Power-Management Architecture of the Intel Microarchitecture Code-Named Sandy Bridge

Modern microprocessors are evolving into system-on-a-chip designs with high integration levels, catering to ever-shrinking form factors. Portability without compromising performance is a driving market need. An architectural approach that's adaptive to and cognizant of workload behavior and platform physical constraints is indispensable to meeting these performance and efficiency goals. This article describes power-management innovations introduced on Intel's Sandy Bridge microprocessor.

Efraim Rotem
Alon Naveh
Doron Rajwan
Avinash Ananthakrishnan
Eliezer Weissmann
Intel

Continuous advances in process technology let designers integrate an ever-increasing number of transistors onto a single die. This increased transistor density has enabled system-on-a-chip (SoC) functionality by integrating graphics engines, memory controllers, and other platform components into modern CPU dies. The Sandy Bridge client CPU contains just over 1 billion transistors, and the Sandy Bridge server contains up to 3 times as many on a single monolithic die.

The second-generation Sandy Bridge core processor has five main domains: Intel architecture cores, a ring interconnect, a shared last-level cache (LLC), a system agent, and processor graphics. Figure 1 describes the Sandy Bridge die.

Sandy Bridge redesign

We significantly redesigned Sandy Bridge from the previous generation, providing increased instruction-level parallelism and multithreading capabilities and improved power management and energy efficiency. We also introduced several architectural changes, most notably the 256-bit Advanced Vector Extension (AVX).

In addition, we moved the processor graphics from the chip set or discrete graphics onto the lead CPU process technology, resulting in much higher transistor counts, lower power, and higher frequencies. We also significantly improved the internal microarchitecture of the processor graphics and media, resulting in much higher-performance execution units and high-performance media functionality.

The ring interconnect and LLC architecture provide high modularity and feature-rich general-purpose, graphics, media, and system integration.

The transistor count increase and integration of platform components into a monolithic die, together with the increase in core frequency, introduce demanding power and

Published by the IEEE Computer Society

energy challenges to modern computers. Because computer systems' power and thermal envelopes aren't increasing, addressing these challenges requires a highly energy-efficient design. Sandy Bridge power management builds on the existing SpeedStep and Turbo Boost technologies significantly increasing performance and energy efficiency. Sandy Bridge's power-management features provide the maximum performance possible within the package and system physical constraints when needed, while consuming very low power and energy when full performance isn't needed. Furthermore, we expanded the power-management features implemented in previous generations of Intel CPUs on Sandy Bridge to give the rest of the SoC power budgeting and prioritization capabilities.

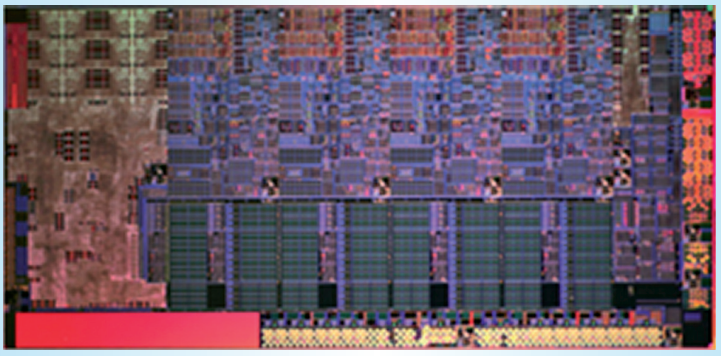

Figure 1. The Sandy Bridge die photo.

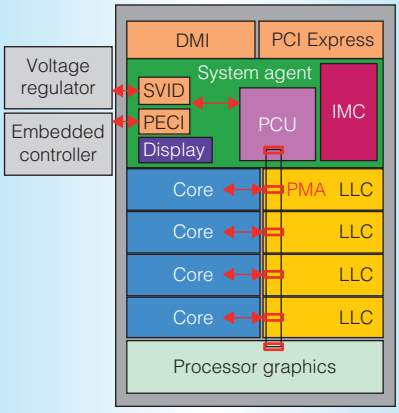

Figure 2. Sandy Bridge's power-management architecture. Sandy Bridge block diagram showing the major functional blocks and the power-management control blocks and interconnect. The platform power-management components and the associated platform interconnect are also shown.

