Power and thermal constraints of modern system-on-a-chip computer

Efraim Rotem*[1], Ran Ginosar[2], Avi Mendelson[2], Uri C. Weiser[2]

[1] Intel corporation, Haifa Israel
[2] Technion, Israel Institute of Technology

* Corresponding Author: efraim.rotem@intel.com

Abstract

Power and thermal are major constraints for delivering compute performance in high-end CPU and are expected to be so in the future. For high end processors, junction temperature has been considered the toughest physical constraint that needs to be tightly managed. Recent trends in form-factors and the increased focus on thin and light systems such as Ultra Book, tablet computers and smartphones, shift the challenge away from junction temperature. Ergonomic thermal considerations and power delivery are becoming the limiters for delivering high computational performance density and need to be managed and controlled. In this paper we describe the major physical constraints, design considerations and modern power and thermal management techniques and demonstrate them on an Intel Core(tm) i7 system.

1 Instruction

Continuous advances in process technology allows for the integration of an ever increasing number of transistors onto a single die. Moors law is expected to deliver even higher transistor density for the foreseeable future. This increased transistor density enables integrating CPU cores, graphics engines, memory controllers and other platform components into modern SoC (System on a Chip) dies. For the last few process generations however, the process technology do not deliver power and energy improvements. A modern CPU contains over a billion transistors, on a single monolithic die. This increase in transistor count and integration of platform components into a monolithic die, together with the increase in core frequency introduce demanding power and energy challenges. Recent market trends toward smaller, thinner and lighter form factors such as Tablet computers and Ultrabook™ drive the power and thermal envelopes of computer systems further down. More focus is put on the various aspects of user experience, including responsiveness to user interaction and GUI operations, sustained general purpose compute, rich graphics and media content and ergonomic considerations. Most modern computer systems cannot sustain all the system on a chip components operating at their highest power-performance state, all the time. Power management has become the primary mechanism to maximize user experience within multiple system constraints. Power management features are designed to provide the maximum performance that is possible within the package and system physical constrains when needed, while consuming very low power and energy when full performance is not needed. In this paper we will describe the various constraints of a modern system, evaluate power management techniques and their power performance benefits and evaluate the performance gain achieved by managing power and performance within these constraints.

2 System physical constraints

Computer power management attempt to maximize the user experience under multiple system constraints. The user experience may have various attributes:

- Throughput performance – sustained computational for a long period of time, either general purpose compute, graphics and media, audio etc. User may have preference between various computational engines on a single die.
- Responsiveness – burst performance while executing user interactive actions.
- Battery life and energy bills – active and idle energy consumption.
- Ergonomics - acoustic noise, skin and outlet air temperature etc.

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

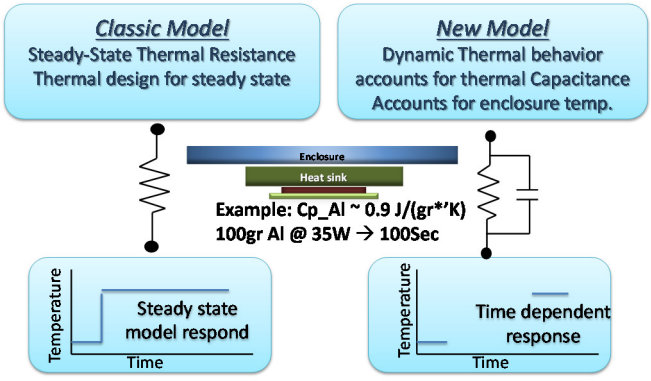
- Silicon capabilities – Voltage and frequency, reliability limitation, current consumption etc.
- System thermo mechanical capabilities – the ability to extract heat from the die junction and the box to the ambient.
- Power-delivery capabilities – Voltage regulator, battery and power supply drive capabilities.
- Software and operating system quality of service requirements

2.1 Thermal limitations

Junction temperature and the ability to cool the die have been considered in the past to be the primary limiter for delivering high performance computation [1]. Fig. 1a describes the classical model that has been considered in computer cooling

systems. CPU thermal design power (TDP) is specified for the worst case application, tested on the worst case manufactured part, measured under the worst case equipment and platform components tolerance at worst case ambient conditions. Computer manufacturers where expected to design the system thermo-mechanical parameters for cooling such worst case conditions for sustained operation. In small form factor such power definition limits the maximum voltage and frequency of the SoC. Most workloads however consume much lower power and can benefit power budget that allows increased frequency and performance [2].

