The Coming Reality for SOC Designers

by
John Bourgoin
Chairman and CEO of MIPS Technologies, Inc.

Abstract
Continued rapid semiconductor technology advances bring with them some great benefits. But the trends of designer productivity continuing to fail to keep up with transistor budgets, the explosive growth in mask costs, the huge increases in the minimum dice quantities created by 300mm wafers and small geometry chips, and the relentless time to market pressures will have a major impact on the choices SOC designers will be making in the next few years. Programmability and high performance will increasingly take center stage as desirable elements in future SOC designs.

It's hard to imagine a more diverse and dynamic marketplace than the market for embedded microprocessors. Opportunities abound and new ones are emerging like never before.

On the other hand, there are challenges. Design costs, mask costs and wafer costs are increasing. Profits must be increased by, not only getting products to market faster, but also extending their time in the market. These challenges can be met by taking advantage of Moore's law and soaring transistor counts, performance headroom, and software programmable solutions. It's a trend that has been around for a while. Today, it's accelerating.

It started with the invention of the programmable computer in the 1940s, when software programmable processors began to subsume tasks that previously required custom logic. By the late 1970s, small microprocessors were being used to replace custom designs. In the '80s and '90s, there was astounding growth for these embedded processors and, as processor performance grew, function after function moved out of custom hardware and into software.

When the technology reaches 90nm, complexity, skyrocketing non-recurring engineering (NRE) costs and greater pressures on time-to-market will drive system OEMs to a new approach. That approach will take advantage of increases in microprocessor performance, and programmability will be used to eliminate custom design wherever possible.

IC options for the system developer

Traditionally, system developers have three basic options in how they approach a design. For minimum hardware design cost and maximum flexibility, they can use a general-purpose processor or other programmable device, like an FPGA. For more performance, they can use special-purpose programmable devices like DSPs. For maximum performance, they can do an

ASIC or full custom design. Today's SOC designers often develop solutions that combine all of these approaches to take advantage of the best features of each in the same SOC.

How they make their decision is based on many issues. All things being equal, they likely choose the most flexible, least risky approach. So, let's start with the most flexible, least risky option — a programmable solution — and ask three questions:

1. Can it perform the function (performance, power, feature set, etc.)?
2. Can it meet the cost needs of the market?
3. Can I get it to market early, or at least early enough?

Question #1: Can it perform the function?
As silicon technologies approach 0.1-micron, something is becoming increasingly apparent: cheap transistors. Transistors, even in the hundreds of thousands or millions, will cost so little that processor or programmable device developers will readily trade them to enhance performance anywhere they can.

In fact, extensions to standard processors are becoming increasingly commonplace. These additions, though standard, will be modular and can be selected in or out of the processor depending on market needs. The key point is that their presence will accelerate the performance of the base processor by one or even two orders of magnitude on many important applications. Examples include the MIPS32™ 4KSc™ core for smart cards, which utilizes extensions that increase the performance of encryption and decryption by 5 times. Intrinsity's new FastMATH product, which utilizes an innovative circuit approach and matrix processing extension, is expected to operate at 2 GHz in 0.13-micron technology and perhaps 4 GHz in 0.10-micron.

Each generation of products has more features and functions, and there is a constellation of applications within reach of programmable devices. Increasingly, each new SOC subsume more functionality from other ICs on the board because of the huge economic advantages of integration. Among the challenges of new systems today:

- streaming media requires high bandwidth to and from the chip, as well as high levels of computational power