Power-management architecture

The Package Control Unit (PCU) is the brain behind the Sandy Bridge power-management features. The power-management architecture is highlighted over the block diagram in Figure 2. The PCU resides in the system agent and is a combination of dedicated hardware state machines and an integrated microcontroller. A power-management link connects the PCU to different cores and functional blocks on the die via power-management agents (PMAs). PMAs collect telemetry information such as power consumption and junction temperature, and perform control functions such as P-state and C-state transitions. The PCU communicates to the external voltage regulator and embedded controller that perform system power-management functions. The PCU runs firmware that constantly collects power and thermal information, communicates with the system, and performs various power-management functions and optimization algorithms.

Sandy Bridge's package implements two independent variable power planes. One shared power plane feeds all CPU cores, the ring, and the LLC. Embedded power gates turn each core on and off individually. The LLC's power gates can turn on or off portions of the cache in shallow package sleep states or all of the cache in deeper sleep states. All the cores and the ring share the same clock and perform dynamic voltage and frequency scaling together. The graphics processor has an independent power plane, whose voltage and frequency can be varied independently. It can also be turned off completely when the graphics are inactive. Additional fixed power planes control the system agent and I/O.

The Sandy Bridge power management maximizes the user experience under multiple constraints. The user experience has the following attributes:

- throughput performance,
- responsiveness (burst performance),

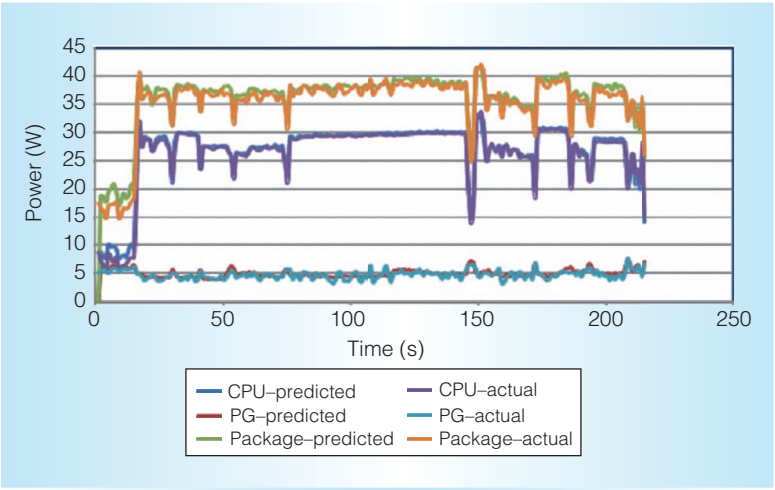

Figure 3. Power meter: predicted and actual power of the CPU, processor graphics, and total package. The figure shows a power snapshot of combined CPU and graphics workload. The chart presents the actual measured power and the architectural power meter reporting for the IA core, processor graphics, and total package. The actual and reported power correlate accurately.

- CPU and graphics performance,
- battery life and energy bills, and
- ergonomics (acoustic noise, heat, and so on).

To meet user preferences, the power-management algorithms optimize around the following physical constraints:

- silicon capabilities, including voltage, frequency and power characteristics;
- system thermomechanical capabilities;
- power-delivery capabilities;
- software and operating system explicit control; and
- workload and usage characteristics.

The system designer can control the power-management functionality's behavior and preferences via basic input/output system (BIOS) settings, runtime software, or an on-board embedded controller. At runtime, the system reads and controls parameters such as power, maximum current consumption, and die temperature.

Intel Turbo Boost technology 2.0

The power and frequency of the CPU and processor graphics are defined by a scenario of concurrent CPU and processor graphics running a heavy workload at the same time at worst-case conditions.[1] In most cases, the CPU is running a less-demanding application and the Intel Turbo Boost technology uses this power headroom to extract higher performance when possible.[2,3] Sandy Bridge's power performance control is performed primarily through dynamic voltage and frequency scaling (DVFS). When the operating system identifies a need for high performance, it issues a high P-state request. Whenever power and thermal headroom exist, the PCU increases the voltage and frequency to the highest point that is lower than or equal to the operating system request, that still meets all physical constraints. Sandy Bridge implements architectural power meters. It collects a set of architectural events from each Intel architecture core, the processor graphics, and I/O, and combines them with energy weights to predict the package's active power consumption. Leakage information is coded into the die and is scaled with operating conditions such as voltage and temperature to provide the package's total power consumption. The system uses architectural power predictor output, which is also exposed externally to software, to decide the amount of turbo upside available for the current workload (turbo upside is available higher frequency that the CPU can go up to and use for higher performance).[4,5] The architectural power predictor provides a consistent turbo behavior while minimizing the die-to-die variations and dependency on ambient temperature. Figure 3 describes the actual versus predicted power of the CPU, processor graphics, and total package.