Physical behavior of cooling system is characterized not only by the steady state conductivity but also by heat capacity (Fig 1b.). Typical heat sink can absorb a substantial power surge until the heat sink heats up while keeping the junction temperature within specifications.



1a. Static model *1b. Dynamic model*

Figure 1: Static steady state model and dynamic thermal model

Junction temperature is only one possible thermal limitation. Small form factors such as Smartphones and Tablet computers are sensitive to the enclosure skin temperature – the outer surface temperature that is in contact with the user. High temperature may cause discomfort or even damage to the human skin. A typical acceptable temperature of hand held device is less than 45°C. The combined thermal limitation of a small form factor device is a combination of the two (Fig 2). Y axis is the time it takes the device to reach its constraining thermal limit as a function of total power consumption of the CPU (X axis). The red line describes the case thermal limit. The steady state cooling capability is ~1W (infinite time). It is possible to consume 1.5W for 7200Sec, 4W for 240Sec or 6W for 100Sec before the device heats up to its steady state and the skin reaches its temperature limit. A similar behavior is observed on the die junction temperature (Blue line), but with a much shorter time constants. The junction temperature steady state limit occurs at a power slightly lower than 3W but at these power levels the skin temperature is more constraining. The steady state maximum power would therefore be ~1W with die junction temperature much lower than the specifications limit. It is also possible to burst the CPU as high as 6W for 5 seconds before the junction overheats. This time is much too slow for the skin to experience any meaningful change in temperature.

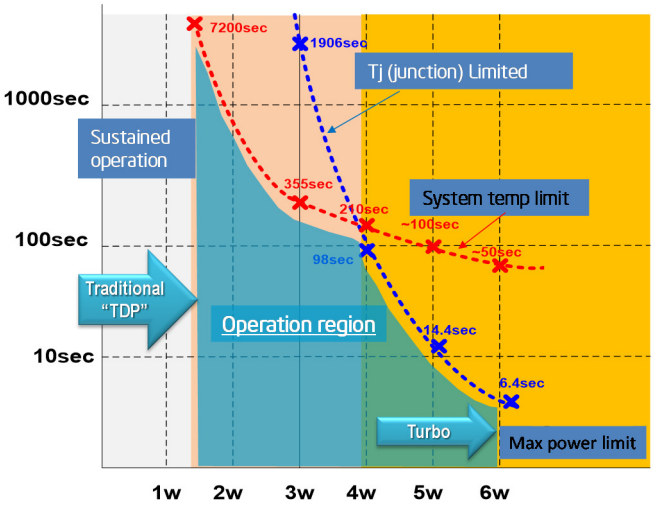


Figure 2: Small form factor dynamic cooling capability profile constrained by the highest of junction temperature or case temp.

Modern processors such as the Intel® Core™ 2 duo [3] make use of this thermal profile. The solid line in Fig.3 describes frequency profile often referred to as Turbo. The operating system tracks and controls the power and performance states of the device [4]. After an idle period, the heat sink cools down. Activation of the device e.g. interactive user activity initiates a burst of high power that is absorbed by the cool heat sink. This enables a responsive behavior to user interactive action that is not possible for long period of time. After a period that is defined by the heat sink thermal constant, the power can stabilize around steady state cooling capability. The red and green dashed lines show conceptual junction and skin temperatures respectively. The junction is heated much faster but the short periods of high power are getting filtered and skin is impacted by energy that is accumulated over long time into the bigger thermal mass.

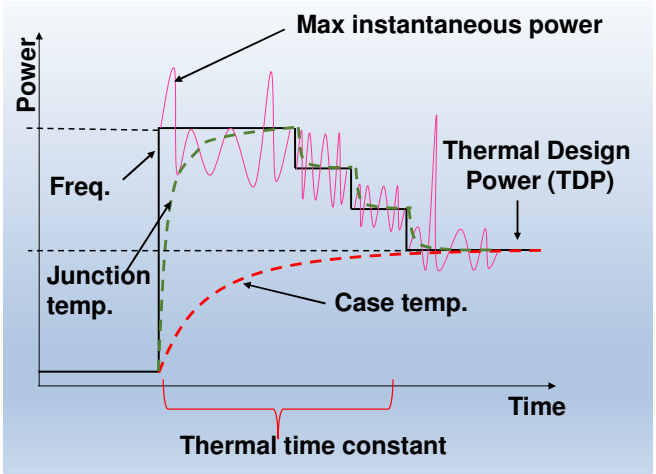


Figure 3: Turbo power profile - After idle period a burst of high power is allowed for thermally significant time, stabilizing to steady state thermal design power.

