Platform 2015: Intel® Processor and Platform Evolution for the Next Decade

Executive Summary

Without a doubt, computing has made great strides in recent years. But as much as it has advanced in the last ten years, in the coming decade, the emergence and migration of new workloads and usage models to mainstream computing will put enormous demands on future computing platforms: demands for much higher performance, for much lower power density and for greatly expanded functionality. Given these seismic shifts in the uses of computing, how we define and architect future computing platforms will have to change dramatically, holistically comprehending and satisfying not only computation, but interface and infrastructure requirements as well. Intel's long-range vision for the evolution of these three fundamental platform elements, and the architectural innovation and core competencies enabling that evolution, is what we call Platform 2015. In addition to examining relevant trends and usages and their platform processing implications, this paper will focus on the computational element of Platform 2015, describing the evolution of Intel's microprocessor architecture over the next 10 years.

Shekhar Y. Borkar
Intel Fellow
Corporate Technology Group
Director, Circuit Research
Intel Corporation

Pradeep Dubey
Senior Principal Engineer and Research Manager
Platform Architecture Lab
Intel Corporation

Kevin C. Kahn
Intel Senior Fellow
Corporate Technology Group
Director, Communications Technology Lab
Intel Corporation

David J. Kuck
Intel Fellow
Software and Solutions Group
Director, Parallel and Distributed Solutions Division
Intel Corporation

Hans Mulder
Associate Director
Intel Research Network of University Labs
Intel Research and Director
Intel Corporation

Stephen S. Pawlowski
Intel Senior Fellow
Enterprise Platforms Group
Chief Technology Officer and Director,
Platform Planning, Architecture and Technology
Intel Corporation

Justin R. Rattner
Intel Senior Fellow
Corporate Technology Group,
Systems Technology Lab
Senior Director
Platform Planning, Architecture and Technology
Intel Corporation

Edited by R. M. Ramanathan and Vince Thomas, Intel Corporation

Contents

Introduction

The microprocessor is on the verge of undergoing perhaps the most significant transformation since its creation requiring advances in addition to achieving the predictions of Moore's Law. Continuing increases in transistor count are vital, of course, but the next leap is also going to take some thorough rethinking of basic foundations such as process technology, architecture and software. In fact, that rethinking is already in progress; the foundations for the next decade are even now being laid. And from these beginnings we can start to anticipate what processors will look like ten years down the road. We will provide a glimpse of that future, describing microprocessor architecture over the next 10 years, the trends and the applications driving this outcome and the technical challenges that only a very few companies are equipped to address.

As demonstrated throughout Intel's history of processor and platform innovation, this is an evolutionary approach in which Intel integrates new features into our chips to bring more performance and capabilities to the platform while maintaining compatibility with thousands of already existing applications and support for the ecosystem surrounding them.

Macro Trends

Processors and platforms of 2015 will be a direct outgrowth and response to social and technological trends already evident today. Examples of these trends include:

Pervasive Connectivity and Proactive Computing

Omnipresent connectivity is becoming a fact of daily life. There are more than 1.4 billion mobile phone subscribers in the world today – more than 20 percent of the world's population. Increasingly, people expect anytime, anywhere access to information, services and entertainment. That expectation is driving a profusion of wireless mobile devices: notebooks, handhelds, "wearable" systems and other go-anywhere gadgets. Likewise, "implicit" (i.e., embedded) computing is increasingly commonplace, with processing carried out invisibly in all kinds of products and environments: appliances, cars, buildings, toys – even the human body (via medical implants). At the same time, computing is becoming more proactive – capable of automatically anticipating, adapting to and serving the needs of users. Corporations are already piloting networks of wireless sensor chips to better manage everything from factories to vineyards, and provide new kinds of healthcare services. And computers are evolving towards more natural interfaces such as speech, handwriting and image recognition.
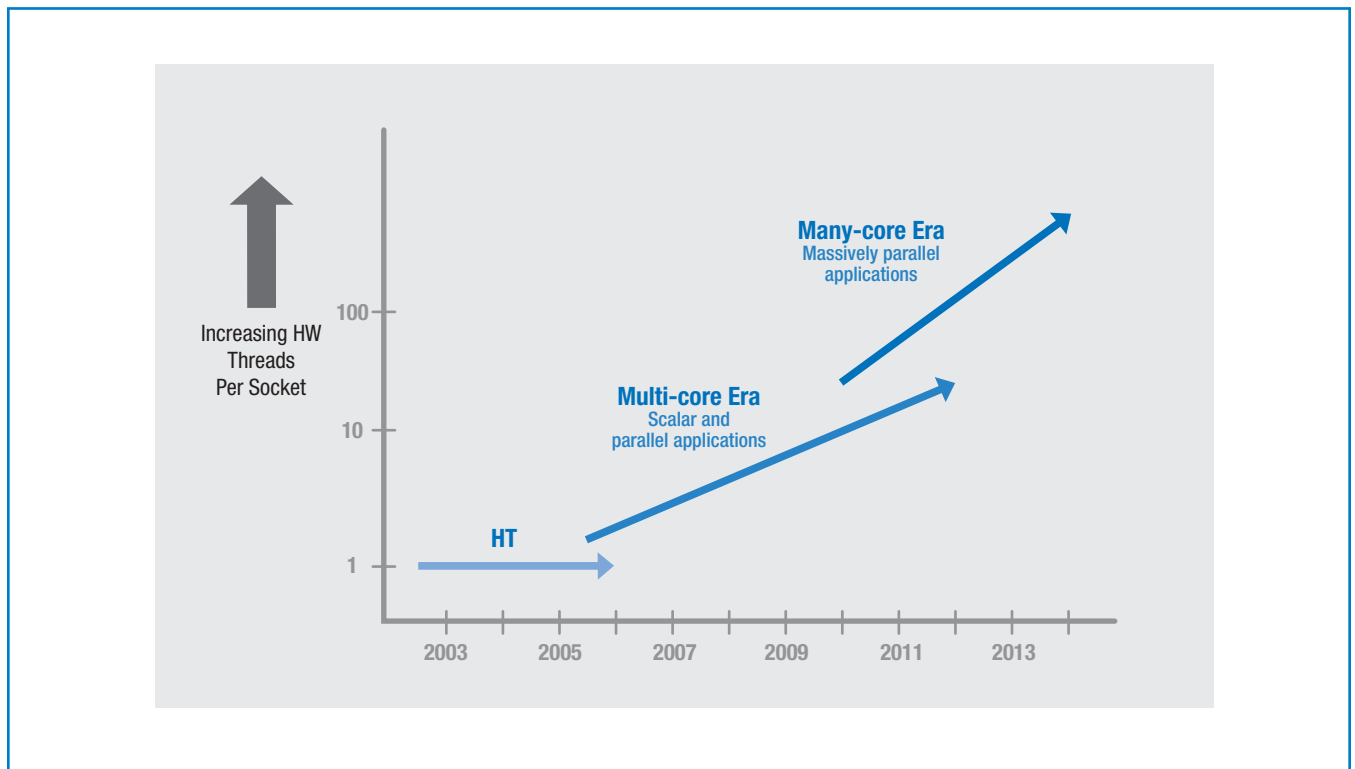