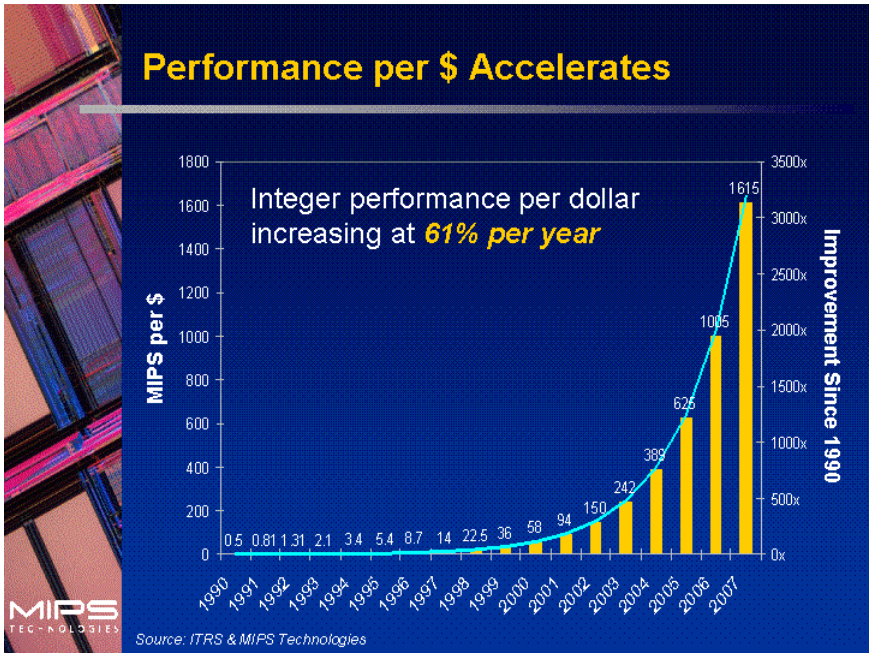
- the need to process packets at wire speeds and service video and audio applications demands high throughput

- most systems require that multiple tasks be serviced simultaneously, often with short latencies and good performance

- applications are increasingly driven by real-time constraints, which demand short and deterministic latencies and high computational power

- security is likely to become crucial in most systems, and so will the need for high computational power for encryption and decryption

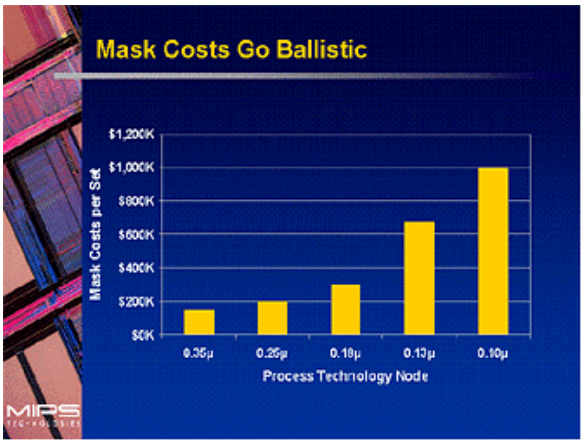
Meeting these challenges with a programmable solution requires very high performance. Fortunately, Moore's law shows no signs of slowing down, which gives us ample opportunity to use more transistors to solve market problems. According to the International Technology Roadmap for Semiconductors (ITRS), both transistor density and chip frequencies are increasing more than 25% per year. MIPS Technologies estimates that integer performance of microprocessors is increasing 57% per year, and integer performance per dollar is increasing at a 61% compound annual growth rate. That adds up to a 500x improvement in performance per dollar since 1990. Simply projecting this forward, without the accelerations in performance that we believe will come from augmentations, says that processor MIPS per dollar will climb to over 1000 by 2006.