In this example (Fig 2) the instantaneous burst power can get as high as six times the steady state. Furthermore a modern CPU is characterized by high dynamic power between its normal operation and highest possible load. This dynamic profile is described in red chart of Fig 3.

At steady state, the die runs at lower than its maximum junction temperature. This reduces the reliability stress of the silicon. It does however experience occasional bursts of performance and junction temperature that impact the lifetime behavior of the part but less than a part that experiences high temperature for extended periods of time.

Although these instantaneous short bursts have little impact on junction or skin temperatures, previous study [5] showed that they are highly constraint by power delivery network. Power supplies are also constrained in multiple time constants. Electrical limits are instantaneous and should never be exceeded while conversions losses that are dissipated as heat are thermally limited over longer periods of time. The power supply limitations are described in more details in [5]. Power density also impacts the junction temperature. Multi core processors are becoming ubiquitous all the way from high end server platform to small hand-held smartphones. Single core generates high power density as described in Fig 4. A single core however generates lower power then the power of multiple core running simultaneously even with Turbo operations. A single core in the system we have tested reaches the maximum allowable voltage before reaching junction temperature.

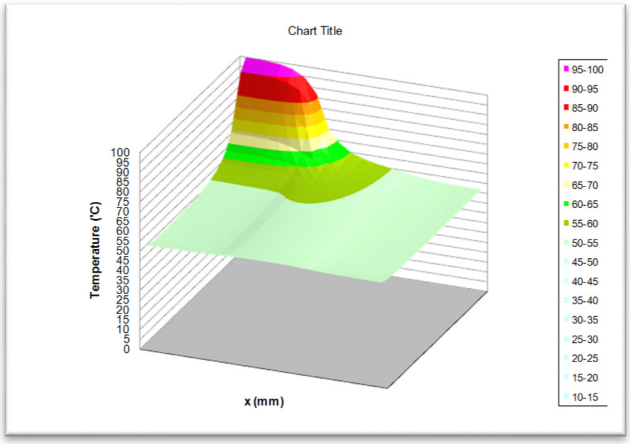


Figure 4: A single core active crates high power density, resulting with thermal hot-spot

Modern system on a chip such as the Intel® Core™ 2 duo incorporate a power management unit, in form of embedded controller, software module or hardware. The power management unit collects power consumption and temperature of various telemetry points and at any given time constrains the power and performance of the system to meet the highest constraint of all.

3 Evaluation of power and thermally constrained computer performance

We instrumented a quad core 32nm Intel® Core™ 2 Duo 2860QM [6] system with power and thermal measurement capabilities of various system components - CPU, graphics processor, DDR memory etc. The lab setup is shown in Fig 5.

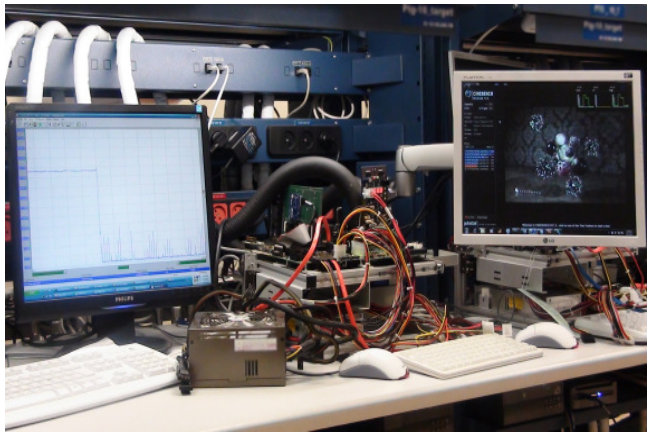


Figure 5: Experimental setup of a quad core Intel® Core™ 2 Duo 2860QM system

3.1 Sustained power management

At this study we evaluate the steady state power limited performance of a computer system (Fig 6).