Figure 1: Current and expected eras of Intel® processor architectures

Ubiquitous device-to-device connectivity and proactivity require that devices receive and handle a huge variety of data from a vast array of sources. Consequently, data will have to be self-describing – containing not only the information to be processed, but also the descriptive "meta data" needed to interpret and process it correctly. For example, imagine a camera that automatically aggregates information related to location, environment, time, and the activities happening in its vicinity – builds the context relevant to a picture/video-frame and builds the "meta data" part of that picture/video-frame itself. With this approach, a video camera in a gas station could issue an alert of a potential theft in real time. Today's computing devices lack the computational power to perform these tasks in a cost effective way. Computing devices must be developed that continuously learn and aggregate information relevant to the data objects they work on. These machines will need enormous processing power and memory in order to learn on their own, behave intelligently and become more and more proactive.

For instance statistical modeling methods like Bayesian networks could be efficiently used to calculate posterior probabilities and decision support systems could be developed based on those probabilistic relations. Machine learning technologies based on such mathematical principles along with increased processing power could further enhance several applications such as medical diagnostics, genomics, audio-visual processing and detection, and analysis and testing of models.[1]

Data Growth and High-Performance Computing

Pervasive connectivity and proactive computing is leading to a tremendous increase in the amount of data used by individuals and organizations. The amount of data is doubling every 18-24 months as a result.[2] That data must be transmitted, stored, organized, and processed in real-time to be made useful. Today we have gigabytes of photos, music, text and video on our systems being processed by gigahertz platforms. Soon, our systems will host terabytes of digital data, requiring teraflops and TIPS (trillions of floating point operations and trillions of instructions per second) of performance.

These levels of performance will enable everyday use of new types of workloads and applications. One such workload is called "RMS" (recognition, mining, synthesis) and is made up of: 1) the ability to *recognize* and *model* information of interest to a user or application on an ongoing basis; 2) the ability to *mine* large amounts of real-world data in real-time for the information; and 3) the ability to *synthesize* large datasets or a virtual world scenario, interactively and in real-time, based on the informa-

tion. These RMS workloads, when applied to the massive data streams that are becoming commonplace, will require super-computer-level performance in ordinary servers, PCs, and even compact mobile devices and embedded systems.[3]

This level of performance enables unimaginable new applications on desktop and mobile devices. These applications will exploit the tremendous processing power of these future platforms as well as their ability to access data anywhere in the world. Some of the examples will include applications like real-time decision-making analysis, intelligent search, real-time virtual rich media and photorealism, interactive story nets, multi-party real-time collaboration in movies, games and strategy simulation and real-time medical imaging.

Internet as Computer and Conduit

While the Internet will continue to function as a communication pipeline, it will also become smarter, grow substantially in bandwidth, and deliver larger and larger amounts of intelligent data and applications to connected devices (thus adding to the performance demands to process and search this data as described above). The number and variety of these devices will snowball, as a multitude of new implicit systems and sensor nets come online. In order to provide access to rural users in the developing world, the Internet will need to support new forms of device connectivity and mechanisms for delay tolerance (due to the unreliability of the connections) and transcoding (converting content to a form appropriate to various devices). These new Internet-connected devices will eventually number in the tens of billions, driving data traffic to mind-boggling levels. At the same time, the Internet will increasingly serve as a kind of global computing platform, providing the connectivity to a vast array of new services and applications.[4] This will require a massive infusion of processing power around the world to support the computing requirements of these new services and will be delivered in several forms including new and more powerful servers.

Globalization

The Internet, high-speed data communication, and increasing wireless connectivity are bringing the world together into a corporate and consumer global village. Work teams, commerce and social exchange are increasingly spread across disparate geographies, cultures and languages. As people and organizations worldwide collaborate and interact more closely together, they will increasingly rely on computation to help them bridge those gaps. Imagine, for example, headsets that perform real-time language translation. Globalization also drives the need for new form factors, price points and intelligent interfaces –

to reach a greater range of users for whom conventional computers are impractical, too expensive or too complex. For example, new performance rich, low power and low cost mobile devices have been envisioned for rural populations in the developing world with variable literacy rates and limited access to reliable power grids and networks.