In addition to the Intel Turbo Boost technology already implemented in previous generations of Intel processors, Sandy Bridge offers two new functionalities: total package power control and responsiveness via dynamic turbo. Sandy Bridge is an SoC monolithic die. The power is specified in terms of the entire package's total power consumption. The real workload uses the die's different computational and communication resources. The PCU continuously monitors the individual functional blocks' power consumption and performs dynamic budget allocation to the various components. One such example is power sharing between the

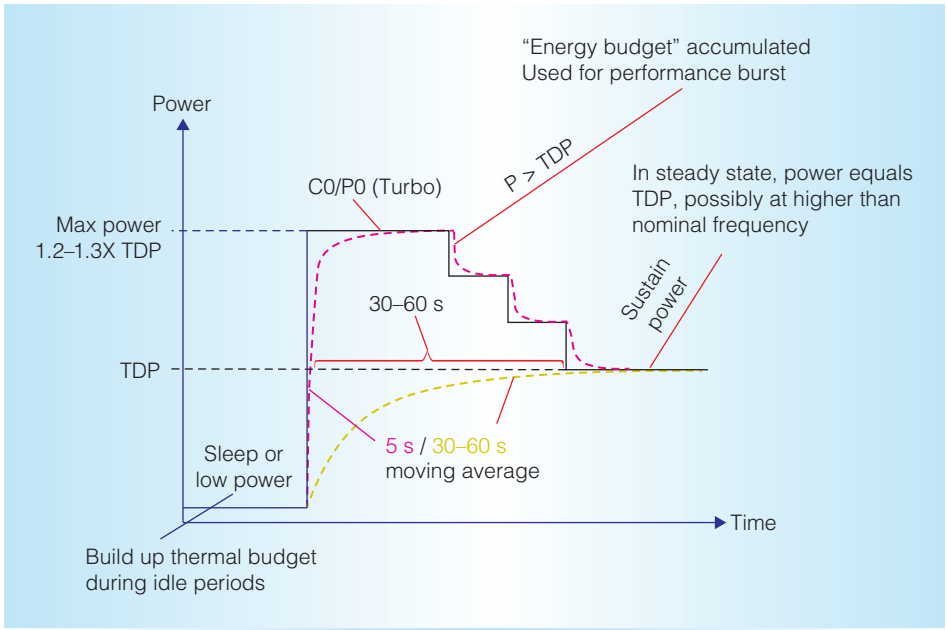

Figure 4. Dynamic behavior of the Intel Turbo Boost. After a period of low power consumption, the CPU and graphics can burst to very high power and performance for 30 to 60 seconds, delivering a responsive user experience. After this period, the power stabilizes back to the rated TDP.

CPU and the processor graphics. In workloads that heavily use the CPU and don't perform graphics operations, the PCU shifts the power budget to the CPU, and vice versa. For balanced workloads, the system designer can use a preference knob provided via BIOS or the runtime driver to set the desired processor graphics and CPU preference policy.

We defined the package's power specifications based on steady-state conditions. A typical heat sink has a thermal mass, and its heat capacity can sustain short periods of high power before it reaches steady-state temperatures. We use this phenomenon in Sandy Bridge to deliver responsiveness on interactive workloads. The PCU uses an exponential weight-moving average (EWMA) algorithm over multiple seconds to control the power. This results in short periods of time that the CPU exceeds its rated power as long as the rolling average is within the platform's thermal specifications. The averaging time window is a function of the system's thermomechanical design, and therefore the system designer can configure it. Figure 4 describes the new Intel Turbo Boost technology behavior. We achieve a boost in responsiveness by managing energy budgets. The EWMA algorithm tracks energy consumption over a predefined time window; during periods in which the CPU is idle or consuming less than the thermal design point (TDP) power, it accumulates an energy credit. The PCU then uses these energy credits to burst above TDP for short periods until the energy credit is fully consumed. This provides responsiveness and an instantaneous performance boost for the user.