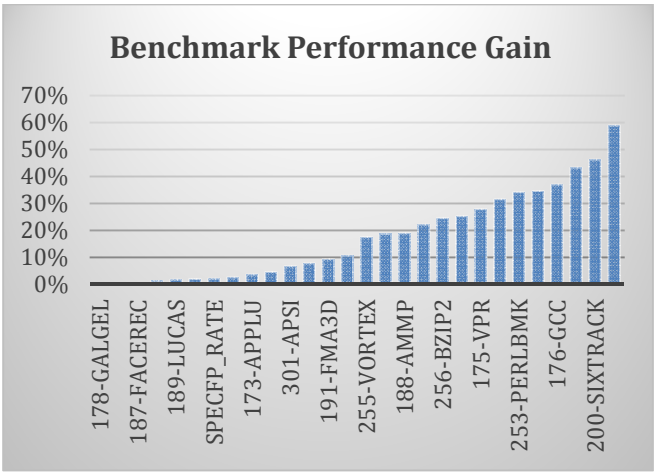


Figure 6: Performance gain of SPEC benchmark components from fully utilizing thermal headroom.

We ran 28 components of SPEC benchmark [7]. These benchmarks are characterized by a long run time of heavy workload. We first run the benchmarks at the fixed 2.5GHz, which is the guaranteed frequency of the Intel® Core™ 2 Duo 2860QM and measured the benchmark score. We then repeated the test allowing the frequency to increase performance of the same benchmarks with variable frequency until it reached the part specification limit of 45W. We use the internal power management features of the Intel® Core™ 2

duo to turbo up to the specified power limit. The junction temperature is monitored internally and maintained lower than the maximum allowable junction temperature. We measured the benchmark scores and compared to the scores achieved at 2.5GHz. The performance benefits from the increased thermal headroom is described in Fig 6. An average of 17% with up to 59% performance gain is achieved by allowing the CPU to run at the high frequency, while maintaining steady state power constraints.

3.2 Instantaneous power bursts

Human interactive workloads are characterized by burst of activity separated by idle period for idle time. The heat capacity of the heat sink allow burst of high performance and power for those short period of times of several seconds without violating junction or case temperature. We use Sysmark 2007 [8] and 3DMark Vantage [9] which are intended to represent user interactive scenarios. We observed an average of 32% performance with up to 41% performance gain over guaranteed frequency (Fig 7).

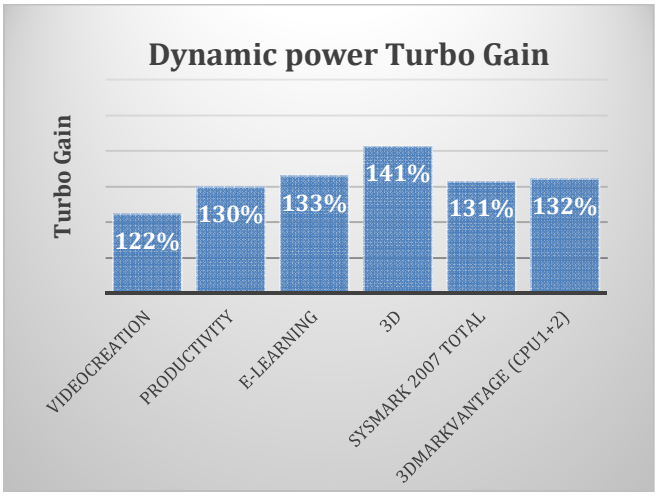


Figure 7: Burst performance gain over guaranteed frequency

3.3 Balancing power budget of on SoC computation engines

The power specification of the device defines the total power of the SoC. The main power consumers of the Intel® Core™ 2 Duo are the General Purpose IA processor and the Processor Graphics (PG). Workloads that use both IA and PG simultaneously can assign different power budget to the two while keeping the sum of all SoC components power constant. We evaluated the impact of assigning different power budget to the different components (Fig 8).