Processing Implications on Converged Platforms

These trends have a number of clear implications for future computing platforms.** Specifically, processor architectures need to evolve to be able to support increasing performance, functionality and market requirements over the next 10 years. We believe there are at least six key requirements to be met:

General-purpose Performance

It almost goes without saying that performance must scale steeply to support ever increasing massive amounts of data and complex workloads. "Tera" level data sets, RMS applications, advanced multimedia, natural interfaces and real-time language translation, increased device intelligence – to name a few examples – need much, much higher computational horsepower than today's processors can manage. Pro-active computing – in which devices act intelligently on our behalf – ups the performance ante further.[5] So too does the increasing reliance on self-describing data, since it requires the processor to handle not only the basic data but also the meta data attached to and describing it. And the transformation of the Internet into a services-hosting platform will generate its own additional demand for increased processing power. Together these trends call for a substantial, order-of-magnitude leap in overall processing performance, relative to current-generation computing platforms.

Power Management

While the performance requirements are soaring, power and heat limitations remain. With the rise of mobile computing, pervasive connectivity and proactive computing systems, devices are becoming smaller and more mobile, which means platforms must consume less power and produce less heat, not more. Hence the power density must be far lower than it is for today's processors and platforms. As processors and platforms evolve over the next 10 years, they will require a variety of advanced and active power management and energy-conserving technologies.[6,7]

Special-purpose Performance and Adaptability

Many of these future applications and usage models will require special types of processing in addition to high-performance, general-purpose computation. Future processors will need fast on-chip circuitry for specialized functions such as media processing, new user interfaces (e.g., speech recognition) and communications protocol processing (e.g., network packet processing and XML). Moreover, even general-purpose processors will have to be far more versatile and adaptable than they are today, able to dynamically match their capabilities to wide range of changing workloads.

Reliability, Security, Manageability

As computing devices are always connected and perform more mission-critical functions – maintaining our homes, conducting our financial transactions, monitoring and attending to our medical needs – security and reliability take on a life-and-death importance. Meanwhile, security threats are on the rise, with a growing pool of attackers, new kinds of attack and many new points of attack (e.g., new mobile devices, sensor nets and wireless networks). The burden of managing computer systems is likewise skyrocketing. In 2003, almost half of the $1.1 trillion spent on IT went to manageability, which remains heavily reliant on costly human intervention. To prevent these problems from spiraling out of control, the new platforms will incorporate world-class reliability, security and manageability features into the silicon itself – and much of that directly into the processor.

Ecosystem Support and Stability

Providing processors and platforms with teraflops of super-computer-like performance for new applications and workloads without a broad-based development environment and ecosystem behind them will greatly limit their technical and market potential. Therefore, maintaining support and stability for the established software development ecosystem and installed base of applications will be critical to the success of any future processor and platform architecture.

Mass-market Economics

All of this functionality must be delivered at mainstream price points, accessible to billions of global consumers.

** These trends will exert influence on the evolution of all key computing platform elements not just computation and the microprocessor. This includes IO, memory, interconnect, storage, peripherals, communication, OS and software. In this paper, these topics will be addressed only in the context of the processor as computational core.

Microprocessor Architecture in 2015: The Intel Roadmap

Looking ahead, Intel® processors and platforms will be distinguished not simply by their raw performance, but by rich and diverse computing and communications capabilities, power-management, advanced reliability, security and manageability, and seamless interaction with every other element of the platform.[8] Intel's roadmaps encompass the following key characteristics of processor architecture circa 2015, or "Micro 2015" for short:

1. Chip-Level Multiprocessing (CMP)

Intel continues pioneering in one of the most important directions in microprocessor architecture – increasing parallelism for increased performance. As shown in Figure 2, we started with the superscalar architecture of the original Intel® Pentium® processor and multiprocessing, continued in the mid 90's by adding capabilities like "out of order execution," and most recently with introduction of hyper-threading in the Pentium® 4 processor. These paved the way for the next major step – the movement away from one, monolithic processing core to multiple cores on a single chip. Intel is introducing multi-core processor-based platforms to the mainstream. These platforms will initially contain Intel processors with two cores and evolving to many more. We plan to deliver Intel processors over the next

decade that will have dozens and even hundreds of cores in some cases. We believe that Intel's chip-level multiprocessing (CMP) architectures represent the future of microprocessors because they deliver massive performance scaling while effectively managing power and heat.

