going to take a lot of design engineering finesse to manage it.

DITZEL I was on a panel at the Kyoto VLSI Symposium recently where there was a combination of people who make semiconductors: the fab guys and the people who design with them, the circuit and logic guys. The fab guys were all saying, "Hey, we solved this problem over the last 20 years. Now it's somebody else's turn because the physics just is limiting us in what it can do."

DOBBERPUHL It is. As you said, we now have vertical dimensions that are on the order of 12 or 14 angstroms,

only a few atomic layers.

DITZEL So will that change who is willing to go to the newer technologies and how fast they'll go to it? I've heard that as people are switching from today's technology—where we're currently at about .13 microns—as we go into 90- and 65-nanometer here, that maybe it's going to be a different kind of customer who will go there first. Maybe it won't be the standard company doing ASICs [application-specific integrated circuits] that will go, but somebody else. How do you think it's going to shift the patterns of usage?

DOBBERPUHL It's hard to say. There's a slowdown in the rate of transition to the more advanced technologies. It's caused by a number of factors, including the technical design issues—of which the leakages are a very significant one. There are others as well. Also, as just a financial issue, mask costs are extraordinarily high below 0.1 micron—on the order of \$1 million for one mask set.

DITZEL Where do you see technology taking us? Today I guess one of the complaints we have is people have PCs with four or five fans in them when the CPUs are dissipating 50 or 60 watts. Should we expect that our PC desktops are suddenly going to turn quiet in the next 12 months? Or are people going to use more power to get higher megahertz. What do you have in your crystal ball for the next five years?

DOBBERPUHL I think you're in a better position to answer that than I am, Dave. This is a kind of a marketing and business trade-off. There are a bunch of technical constraints, and right now we do the best we can within those constraints.

You can have high performance or you can have low power or you can have a combination. But you can't have the extreme of either. So how you choose to parameterize a given system is really a function of what you think the customer wants.

DITZEL I think it's almost a psychological factor of when people believe they have enough performance. Because if they don't have enough, they're going to push for higher and higher megahertz, which is probably going to come at the expense of power continuing to go up. But wouldn't it be true that once people feel like they have enough performance, if you just need to sustain the performance levels, technology goes forward. Shouldn't that give us opportunities to reduce power?

DOBBERPUHL No question about it. If all you want to do is sustain current levels of performance, then advances in technology make that easier and easier. The problem

comes when you try to extract the maximum amount of performance from the advanced technology.

DITZEL So I guess—given the operating system you have today and maybe the PC you have today—if you're happy with the performance, the real problem for business is you're not buying a new PC or a new operating system.

In some sense one can say that the people who design operating systems and new processors are looking for ways that you won't have sufficient performance with the machine you have today, so that you'll need to buy something new. Is there any way out of this vicious cycle?

DOBBERPUHL I don't know. From an engineering point of view, we'll continue to try to improve the performance-per-watt ratio. It's interesting that a low-power processor can deliver on the order of 1,000 MIPS per watt. And maybe 2,000 for a really good one. It depends upon the absolute performance level.

And the high-performance processors are about an order of magnitude worse than that. They're on the order of 100 MIPS per watt. So there's a huge spread, and that's mostly a function of the same things we were talking about earlier with the Alpha and StrongARM.

It's where you choose the design point—whether you go with 100 million transistors or 20 million transistors

and whether you choose two gigahertz or one gigahertz. These things are all parameters the designer chooses, but then he has to live with the consequences.

DITZEL When you build chips, do they all come out at the same speed and the same power?

DOBBERPUHL Don't we wish? The semiconductor process variation is quite wide and follows a statistical distribution that designers target anywhere from three to six sigma points of the variations, depending on what kind of yield they want to get. In terms of frequency, it's not unusual to have a 40 or 50 percent spread from the

slowest to fastest coming out of the same process line.

DITZEL So you could actually think of buying a computer, and one might have a different battery life than another just because of the distribution of chips and how they're made.

DOBBERPUHL Absolutely. The power dissipation varies, just as the performance does. The dynamic part of power dissipation is not that variable. But the static power dissipation is highly volatile and can vary dramatically from device to device.

DITZEL I notice you have a PDA device with you here. It seems to me in many ways people had to change the operating system and the user interface with PDAs because they couldn't fit what was in their PCs into something that was handheld. Do you think we'll be able to get the performance of a desktop PC into your hand in a few years?

DOBBERPUHL To a large degree, yes, I do. I think it's mostly going to come from design finesse—somewhat from technology, but just a lot of continuing design development because we're kind of at the wall in terms of the physics.

DITZEL One of the other issues for power is not only the power the chip consumes computing things, but also the power it takes communicating to the other chips on the board—the I/O power. Are we going to have to do something about that problem in the future as well? For example, will we have to do more integration on chips? How do we get I/O power to go down?