Sandy Bridge's turbo algorithm differs significantly from those of prior-generation CPUs[3] in that the amount of turbo upside is a function of the accrued energy credits over a thermally significant time period. This lets Sandy Bridge turbo up in frequency during periods of high activity while still meeting the long-term system power and thermal constraints. Figure 5 illustrates measured gain in overall performance between the similarly configured four-core Sandy Bridge platform and the Clarksfield platform. The IPC gains between Sandy Bridge and Clarksfield average between 10 and 15 percent. Sandy Bridge's additional performance gain

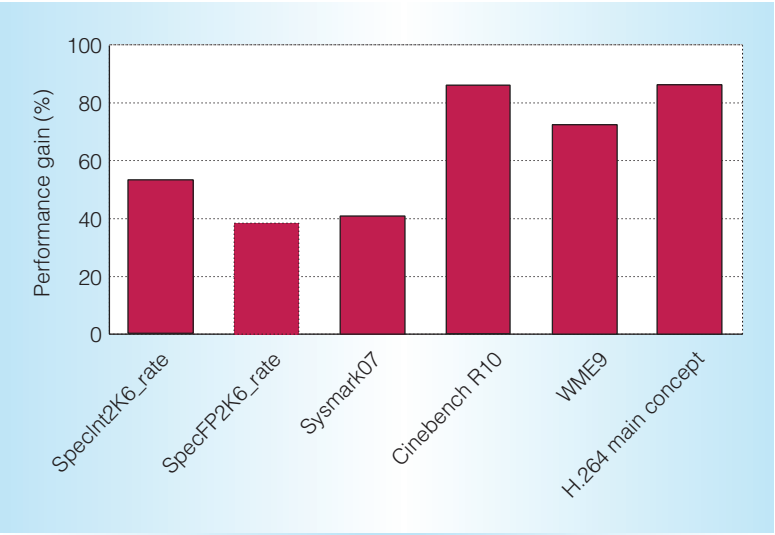

Figure 5. Sandy Bridge performance gain over Clarksfield. The figure shows Sandy Bridge's measured benchmarks performance compared to Clarksfield with Turbo enabled. Up to 88 percent performance gains can be observed. Most of the performance gain comes from improvements in the Turbo algorithms and power-management features.

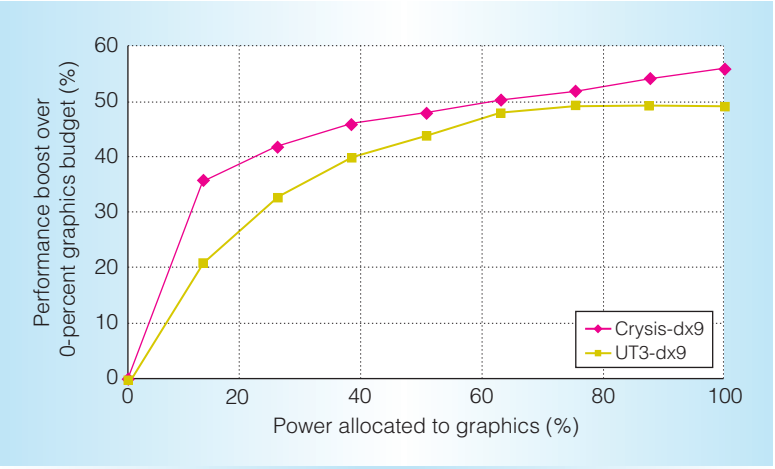

Figure 6. Performance versus power balancing. This figure shows the performance of two graphics-intensive games as a function of power budget assigned to the graphics processor. The baseline is all power budget assigned to the IA core and no power budget to the graphics. The more power allocated to the graphics processor, the higher the performance. There is a diminishing return on the power allocation at the high ratios because this budget is taken from the IA core, which needs to feed the graphics processor.

comes from intelligent power management and turbo algorithms.

As Figure 5 shows, Intel's Turbo Boost 2.0 technology provides a significant performance boost on short encoding, video editing workloads by allowing the CPU to burst above TDP for 20 to 30 seconds at a time before settling back to run within the rated TDP envelope.