In the past, performance scaling in conventional single-core processors has been accomplished largely through increases in clock frequency (accounting for roughly 80 percent of the performance gains to date). But frequency scaling is running into some fundamental physical barriers. First of all, as chip geometries shrink and clock frequencies rise, the transistor leakage current increases, leading to excess power consumption and heat (more on power consumption below). Secondly, the advantages of higher clock speeds are in part negated by memory latency, since memory access times have not been able to keep pace with increasing clock frequencies. Third, for certain applications, traditional serial architectures are becoming less efficient as processors get faster (due to the so-called Von Neumann bottleneck), further undercutting any gains that frequency increases might otherwise buy. In addition, resistance-capacitance (RC) delays in signal transmission are growing as feature sizes shrink, imposing an additional bottleneck that frequency increases don't address.
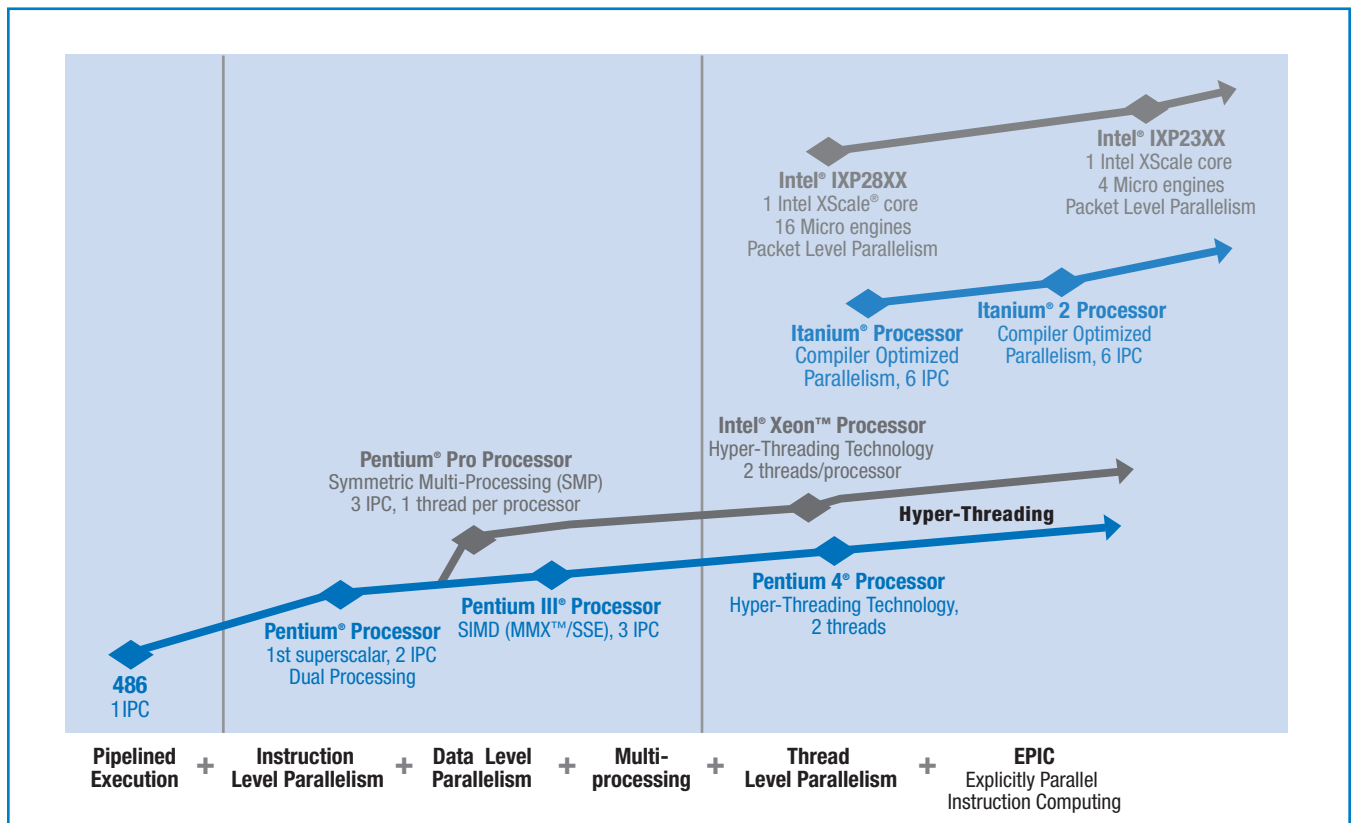

Figure 2: Driving increasing degrees of parallelism on Intel® processor architectures

Therefore, performance will have to come by other means than boosting the clock speed of large monolithic cores. Instead, the solution is to divide and conquer, breaking up functions into many concurrent operations and distributing these across many small processing units. Rather than carrying out a few operations serially at an extremely high frequency, Intel's CMP processors will achieve extreme performance at more practical clock rates, by executing many operations in parallel.[9] Intel's CMP architectures will circumvent the problems posed by frequency scaling (increased leakage current, mismatches between core performance and memory speed and Von Neumann bottlenecks). Intel® architecture with many cores will also mitigate the impact of RC delays.[10]

Intel's CMP architectures provide a way to not only dramatically scale performance, but also to do so *while* minimizing power consumption and heat dissipation. Rather than relying on one big, power-hungry, heat-producing core, Intel's CMP chips need activate only those cores needed for a given function, while idle cores are powered down. This fine-grained control over processing resources enables the chip to use only as much power as is needed at any time.

Intel's CMP architectures will also provide the essential special-purpose performance and adaptability that future platforms will require. In addition to general-purpose cores, Intel's chips will include specialized cores for various classes of computation,

such as graphics, speech recognition algorithms and communication protocol processing. Moreover, Intel will design processors that allow dynamic reconfiguration of the cores, interconnects and caches to meet diverse and changing requirements. Such reconfiguration might be performed by the chip manufacturer, to repurpose the same silicon for different markets; by the OEM, to tailor the processor to different kinds of systems; or in the field at runtime, to support changing workload requirements on the fly. Intel® IXP processors today provide some of these capabilities for special purpose network processing. As shown in Figure 3, IXP 2800 has 16 independent micro engines operating at 1.4 GHz along with an IntelXScale® core. Another related Intel research area is focused on development of a reconfigurable radio architecture, enabling processors to dynamically adapt to different wireless networking environments (such as 802.11b, 802.11a, and W-CDMA).