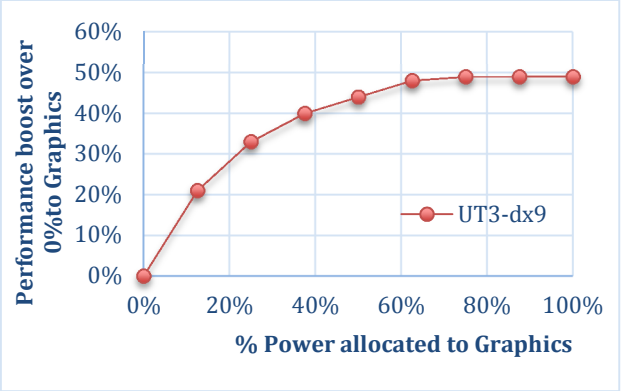


Figure 8: Power budgeting between CPU and Processor Graphics

We use a high end game - Unreal Tournament 3 that uses heavily both IA and PG. We run it in demo mode and use frame rate as a measure to overall performance. In this test we used the built in capability of the Intel® Core™ 2 Duo to control the power of each power rail independently. We started by assigning no power budget to the PG and gradually increased its power limit. As we gave more budget to the PG, the overall performance increased as long and the IA had enough performance to prepare work for the PG (Fig 8). At some point, 60% in this particular example, the PG saturated and additional power budget did not allow more performance. The proper balance is workload specific and power management algorithms in Software or firmware are responsible for selecting the optimal operation frequency and power [3].

3.4 Dynamic thermal analysis

We modeled the dynamic thermal respond of a notebook system designed to cool a 45W CPU (Fig 9). The model was tuned to reach a steady state maximum junction temperature of 100°C while dissipating the rated power of 45W. The dashed blue line describes a power step function of 45W. The solid red line describes the increase of junction temperature over time until reaching the steady state. The time to reach the steady state in this thermal solution was over 200 seconds. We applied a burst of 90W (green dotted line) and as a result, the junction temperature increased much faster, reaching the max allowed junction temperature within 20 second. The power control algorithm reduced the power gradually until it stabilized on 45W steady state power and 100°C junction temperature. It is possible to consume as much as twice the power for 20 seconds, delivering 34% higher performance for this period of time.

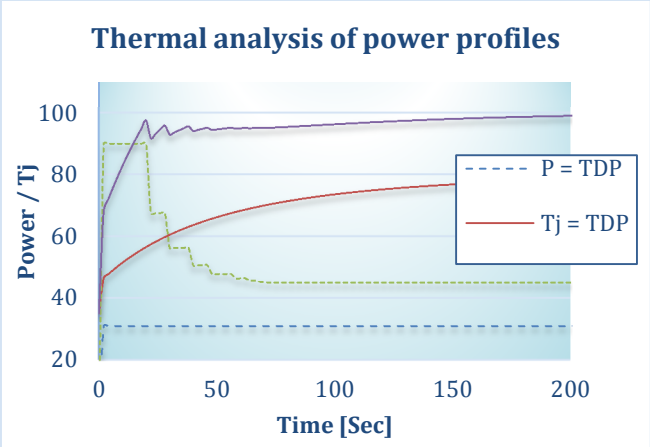


Figure 9: Dynamic thermal response of steady state and turbo power profile

For a steady state workload, this increased frequency is averaged over a long time and not noticeable. For interactive workloads however (Fig 7), the performance increase is highly noticeable. Idle periods between bursts of high activity allow the heat-sink cools down, enabling additional performance bursts, as long as the average power over thermally significant period of time, does not exceed the steady state cooling capability.

3.5 Enclosure skin temperature limit analysis

Thermal limitation of small form-factors are often limited by the enclosure skin temperature (Fig 2). In this study (Fig 10) we apply a burst of high power for ~50 seconds until the enclosure reaches the temperature limit and then lower the power to its steady state (Green dotted line). The junction temperature rises to high values before stabilizing to steady state of 80°C. Although the steady state temperature is lower than the maximum specification, the silicon experiences periods of high temperature that impact the parts lifetime stress.