Graphics and core integration

Sandy Bridge is the first Intel microprocessor to fully integrate a graphics engine and a traditional Intel architecture processing core on a single die. As Figure 2 shows, the graphics and Intel architecture cores share the LLC, interconnect resources, and the integrated memory controller present on the die. This integration poses interesting opportunities from both a power-management and performance-optimization viewpoint.

Because the processor graphics and the Intel architecture cores are integrated on the die, they share a common power and thermal envelope. This implies that running the cores at a higher frequency and hence at a higher power level takes away power headroom that we could otherwise allocate to the processor graphics to boost its frequency, and vice versa. The EWMA algorithm implemented in the PCU tracks the energy consumed over a certain configurable time window and the headroom available to the entire die at a given time. On workloads that keep predominantly only one of the two compute engines busy (that is, either the Intel architecture cores or the graphics core), how to allocate the available headroom to maximize performance is a trivial decision. Real-world applications, especially high-performance gaming workloads, use the Intel architecture cores and the processor graphics simultaneously, and the usage of each component varies dynamically over the course of the workload.

The PCU on Sandy Bridge exposes a software interface through which an operating system or graphics driver can choose how to partition the available energy headroom between the Intel architecture cores and processor graphics.[5] Figure 6 illustrates the sensitivity of performance to how power is allocated to processor graphics on two DX9 gaming workloads.

Software performs this repartitioning dynamically, accounting for workload-specific

characteristics, and hence further improving performance under a power-constrained envelope. It repartitions unused energy budget again such that it's not being lost.

In addition to power and thermal dependency between the CPU and the processor graphics is a functional dependency. The CPU and processor graphics share the same LLC and ring interconnect. As a result, graphics performance strongly depends on how much interconnect and cache bandwidth is made available to the graphics engine. Furthermore, Intel architecture cores and the ring interconnect are part of a single voltage and frequency domain. This implies that the cores, when active, will run at the same frequency as the ring. Sandy Bridge exposes a set of performance counters for the software and the graphics driver to help select the optimal CPU and interconnect frequency at graphics workloads.

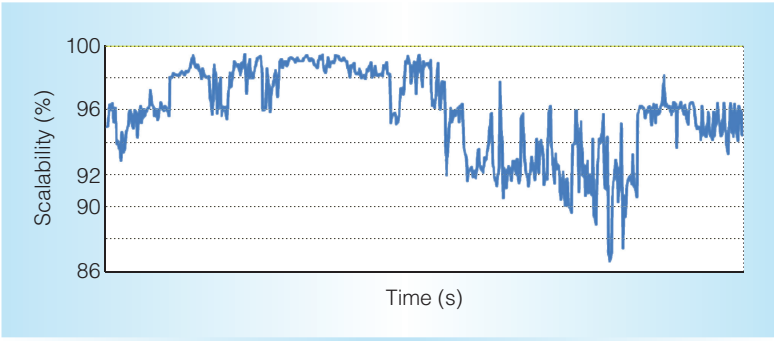

Figure 7. CPU and memory-bound profile. The figure shows scalability snapshot as a function of time, from the SPEC benchmark. Scalability of 100 percent means that the IA core performance gains performance proportional to the frequency increase. Lower values imply that there are memory accesses and increasing the IA core frequency will only partly improve the overall benchmark runtime.

Energy-aware CPU

We designed Sandy Bridge to deliver high performance when needed while consuming minimum active and idle energy when the CPU or processor graphics aren't active. Running at the highest possible frequency delivers the best performance but isn't energy efficient because of the cube dependency of power in frequency and voltage. Some memory-bound workloads that run at a high frequency consume high power but don't fully gain the performance from the increased frequency. If a user sets a balanced power preference, Sandy Bridge activates an energy-aware turbo algorithm that profiles the memory access pattern at runtime and performs a less-aggressive turbo during memory-bound intervals while increasing the frequency during CPU-bound phases. Figure 7 demonstrates a real-time application capture, showing the memory access patterns and predicted performance gain from higher frequency. At low-scalability periods, the PCU performs a DVFS transition to a less-aggressive turbo frequency, gaining power and energy for a small performance loss.