2. Special Purpose Hardware

Over time, important functions once relegated to software and specialized chips are typically absorbed into the microprocessor itself. Intel has been at the forefront of this effort which has been the driving force behind our business model for over 35 years. By moving functions on chip, such capabilities benefit from more efficient execution and superior economies of scale and reduce the power consumption drastically. Low latency
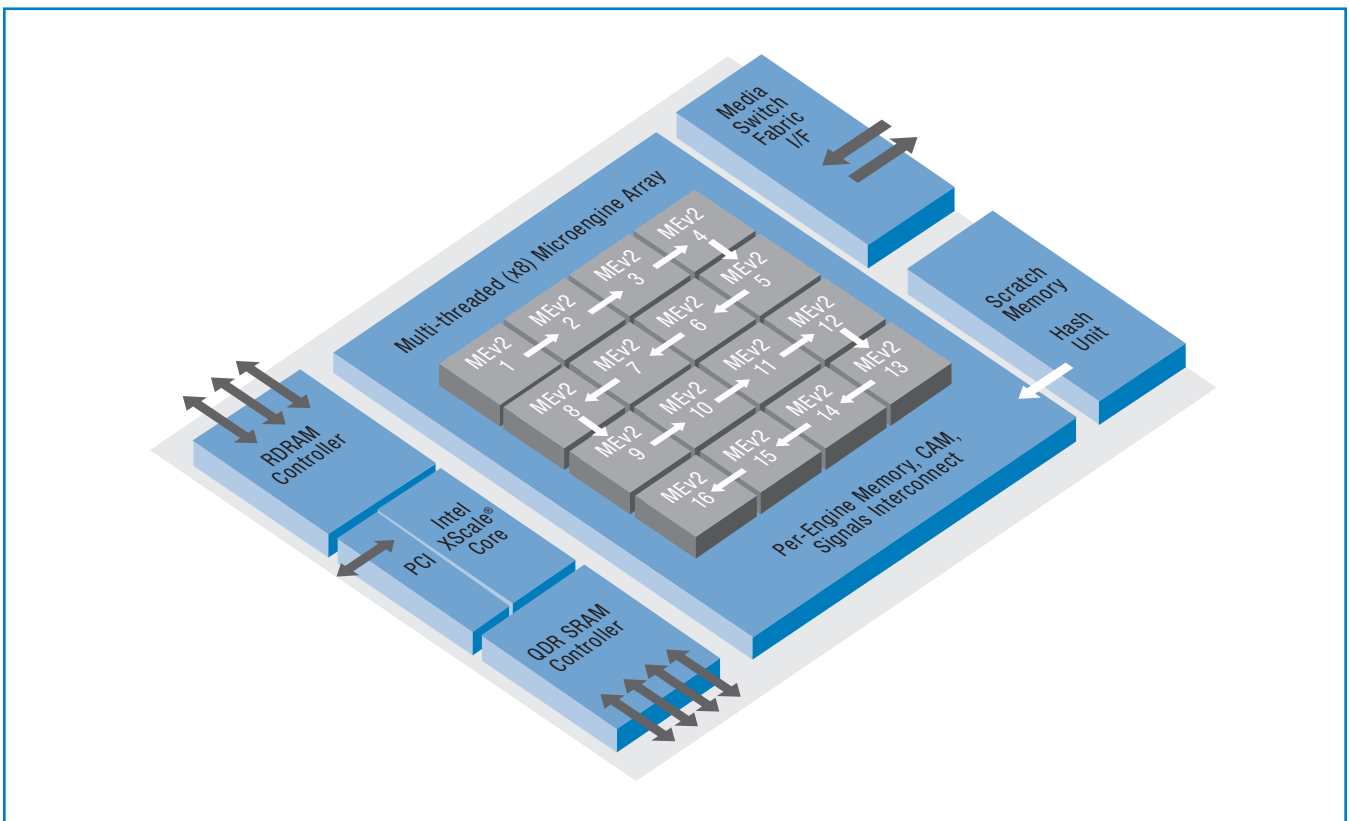

Figure 3: Block level diagram of Intel® IXP 2800 with 16 independent micro engines and one Intel XScale® core

communication between special purpose hardware and general purpose cores will be especially critical to meet future processor architecture performance and functionality expectations.

Special purpose hardware is an important ingredient of Intel's future processor and platform architectures. Past examples include floating point math, graphics processing and network packet processing. Over the next several years, Intel processors will incorporate dedicated hardware for a wide variety of tasks. Possible candidates include: critical function blocks of radios for wireless networking; 3D graphics rendering; digital signal processing; advanced image processing; speech and handwriting recognition; advanced security, reliability and management; XML, and other Internet protocol processing; data mining; and natural language processing.

3. Large Memory Subsystems

As processors themselves move up the performance curve, memory access can become a main bottleneck. In order to keep many high-performing cores fed with large amounts of data, it is important to have a large quantity of memory on-chip and close to the cores. As we evolve our processors and platforms towards 2015, some Intel® microprocessors will include on-chip memory subsystems. These may be in the gigabyte size range, replacing main memory in many types of computing devices. Caches will be reconfigurable, allowing portions of the caches to be dynamically allocated to various cores.

Some caches may be dedicated to specific cores, shared by groups of cores or shared globally by all cores, depending on the application needs. This flexible reconfigurability is needed to prevent the caches themselves from becoming a performance bottleneck, as multiple cores contend for cache access.

4. Microkernel

Microprocessors will need a sizable dose of integrated intelligence to coordinate all this complexity: assigning tasks to cores, powering up and powering down cores as needed, reconfiguring cores to suit changing workloads, and so on. In Intel's CMP architectures, with their high capacity for parallel execution, the processor itself will have to perform some amount of invisible user-level threading breaking applications into multiple threads that can be processed simultaneously. One way to efficiently handle all this is through a built-in microkernel, relieving higher-level software of these complicated hardware management tasks.