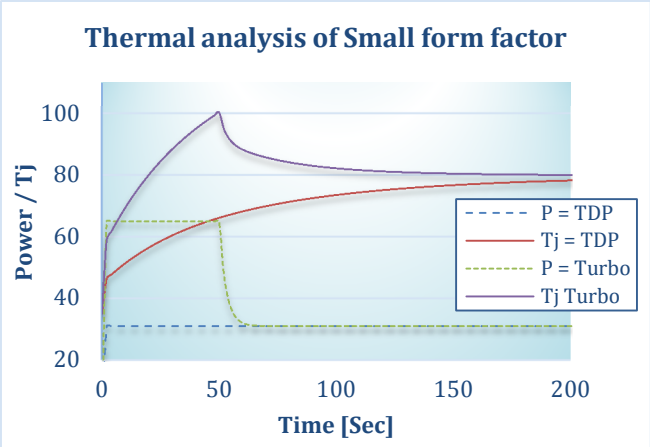


Figure 10: Thermal analysis of skin limited enclosure. Power burst drives high T_J and then stabilize on lower steady state temperature

3.6 Power delivery and reliability limits

Dynamic thermal behavior that is obtained by the heat capacity of the cooling system's thermal mass allow bursts of very high power for short periods of time (Sec. 3.4 and 3.5). This power however is further constrained by additional platform parameters:

1. Voltage regulator's (VR) ability to deliver this high current. Exceeding the voltage regulator rated values may cause permanent damage or trigger over current protection that shuts down the system. Previous work [5], [10] discussed in more details the effect of power delivery capability and topology.
2. Junction temperature rises much faster than the enclosure temperature and need to be maintained within safe values. In particular, Fig 4 describes example of a single core running at its highest voltage and frequency. A single core generates local junction temperature hot spot, reaches the silicon specification limit but consumes less power than multi-core scenario.
3. Reliability limits. The primary factors that impact silicon aging are voltage and temperature.
   a. The maximum frequency is a function of voltage, and cannot increase beyond the max rated voltage and frequency, even if the overall power is low and thermal headroom exists.
   b. Fig 10 exemplifies a case where steady state junction temperature is 80°C with occasional bursts to 100°C with high voltage and frequency. Overall reliability stress models cannot use steady state scenarios and need to account for such bursts.

4 Summary and conclusions

Junction temperature at steady state power has been considered the primary limiter for delivering high computational performance. Recent trends of small form factors and ever growing focus on user experience scenarios increase the importance of managing compute performance within multiple physical constraints. We have demonstrated in this work up to 59% performance gain using existing static and dynamic thermal headroom. The thermal capacity of the enclosure allows user perceived responsiveness without compromising ergonomic limitations of an enclosure skin temperature.

Acknowledgements

This work was supported by "ICRI-CI" – Intel Collaborative Research Institute for Computational Intelligence"

References

[1] C. Isci, A. Buyuktosunoglu, C. Cher, P. Bose and M. Martonosi, “An Analysis of Efficient Multi-Core Global Power Management Policies: Maximizing Performance for a Given Power Budget,” In Proc. 39th Annual IEEE/ACM Int. Symp. on Microarchitecture, 2006.

[2] Mobile 4th Gen Intel® Core™ Processor Family: Datasheet, Vol. 1, [online], Available: www.intel.com

[3] E. Rotem, A. Naveh, A. Ananthakrishnan, E. Weissmann, and D. Rajwan, "Power-Management Architecture of the Intel Microarchitecture Code-Named Sandy Bridge," IEEE Micro, vol. 32, no. 2, pp. 20-27, March-April 2012

[4] Advanced Configuration and Power Interface (ACPI) Specification, [online], Available: www.acpi.info/

[5] E. Rotem, A. Mendelson, R. Ginosar, and U. C. Weiser. 2009. Multiple clock and voltage domains for chip multi processors. In Proceedings of the 42nd Annual IEEE/ACM International Symposium on Microarchitecture (MICRO 42)

[6] Intel® Core™ i7-2860QM Processor, [online], Available: www.intel.com

[7] Standard Performance Evaluation Corporation, [online], Available: www.spec.org/

[8] Bapco®, SYSMark 2007, [Online], Available: www.bapco.com/

[9] 3DMARK® Vantage, [Online], Available: http://www.futuremark.com/benchmarks/3dmark-vantage

[10] U. Y. Ogras, R. Marculescu, P. Choudhary and D. Marculescu, “ Voltage-frequency island partitioning for GALS-based networks-on-chip,” In Proc. 44th Annual Design Automation Conference, June 2007.