Smart average power support

During low-activity periods, when there are no active jobs pending to be executed, the operating system puts the processor into a sleep state. In ACPI terminology,[6] this sleep state is called the C-state. The operating system controls each core individually, and the PCU coordinates between the cores and threads. Sandy Bridge introduced new PCU-managed C-states (see Figure 8).[2,3] Deeper C-states offer more power savings,

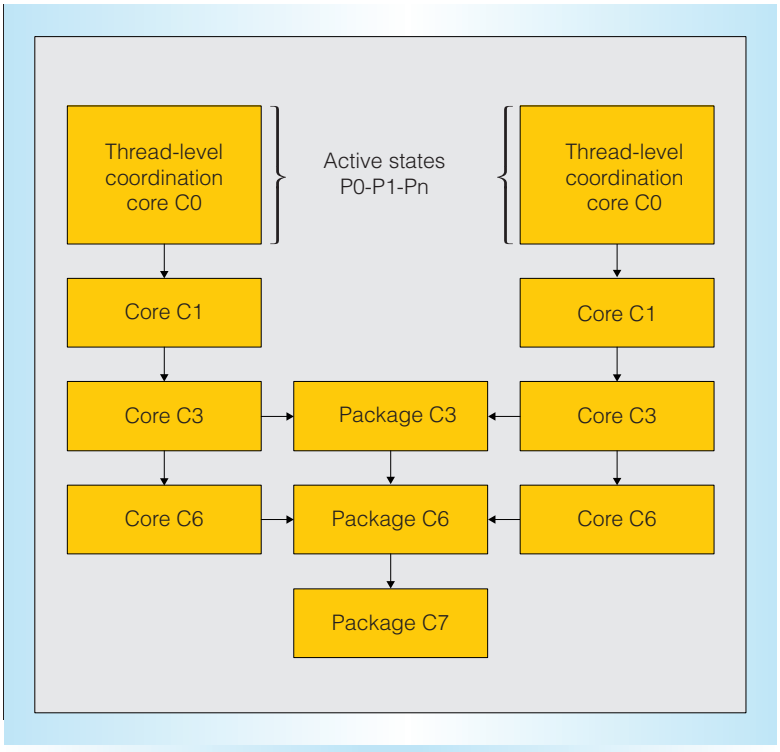

Figure 8. Sandy Bridge package C-state coordination.

Table 1. Demotion and un-demotion algorithm results.

	Demotion algorithm vs. no optimization		Un-demotion benefit over Demotion only	
Platform	**Performance**	**Power**	**Performance**	**Power**
SYSmark 07	+4%	+122 mW	+0.3%	–110 mW
Mobile Mark 07	+35%	–17 mW	–2%	–21 mW

but at the cost of longer latency to enter and exit the C-state. Proper management of the C-states is a fundamental capability for low power consumption and long battery life.

The ability to save power by entering into deeper idle states includes two major issues:

- Entering and exiting a deeper core-level state involves microarchitectural operations and voltage transitions, which result in long latency. For example, exiting a deeper C-state in Sandy Bridge typically takes a few tens of microseconds. During this time, the CPU is not executing useful code. Too-frequent transitions could therefore affect the workload's overall performance.
- The entry and exit idle states also consume energy. Frequent C-state transition might cost more energy than the energy saved while in the deep sleep state.

To deal with these two issues, Sandy Bridge includes a Demotion algorithm from a deeper C-state into less aggressive C-states. Typically, the operating system frequently uses the sleep state in bursts. Sandy Bridge's PCU C-state algorithm performs runtime C-state entry and exit profiling. It then calculates the energy savings achieved by using these C-states. If too-frequent entry and exit transitions to a deep C-state result in net performance or energy losses, the Demotion algorithm will override the operating system request and keep the cores at a higher power state.

During idle periods following these C-state bursts, the CPU will promote the package again to deeper C-states to minimize energy consumption.

Table 1 shows the added value of the C-state promotion/demotion algorithm. The Demotion algorithm improved Mobile Mark 07's performance by 35 percent but had a negative power effect on SYSmark 07.[7] The promotion method gained back this power with minimal performance impact.

System power management

The Sandy Bridge power-management architecture enables a hierarchical control for total system power and energy consumption. A new system serial bus (PECI) connects Sandy Bridge to a system-embedded controller.[8] This power-management bus lets the embedded controller read and manage the die's power, energy, and temperature.