5. Virtualization

Future microprocessors will need several levels of virtualization. For example, as shown in Figure 4, virtualization is needed to hide the complexity of the hardware from the overlying software. The OS, kernel and software should not have to deal with the intricacies of many cores, specialized execution hardware, multiple caches, reconfiguration and so on. Rather, it should see
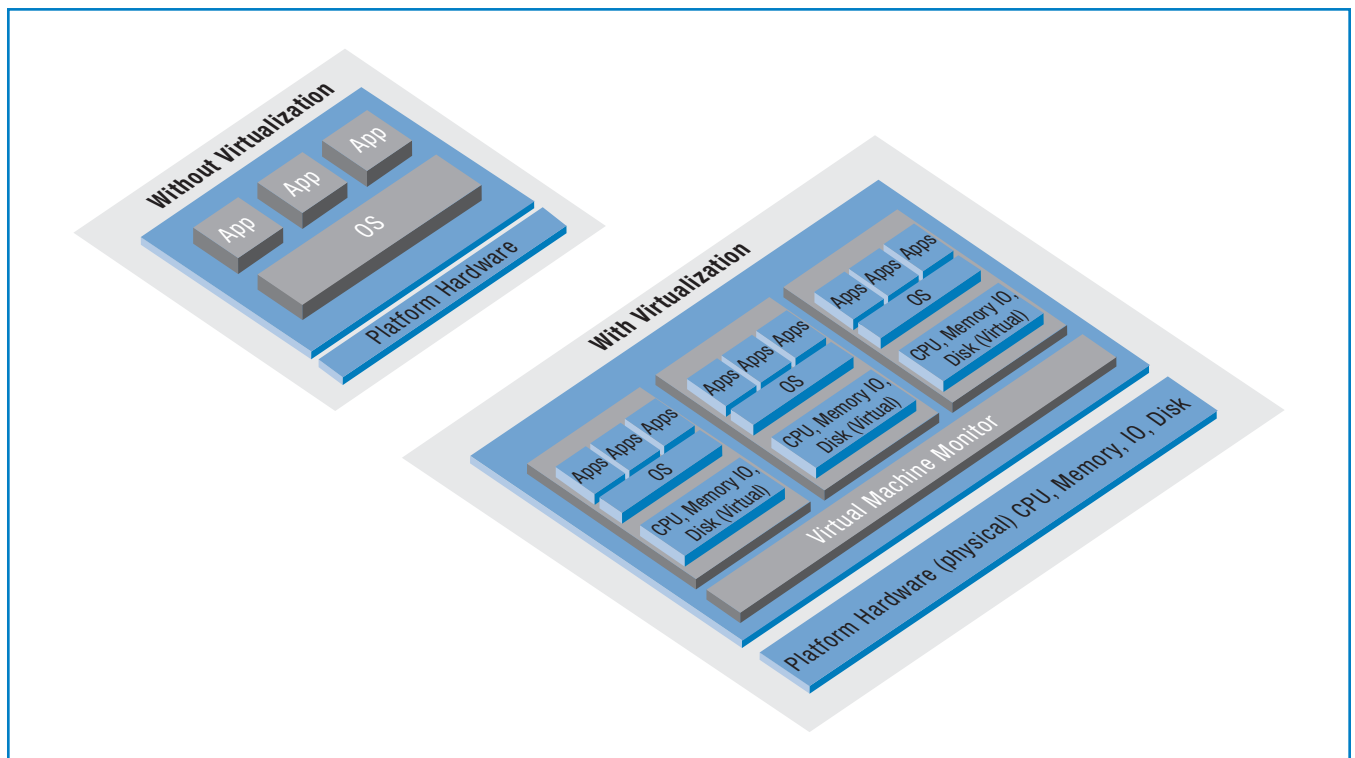

Figure 4: Intel virtualization scheme providing an abstraction layer to hide hardware complexity from software

A Glimpse of Evolving Processor Architecture

How all this might come together in a hypothetical future Intel processor architecture is what we call "Micro 2015" for short. This hypothetical Micro 2015 processor features include:

- Tens of billions of transistors in a single chip (as predicted by Moore's Law).

- Reliable and reconfigurable circuit blocks with a built-in management infrastructure.

- Parallelism at all levels that will be handled through an abundant number of software and hardware threads. Chip Level Multi-processing (CMP) will provide true parallelism with multiple low power IA cores in a reconfigurable architecture with a built-in microkernel.

- Special-purpose, low-power hardware engines for fixed functions including, but not limited to, real-time signal processing and graphics.

- A relatively large high-speed, global, reconfigurable on-chip memory. The memory will be shared by groups of cores, the OS, the micro-kernel and the special-purpose hardware. Parts of the memory may be dedicated to special-purpose cores or applications. We can expect gigabytes of such global on-die memory in the processor coupled with other shared cache resources complementing the main memory.

- High-speed interconnects linking cores within groups and among groups, as well as special-purpose hardware and memory. The memory interconnect bandwidth will match the performance requirements of the processor and be in the multiple gigabytes/second range.

- Built-in virtualization and trust mechanisms providing layers of abstraction to the applications and the OS to meet security, reliability and manageability requirements.

- Compatibility with existing software while providing teraflops of supercomputer-like performance and new capabilities for new applications, workloads and usage models.

This is just one example of many possible architectural scenarios since Micro 2015 is a composite of many capabilities that may or may not be incorporated into Intel's roadmaps based on current and future trends and requirements, and technological constraints, Nonetheless, we believe it fairly represents the overall shape of things to come.