DOBBERPUHL Yes, I/O power is an issue. And I think it's a piece of a bigger issue, which is system-level power overall. We've been concentrating primarily on microprocessors, but really the same issues apply to all silicon chips.

People have spent many years doing system architecture and micro-architecture based on performance. Now I think it's very relevant to look at system architecture from a power point of view. And it may cause us to make some changes in the way we architect systems. We design memory systems to optimize performance. In these power-constrained systems, it makes a lot of sense to design memory systems to optimize power, or a combination thereof. So I think, then, the same thing goes for I/O.

Off-chip clock rates in the last few years have increased dramatically with the DDR [double data rate] concepts and things like that. Consequently, power has increased as well. Managing I/O power is an issue, and again it's the kind of physics issue that can't be avoided, and so it behooves the designers to deal with it and make trade-offs.

DITZEL One of the things I've heard is that in order to reduce I/O power, people are moving to the system-on-a-chip approach where the functionality of many chips is integrated into one. How does that affect the design challenge? I presume it's harder to design one mega-chip than individual chips with many fewer gates per chip.

DOBBERPUHL Well, you're absolutely right in terms of integration saving power. You can save a lot of power if you can move interconnects from chip-to-chip to on-chip.

You can have higher performance as well. So it's a double win. But the problem is that as you integrate more functions onto a chip, the design problem grows somewhat linearly and the verification problem grows exponentially. And perhaps in many situations, the general applicability of the chip decreases as you specialize it more and more. But those are engineering trade-offs.

DITZEL In some sense designing a custom hardware chip for every specific use might be very well powered in some optimal sense. But if the chips are getting so complicated, it seems to me you'd want a more programmable chip.

You might want to use a chip for many different things. I'm thinking of some of the future video chips—for example, where the algorithm itself keeps changing often. Does that force us to do kind of these super mega-chips that are trying to cover many bases all at once? And can you really cover the power ranges people want from super-low power in your wristwatch all the way up to what your notebook computer would be?

DOBBERPUHL Hardwired logic is considerably more power efficient than programmable logic. We talked about high-performance processors being 10 times less efficient than low-power processors. Hardwired logic is probably on the order of 10 times more efficient from a power point of view than a good low-power processor. So

there are serious advantages to hardwired logic in chips. The obvious disadvantage is that it's hardwired. Again, these are engineering trade-offs, and different situations demand different applications.

But it is generally true, I think, that programmability has been the direction that the industry has taken over time. Things that were originally hardwired eventually became programmable—because of the flexibility, the adaptability, the ability to fix bugs. There are a lot of advantages to programmability, but power is not one of them.

DITZEL Do you think we're going to go through a period of evolution in the next several years? Or do you see some maybe unexplored technologies like a new circuit family or a new kind of transistor that's going to give us a big breakthrough? Or are we going to follow a Moore's-law curve?

DOBBERPUHL I think it's going to be pretty much evolutionary. There are improvements in the technology coming beyond the scaling that I think will be helpful. But they're not going to be orders-of-magnitude helpful. And we really have an orders-of-magnitude problem.

DITZEL There's a fairly famous chart I've seen that Pat Gelsinger at Intel did that showed chip power in PC processors from the original Intel 8086 chip up through the future. It unfortunately shows power drawn on a log scale. If we were to continue this trend, the heat density of these chips would be as hot as a nuclear reactor.

DOBBERPUHL That is correct.

DITZEL So what people say is, "Well, technology marches on, but obviously this can't continue." Where are we in terms of heading ourselves toward the big problem, and what are people going to do about it?

DOBBERPUHL I think we're there; we have a big problem. Power dissipation is limiting performance across the board, both on high-end devices, as well as on mobile devices. And we've got a lot of smart people working on it. There haven't been any super breakthroughs. But there have been evolutionary improvements.

DITZEL So you don't think it's practical for people to be building one-centimeter-square, one-kilowatt chips that go into consumer electronics devices?

DOBBERPUHL No, I do not. I don't remember the exact number, but on the order of a two-centimeter-square chip, you can't mechanically extract more than about a 100 watts without the temperature going up way too high.

So we're very much constrained in terms of how much

total power an individual chip can dissipate per square centimeter on the high end. And on the low end we're constrained by thermals and batteries.

DITZEL So, Dan, a lot of the issues with power for the past 10 years have been worked on sometimes by what I'll call the circuit engineers, who have been trying to run at lower voltages and just take advantage of the natural scaling of semiconductor processes. Do logic designers, architects, and software programmers have anything to contribute in the future?

DOBBERPUHL Most definitely. We've developed lots of tools to do performance analysis for software development—and to understand the hot spots from a performance point of view in code and make improvements to tune code, to improve performance, or to reduce memory footprint, etc.

I think it's appropriate at this point to think about tools that would allow software programmers to analyze and optimize code for power dissipation. So far, as you say, most of the power conservation work has been done by the hardware engineers.