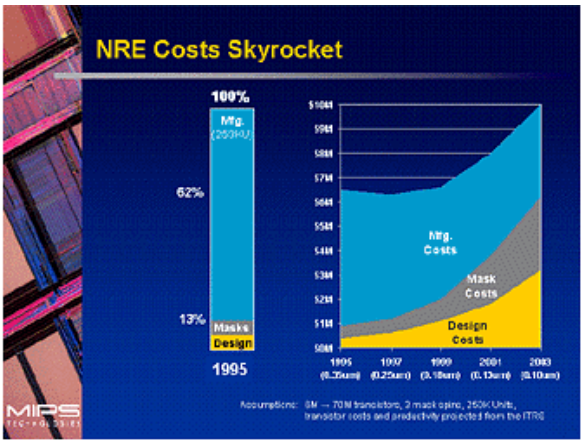


Effective performance will increase even faster. Many of the future advances will come from architectural techniques that are already known. Architecture enhancements will boost streaming media performance well beyond the pace of traditional integer MIPS performance growth. Advanced micro-architecture techniques will boost processor throughput and response times and increase tolerance to slow memories. In addition, chip multiprocessing techniques in many applications highly suitable to parallel processing raise performance levels by large factors. And, floating-point performance is rising even faster than integer performance, making many calculations far simpler to implement at high performance levels.

So, the answer to the question is yes, for a rapidly increasing number of applications, programmable products will perform the function needed. Microprocessors, in particular, will extend their performance capabilities well up into the range of applications that formerly required DSPs or, in some cases, even custom hardware designs, because the extremely low cost of fast transistors permits more and more sophisticated extensions and specialized processor support on new processor engines.

Question #2: Can it meet cost needs?
For years, the big cost-related question has always been, "How large is the die?" Soon, it will be, "How much will it cost to get the die into production?" The trend is crystal clear: The cost of developing advanced SOCs is skyrocketing.





Just a few years ago the cost of a mask set seemed very high at a few hundred thousand dollars. Now, it's roughly a million dollars for a 0.10-micron set. If you've got more than two or three spins of silicon, your costs are out of control. And look at what's happening in the area of non-recurring engineering (NRE) charges. Around 1995, mask and development costs for a typical SOC, including three mask spins and a total run of 250,000 units, ran to about 13% of the total cost of the product. Not so by 2003. That 13% is rising to 62%.

That said, a wise strategy would be to leverage a single design across as many products as possible. A key advantage of programmability is that different products can be serviced from the same integrated circuit.

Suppose, for example, that a product is designed for use with a cable modem. It might have a microprocessor to act as the controller, security hardware, a vocoder, a fax capability, an Ethernet port, a series of standard peripherals, and hardware to decode the data coming down the cable. As more powerful processors become available and are implemented in the SOC from the beginning, many functions can be subsumed into software run on the processor. What's particularly interesting is that the chip now has significantly more flexibility, which allows it to be used across a range of applications; it is no longer limited to the cable modem. The software does the tailoring, not the hardware. All this means you can get the product out, right the first time, with a simpler design, applicable to more products. And, there's a greater likelihood that you can reuse some, most, or all of the design.

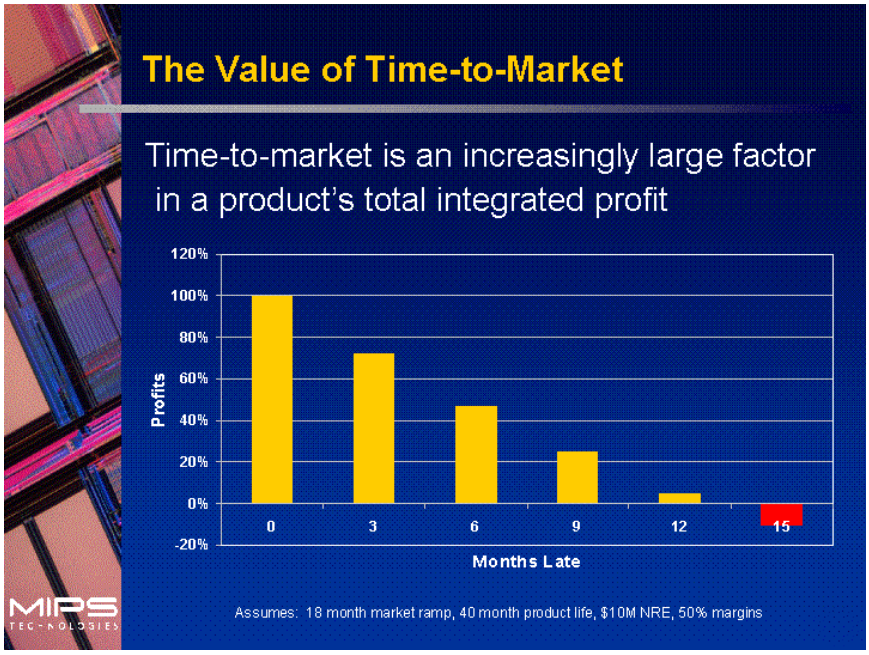
Programmability also lowers NRE by allowing fixes to be implemented in software. And late-breaking changes of protocols or new features can be added without changing the silicon. This adds up to fewer silicon spins at a million dollars apiece.