A voltage regulator module (VRM) delivers voltage to the package. Sandy Bridge introduced a serial voltage regulator control called SVID, which lets the PCU control multiple voltage regulators such as the CPU and the processor graphics on die.[9] It also offers configuration capabilities and telemetry features. The PCU directly communicates with the external VRM to change the supply voltage for the various frequencies. Furthermore, the PCU performs fine-grained voltage control to optimize the voltage in order to meet the CPU's exact run conditions (that is, the number of active cores, the die temperature, and the maximum power consumption). The PCU also manages the voltage regulator power states to minimize losses and optimize the total power consumption.

Process technology improvements allow an ever-increasing number of transistors to be integrated into modern microprocessors. The power savings and speed gains resulting from process improvement can't keep up with transistor density. Modern microprocessors such as Sandy Bridge are beginning to resemble SoCs, integrating multiple compute components such as a CPU core with a graphics engine and a memory controller, all within the same CPU die. These platforms face significant

thermal and power delivery challenges. Maximizing performance under these SoC platforms will require architectural power-management techniques to provide improved performance and energy efficiency.

Sandy Bridge's Turbo Boost 2.0 allows the CPU to burst above TDP for a short duration, providing increased performance and responsiveness while remaining compliant to platform electrical and thermal limits. In addition, Sandy Bridge's Dynamic Turbo provides up to 80 percent higher performance over previous-generation CPUs. Finally, Sandy Bridge offers a mechanism that lets software dynamically move power between the traditional Intel architecture and processor graphics computational elements, improving the graphics performance by up to 50 percent.

As the industry moves toward eight hours and longer battery life requirements with no compromise on performance, algorithms such as Sandy Bridge's energy-efficient turbo are the first step toward addressing this need. MICRO

References

1. *Second Generation Intel Core Processor Family Mobile External Design Specification (EDS),* v. 1-2, 2011.
2. S. Gochman et al., ''Introduction to Intel Core Duo Processor Architecture,'' *Intel Technology J.,* vol. 10, no. 2, 2006, pp. 89-97.
3. S. Gunther et al., ''Energy-Efficient Computing: Power Management System on the Nehalem Family of Processors,'' *Intel Technology J.,* vol. 14, no. 3, 2010.
4. *Intel 64 IA-32 Architectures Software Developer's Manual Documentation Changes,* 2011; http://www.intel.com/content/www/us/en/architecture-and-technology/64-ia-32-architectures-software-developers-manual.html.
5. *Intel 64 and IA-32 Architectures Optimization Reference Manual,* 2011; http://www.intel.com/content/www/us/en/architecture-and-technology/64-ia-32-architectures-optimization-manual.html.
6. *Advanced Configuration and Power Interface (ACPI) Specification,* rev. 5.0, Dec. 2011; http://www.acpi.info/spec50.htm.
7. Bapco, *SYSMark 2007;* www.bapco.com.
8. M. Berktold and T. Tian, ''CPU Monitoring with DTS/PECI,'' white paper, Intel, Sept. 2010.
9. Intersil, *Power Management Products—Processor Power,* 2012; http://www.intersil.com/processor_power.

Efraim Rotem is a senior principal engineer in the client microprocessor group at Intel, where he's working on next-generation power-management features. His interests include CPUs, systems on chip, and platform power management. Rotem has MSc in electrical engineering from the Technion, Israeli Institute of Technology.

Alon Naveh is a principal engineer in the client microprocessor group at Intel. His work focuses on processor and platform power management. Naveh has a BS in electrical engineering from the Technion, Israel Institute of Technology, and an MBA from San Jose State University.

Doron Rajwan is a power-management architect in the client microprocessor group at Intel. His work has included power-management algorithms and firmware design. Rajwan has an MSc in electronic engineering from Tel Aviv University.

Avinash Ananthakrishnan is a power-management architect in the client microprocessor group at Intel. His work focuses on P-state optimization, turbo, package power sharing, platform power delivery, and power-performance projections on client CPU platforms. Ananthakrishnan has an MSc in electrical engineering from the University of Michigan.

Eliezer Weissmann is a principal engineer in the software and service group at Intel, where he's working on next-generation power-management features. His interests include the architecture definition of power-management features, optimization of idle and performance states, and software and hardware optimization for CPU and graphics computation. Weissmann has an MSc in computer science from the Technion, Israel Institute of Technology.

Direct questions and comments about this article to Efraim Rotem, Intel Israel (74) Ltd., PO Box 1659, Haifa 31015 Israel; efraim.rotem@intel.com.