---

the processor as one or more unified virtual machines through its global interfaces virtualization provides the necessary abstraction. Virtualization will also be used to ensure manageability, reliability and security. For example, the processor can be partitioned into multiple virtual processors, some dedicated specifically to management and security tasks and some to manage applications. Some of these features are already in Intel's roadmap as disclosed recently.[11]

6. Silicon and Process Technology

Silicon CMOS process scaling is expected to continue at its current pace at least through 2015 and beyond. The ongoing trend of introducing new materials and new structures will continue. Examples under development are high-k/metal gate dielectrics and tri-gate transistors. Farther out, III-V transistors, carbon nanotubes and silicon nanowires are being investigated. All of these have a goal of continuing to improve device speed, maintaining or reducing power consumption, and further dimensional scaling. In addition, the integration between chip architecture and process technology will become ever tighter as we pack billions of transistors onto a single die. This is very critical – the architects and the process technologists will need to work even more closely in developing the future microprocessor and platform architectures.

7. Compatibility and Ecosystem Enabling

Intel's commitment to maintaining compatibility with existing and older IA software even as we advance the performance and capabilities of our processor and platforms is well proven. Intel will deliver future processors and platforms with teraflops of supercomputer-like performance for new applications and work-loads without sacrificing compatibility with the installed base of software and the vast developer ecosystem around it.

Meeting the Challenges

What will it take to realize this processor vision for 2015? Some major challenges loom, in both software and hardware. It should be mentioned that raw transistor count is not, in itself, likely to be a major hurdle for Intel. No other company has proven better at delivering to the goals of Moore's Law and we can confidently predict that Intel processors will pack 20 to 30 billion transistors on a 1-inch square die: enough to support the many cores, large caches and other hardware described above over next the next years. There are, however, some other challenges.

1. Power and Thermal Management

Currently, every one percent improvement in processor performance brings a three percent increase in power consumption. This is because, as transistors shrink and more are packed into smaller space, and as clock frequencies increase, the leakage current likewise increases, driving up heat and power inefficiency. If transistor density continues to increase at present rates without improvements in power management, by 2015 microprocessors will consume tens of thousands of watts per square centimeter.

To meet future requirements, we must cut the power density ratio dramatically. A number of techniques hold promise. As explained earlier, Intel CMP designs with tens or even hundreds of small, low-power cores, coupled with power-management intelligence, will be able to significantly reduce wasted power by allowing the processor to use only those resources that are needed at any time. In addition, Intel CMP designs will support ultra-high performance without ultra-high clock speeds, thus mitigating some of the current leakage problems that increase with frequency. Further, Intel CMP designs will also exploit variability in transistor speed made possible by future high-density process technologies. Slow and fast transistors can utilize different supply voltages, with time-critical tasks assigned to fast, high-power cores and other tasks to slower, low-power cores. The ultimate goal is to build fully power-aware architectures that can automatically reconfigure to meet power and workload requirements.

Additionally, a variety of techniques at the circuit level – including body bias, stack effect and sleep transistors – can be used to control current leakage. Increasing on-die memory (through large caches) not only increases performance but, by reducing off-chip memory accesses, also slashes power consumption and power density. Specialized hardware such as TCP/IP processing engines can also reduce power consumption because they can execute these functions more efficiently (with less circuitry and fewer clock cycles) than general-purpose hardware.

2. Parallelism

Taking advantage of Intel's future CMP architectures requires that tasks be highly parallelized – i.e., decomposed into subtasks that can be processed concurrently on multiple cores. Today's single- and multi-core processors are able to handle at most a few simultaneous threads. Future Intel CMP processors will be able to handle many threads – hundreds or even thousands in some cases. Some workloads can be parallelized to this degree fairly easily by developers, with some help from compilers, such that the processor and microkernel can support the necessary threading. In image processing, for instance, the image can be subdivided into many separate areas, which can be manipulated independently and concurrent. Around 10-20 percent of prospective workloads fit into this category. A second group of workloads – around 60 percent – can be parallelized with some effort. These include some database applications, data mining, synthesizing, text and voice processing. A third group consists of workloads that are very hard to be parallelized: linear tasks in which each stage depends on the previous stage.

Category 1 represents the early workloads on the parallelism roadmap. The real challenge lies in category 2, which represents the bulk of applications. Parallelizing these tasks requires human intervention and a number of software technologies, currently the focus of multiple industry efforts. These technologies include:

- Arrays of adaptive software libraries and algorithms to handle parallelism seamlessly including sophisticated tools to generate these libraries

- Extensions to existing programming languages, and in some cases a new class of programming languages, to help enable parallelism

- A sophisticated threading model that is compatible with compilers, the operating system and user-level applications and exploits multiple cores

- A set of program expression, performance, and verification tools that advance the parallel engineering process

- Innovative new data representations that support massive multithreading

Intel and its partners are applying an enormous amount of research in this area and we are confident in our ability to develop the software technology that will take advantage of our evolved architectures.[12]

3. Complexity Management

Future Intel CMP architectures represent a big leap in operational complexity. Consider what's needed to coordinate the large number of concurrent processing activities, assign threads to cores, manage power, reconfigure the hardware as necessary on the fly and more – all dynamically at high speed. Efficient operation requires that the processor hardware take on more of these management tasks, rather than leaving them to the operating system and managed runtime software (such as the Java* virtual machine).[11] In other words, the processor itself must have the built-in intelligence to manage and make the best use the underlying hardware.

Such capabilities will be achieved through an intelligent micro-kernel. The microkernel will work with the OS to schedule threads and assign them to cores, turn off cores when idle to conserve power, monitor operations and shift jobs among cores if any are in danger of failing, manage special-purpose hardware and reconfigure cores, caches and interconnects as needed. The microkernel, in turn, will present a unified interface to OS and application software. Operating systems and applications will be able to use the functionality of the microarchitecture, specifying parameters such as power usage or performance requirements, and let the system decide how best to reach those goals using the processing capabilities at hand. Another benefit of this abstraction is that it will allow applications to be ported easily from one class of computing devices to another.