Total design costs also are likely to be lower. Since much of the design can be in software, the size and complexity of the custom portion of the design can be minimized or, in some cases, even eliminated. Reuse of blocks such as processors or I/O circuits means that libraries and drivers can be reused as well as the circuit layout.

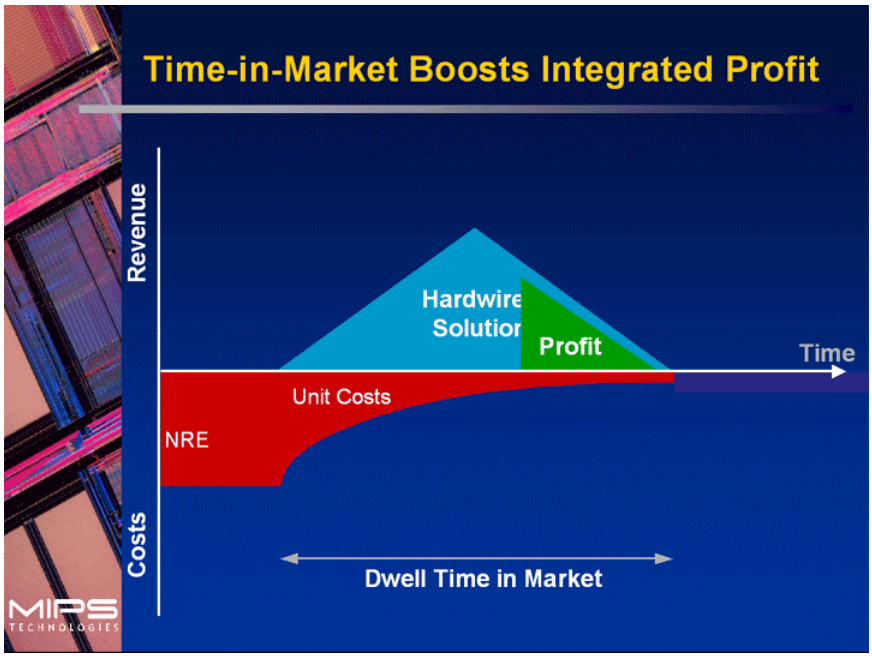
So, in answer to the cost question, future system designers must view their decision by taking into account, not just the unit cost of the IC, but also the development cost and the benefits of reusability.

Question #3: Can I get to market early?