Chip designers have put a lot of hooks into silicon, many of which really haven't been fully exploited. Software really has the big context of what's going on in the system; the hardware has a very small understanding of what's going on. So the software should be able to do the best job of power management. But the hardware has to have the hooks to allow it to do it.

DITZEL Among the chips that are starting to take a lot more power these days are the graphics chips. In fact, they're taking as much or more power than the CPU chips are taking. Is there anything fundamentally different between a graphics chip and a microprocessor in terms of making it go toward low power in the future?

DOBBERPUHL I think it's also true that graphics chips have started to become more programmable. That probably exacerbates the power-dissipation problem. So I think that they really are facing the same problem that the microprocessor guys are in terms of managing performance and power.

DITZEL What do you expect your handheld device to do for you five years from now that your current device doesn't do? Thinking not so much in terms of technology, but if you could imagine a personal handheld device, what do you think we'll be able to do in the future that we can't do today?

DOBBERPUHL I think that, for sure, there's a path toward convergence of voice, video, and data. And we all want to have it with us all the time, and it's definitely possible and feasible. So I think with good engineering we'll have

handheld devices that allow us to communicate in those three dimensions with great efficiency.

DITZEL In terms of technology that might save us, in the last few years we've heard a lot about something called silicon on insulator, a variation of standard CMOS. Is that going to replace standard CMOS technology in the future?

DOBBERPUHL Well, the proponents would say that it will, and the opponents will say that it won't, and only time will tell. The issue I think it struggles with is that it has an advantage over standard silicon in terms of performance and power of about 25 to 30 percent—which is about what you gain from one generation of silicon technology. It is a more complex technology, and it is more expensive. Because it's not in widespread usage, it's not at the same level of development as standard silicon at any point in time. That lag can wipe out its advantage. So it has been a struggle for it to go mainstream. Certainly there are those who predict that it will. But it's not there yet.

DITZEL So we've seen, sometimes, that a mobile chip, like a mobile Pentium chip, might cost more than a desktop chip—even though its megahertz seems to be lower. Why is a low-power chip more expensive sometimes than a higher-powered desktop?

DOBBERPUHL Sometimes it might be for marketing reasons. But there are certainly reasons why it might actually be more expensive to make.

One way to build a lower-power chip is to take one that's at the high end of the speed distribution and reduce its power-supply voltage to get a reasonable clock rate and low power. In that case, you're taking the very best-performance chips and turning them into low-power chips, so that's one reason.

The other reason might be that it uses some process enhancement and may just be fundamentally more expensive to build.

DITZEL Any final comments?

DOBBERPUHL I guess I'd like to turn the tables and ask you a few questions, Dave.

DITZEL Feel free.

DOBBERPUHL Transmeta has certainly made a mark in achieving a very unique approach to power minimization in a standard architecture, the X86 architecture. I'd be interested to hear what you think in terms of how much that architectural approach has contributed to your low-power success and how much of it is just straightforward power minimization through design?

DITZEL I think we get a reduction in the number of transistors. And for the part of the equation that's CV^2f , reducing the transistors is a big component. So I would say at least half of our power advantages come from the simplicity of design and some other issues that come forward.

But I think that the other big part for Transmeta's chips is allowing other engineering teams—in this case, the software engineering teams—to contribute by putting in more sophisticated algorithms in how one might dynamically change the voltage, the megahertz, reading the temperature of the silicon, and using other on-chip parameters to have control of items—something you wouldn't attempt to do in hardware lest you did it wrong.

The fact that we can do some of these sophisticated control items in software lets us attempt a little bit more and attack the problem a little bit more vigorously. I think, as we've seen, reducing power is an issue. It's not that there's any one magic bullet, but it's doing 25 different things all right at the same time and letting each of the different groups—whether it's product engineering or working with the process or the logic designers—be able to contribute in a way where maybe there just hadn't been as much focus on it in previous years.

DOBBERPUHL Are you optimistic that you'll be able to continue making progress without hitting the wall?

DITZEL I think there's another good five to ten years of power reduction in place, where we'll take chips that to date have had to be substantially different architectures, like the StrongARM chips, because there were no low-power PC chips. I think in the future we'll be able to take standard PC architecture applications and make them just as low power as we have other so-called low-power chips today—to the point where we can have handheld devices that can be running full Microsoft Windows XP and where you wouldn't have to have a different operating system on your handheld device from your desktop device from your cellphone.

Then all these things could communicate and run applications. I don't think today about loading an application onto my cellphone. But I think in the future one can do that.

DOBBERPUHL Well, great, in five years let's look back and see how we did.

DITZEL OK, let's make a date. ☺

LOVE IT, HATE IT? LET US KNOW:

feedback@acmqueue.com or www.acmqueue.com/forums

© 2003 ACM 1542-7730/03/1000 \$5.00.