4. Security and Manageability

Intel's microprocessor evolution over the next 10 years will address the need for new, more robust mechanisms to ensure security and manageability. While such functions are handled today by software and human beings, future systems will need more advanced security and manageability built in. New firmware and hardware mechanisms are being developed for Intel® platforms. For example, virtualization can be used to create special partitions for security and management functions, and to implement layers of protection[2] that prevent tampering, intrusions and virus attacks[11]. Hardware reconfigurability will allow such virtualization to be implemented by the chip manufacturer, the device manufacturer and/or during runtime.

5. Variability and Reliable Computing

This is likely to be a major focus of research as we approach the end of the decade. As future transistor sizes drop to 20 nm and below, we are likely to see increasing variability in the behavior of these transistors. Intel is exploring several new mechanisms needed to compensate for this underlying variability in transistor behavior:

- Hardware-based self-monitoring and self-management (which can, for example, detect when a core or a circuit is likely to fail, and preemptively shift the work to another core or circuit). Over the coming years, at least 5 to 10 percent of a processor's 20–30 billion-plus transistors will be used for circuits and logic dedicated to ensuring reliability

- Firmware-based fault prediction based on error history

- Statistical computing techniques, which use probabilistic models to derive reliable results from unreliable components

6. High-speed Interconnects

Intel's CMP architectures will circumvent the bottlenecks and inefficiencies common in other architectures, but may encounter new performance challenges. Chief among these is the communication overhead of shuttling data among the numerous cores, their caches and other processing elements. High-speed interconnects are therefore needed to move data rapidly and keep the processing from bogging down. Intel's approaches include improved copper interconnects, and ultimately, optical interconnects (which move data at light speed). The challenge lies not only in the interconnect material, but also the interconnect architecture. Ring-type architectures are being applied successfully in designs of up to 8 to 16 cores. Beyond that, new interconnect architectures capable of supporting hundreds of cores will be needed. Such mechanisms will have to be reconfigurable, to handle a variety of changing processing requirements and core configurations. Interconnect architecture is an area of active and extensive research at Intel, at universities and elsewhere in the technology industry.

Conclusion

This paper has outlined an ambitious vision which, in some respects, is a wide departure from present-day processors and platforms. But in reality, this vision is based on a continued evolution of Intel's drive for increased parallelism, and our proven investment, research, development, manufacturing, and unparalleled ecosystem enabling capability that, when taken together, will continue to lead us into an era of more powerful versatile and efficient processing engines and platforms containing those engines. Ultimately, this evolution is driven by usage what people want from technology and what they do with it. And though no one can precisely predict the future course of technology, the developments now underway point to some likely outcomes. Based on current requirements and trends, we at Intel believe that processor and platform architecture needs to move towards a virtualized, reconfigurable CMP architecture with a large number of cores, a rich set of built-in processing capabilities, large on-chip memory subsystem and sophisticated microkernel. This architectural evolution, delivered with volume computing economics and an adherence to maintaining compatibility with thousands of already existing applications, will ensure that Intel processors and platforms will continue to power a breathtaking array of sophisticated new applications over the coming years transforming business and daily life in ways we can only begin to imagine.

More Info

For more information, go to
www.intel.com/go/platform2015

References

1. *Machine Learning/Vision White Paper,* Intel R & D, Intel Corporation, 2004.
   www.intel.com/research/machine_learning.htm

2. *"How Much Information," 2003.* Lyman, Peter and Hal R. Varian.
   www.sims.berkeley.edu/how-much-info-2003

3. *A Platform 2015 Workload Model: Recognition, Mining and Synthesis
   Moves Computers to the Era of Tera,* Pradeep Dubey, Intel Corporation, 2005.
   www.intel.com/technology/magazine/computing/recognition-mining-synthesis-0205.pdf

4. *The New Net,* Patrick Gelsinger, Intel Corporation, 2004.
   www.intel.com/technology/magazine/research/rs09041.pdf

5. *Proactive Computing,* Intel Corporation, 2002
   www.intel.com/research/documents/proactivepdf.pdf

6. *Relaxing Constraints – A Model for Architectural Innovation,*
   Joel Emer, Intel Corporation, 2004.
   www.intel.com/technology/computing/mi12021.htm

7. *Dynamic Sleep Transistor and Body Bios for Active Leakage Control for Microprocessors,*
   Shekhar Borker et al, Intel Corporation, 2004.

8. *Intel's T's Deliver New Platform Enhancements Beyond Gigahertz,*
   R M Ramanathan, Intel Corporation, 2004.
   www.intel.com/technology/magazine/computing/it12041.pdf

9. *Mitigating Amdahl's Law Through EPI Throttling,*
   Annavaram, Murali, Grochowski, Ed, Shen, John, Intel Corporation, 2005.

10. *Architecting the Era of Tera,* Intel Corporation, 2004.
    www.intel.com/technology/computing/archinnov/teraera/index.htm

11. *Virtualization – Bringing Flexibility and New Capabilities to Computing Platforms,*
    R M Ramanathan, Francis Bruening, Intel Corporation, 2004.
    **ftp://download.intel.com/technology/computing/archinnov/
    teraera/download/Virtualization_0604.pdf**

12. *Platform2015 Software: Enabling Innovation in Parallelism for the Next Decade,*
    David Kuck, Intel Corporation, 2005.

intel®

0305/VET/HBD/PDF