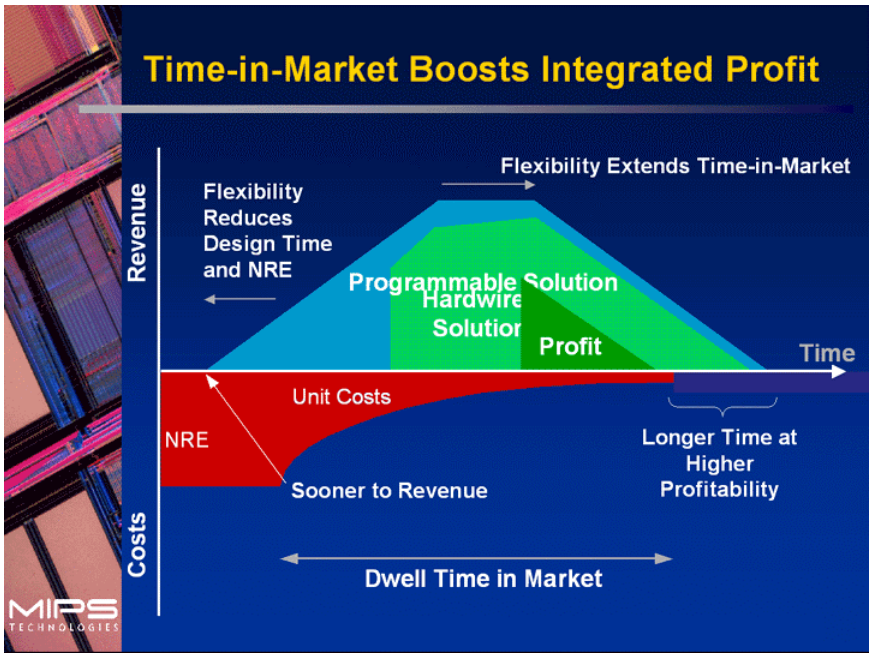
The most important consideration for the system designer is time to market. The chart below shows the percentage of profits lost on a 40-month product life cycle with an 18-month market ramp, $10 million in NRE charges, and 50% margins. Most would probably agree that 18-month product ramp windows are on the long side for most of today's products so the picture is actually worse than this chart paints. Even under these conservative assumptions, half the profit is lost if you are six months late. And if you are late by half the ramp, virtually all profit is lost.



The following slide shows a traditional model of time-to-market versus revenue. Profit is the difference between the revenue and the cost at any point in time and is delayed until the up-front development costs have been returned. The first example here is a hard-wired or custom IC solution:



Shown below is the effect of getting to market sooner with a programmable solution. It reduces the up-front NRE costs, accelerates the time to revenue, and increases the dwell time in the market and may even allow for larger market shares, thus increasing the peak revenue and extending the time in the market place. Programmable solutions can also adapt to changing market environments and, therefore, extend the life of the product *in* the market, resulting in profits that may be several times larger than what you would earn by entering the market a few months later.



Programmable solutions allow product features or standards changes to be inserted after the tape out. And programmability with built-in performance headroom allows for tweaks in the product definition well after the initial concept has been committed in hardware. Even better, the same capability can extend the lifetime of the product by allowing for new features or adjustments to meet changing market conditions, thereby protecting the product from early obsolescence.

So, in answer to the question, yes, you can you get to market early — with flexible system designs that ensure working products early on, which can be modified at the last minute to meet market requirements.

Keys to success

To make all this come true, processor vendors need to apply the falling costs of transistors to lower the rising costs of development.

By 2005, one million transistors will cost 26 cents, according to ITRS 2001, and that includes packages *plus* a 60% gross margin. Using a few million transistors to deal with all the issues mentioned above is a "no-brainer." Increasingly, transistors will be applied to create higher performance processor cores, which will enable more programmable solutions. The result will be lower NREs and faster time to market, resulting in lower costs, higher revenues, and higher profits.

Many techniques will be used to do this. Architectures are progressing from 8-bit to 16/32-, 64- and 128-bit. (64-bit architectures have been used in consumer products for years, and a 128-bit architecture has been a key element of success for the Sony PlayStation® 2.) Wider architectures will utilize the extra bits by applying single instruction, multiple data (SIMD) to performance by increasing the number of data elements operated upon in parallel. Vector processors will soon provide an extended parallel operation capability for cost-sensitive embedded markets. Fine-grain multithreading will keep the processor busy when it's waiting for data from other threads, and it helps mitigate the effects of the growing processor-memory speed gap. And, techniques like chip multiprocessing will dramatically improve the apparent performance of chips where highly parallel applications like data path processors in routers are important.

The challenge — to get products to market faster, to extend their time in the market, to lower development costs, and take advantage of the continued march of Moore's law — is to find ways to leverage transistors for exactly that purpose. Programmability and performance headroom offered by high-performance microprocessor cores will, by necessity, be the preferred solution.