Microelectronics Journal 46 (2015) 1225–1229

Contents lists available at ScienceDirect

Microelectronics Journal

journal homepage: www.elsevier.com/locate/mejo

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

Efraim Rotem [a,b,*], Ran Ginosar [b], Avi Mendelson [b], Uri C. Weiser [b]

[a] Intel corporation, Haifa, Israel
[b] Technion, Israel Institute of Technology, Israel

ARTICLE INFO

Article history:
Received 25 November 2013
Received in revised form
1 September 2015
Accepted 1 September 2015
Available online 20 October 2015

Keywords:
Power
Thermal
Energy

ABSTRACT

Junction temperature has been considered the main physical constraint for high end processors. Recent trends in form-factors and the increased focus on thin and light systems such as Ultra Book, tablet computers and smartphones, shift the focus away from managing and controlling junction temperature. Ergonomic considerations and power delivery are becoming the limiters of high computational density. 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™ i7 system.

1. Instruction

Continuous advances in a process technology enable the integration of an increased number of transistors onto a single die. Moore's law is expected to deliver even higher transistor density for the foreseeable future. This increased transistor density enables the integration of CPU cores, graphics processor, memory controller and other platform components into a modern SoC (System on a Chip). A modern CPU contains over a billion transistors, on a single monolithic die. For the last few process generations however, the process technology does not deliver power and energy improvements. This increase in transistor count, 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™ computers drive the power and thermal envelopes of these computer systems further down. More focus is put on the various aspects of user experience including responsiveness to user interaction, GUI operations, sustained general purpose compute, rich graphics and media content, and ergonomic considerations. Most modern computer systems cannot sustain all the SoC components operating at their highest power-performance state, all of the time. Power management has become the primary mechanism to improve the user experience within multiple system constraints. Power management features are designed to provide the maximum performance that is possible within the package and system power and thermal constraints 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 techniques aim to maximize the user experience under multiple system constraints. The user experience has various attributes:

- Throughput performance – sustained computation for a long period of time, either general purpose computation, graphics and media processing, audio streaming etc. The user may have a preference between the 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 is important to the user, from the battery operated device all the way to the data center
- Ergonomics – acoustic noise generated by active cooling systems, enclosure 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.

* Corresponding author at: Intel corporation, Haifa, Israel.
E-mail address: efraim.rotem@intel.com (E. Rotem).

http://dx.doi.org/10.1016/j.mejo.2015.09.002
0026-2692/

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

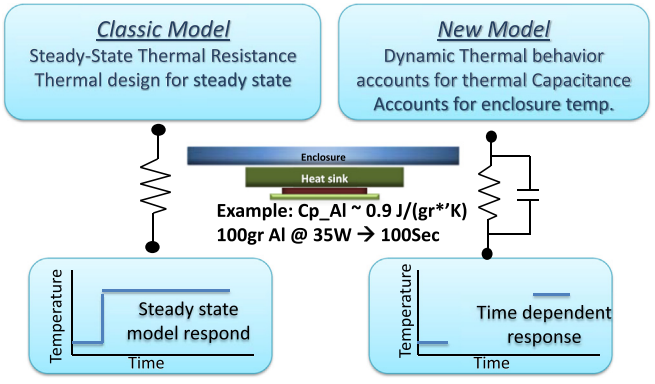
2.1. Thermal limitations

Junction temperature and the ability to cool the die have been considered to be the primary limiter for delivering high performance computation [1]. Processor manufacturers specify the Thermal Design Point/Power (TDP). It is an average power dissipation for the processor when running high power consumer applications, such as word processors, 3D gaming and video transcode. Its intent is as a guideline for platform designers (e.g. designers of laptops). This defines the true thermal envelope that the system need to support. Computer manufacturers are expected to design the system thermo-mechanical parameters for cooling the TDP power at sustained operation. In small form factor systems, such power definition limits the maximum voltage and frequency of the SoC. Most workloads however consume much lower power and can benefit from a power budget that allows increased frequency and performance [2].

Physical behavior of a cooling system is characterized not only by the steady state conductivity but also by heat capacity. A typical heat sink can absorb a substantial power surge until the heat sink heats up while keeping the junction temperature within specifications. For example, a 100 g Aluminum heat sink (~0.9 J/(g °C) that is designed to cool 5 W CPU, starting at 35 °C can sustain a burst of 35 W for 100 s until the processor junction temperature reaches its limit. Modern CPUs [4] calculate this cooling model (Fig. 1) at run time and control the CPU power accordingly.

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 human skin. A typical acceptable temperature of hand held devices is approximately 45 °C. The power-performance profile of a small form factor device is determined by the most constraining limit between junctions and skin temperature (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 ~1 W (infinite time). The SoC may consume 1.5 W for 7200 s, 4 W for 240 s or 6 W for 100 s 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 3 W but at these power levels the skin temperature is more constraining. The steady state maximum power would therefore be ~1 W with die junction temperature much lower than the limit. It is also possible to burst the CPU as high as 6 W for 5 s before the junction overheats. This time is much too slow for the skin to experience any meaningful change in temperature.

Modern processors such as the Intel® Core™ 2 duo [3] make use of this thermal profile. The solid line in Fig.3 describes a 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 periods 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



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

Fig. 1. Static steady state model and dynamic thermal model. (a) Static model. (b) Dynamic model.

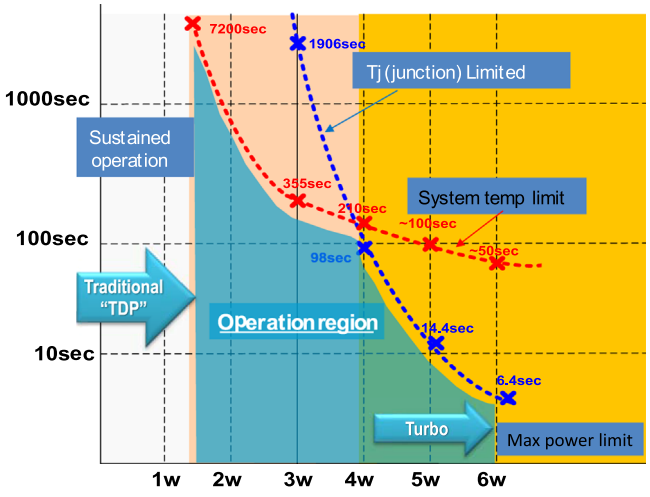


Fig. 2. Small form factor dynamic cooling capability profile constrained by the highest of junction temperature or case temp. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)

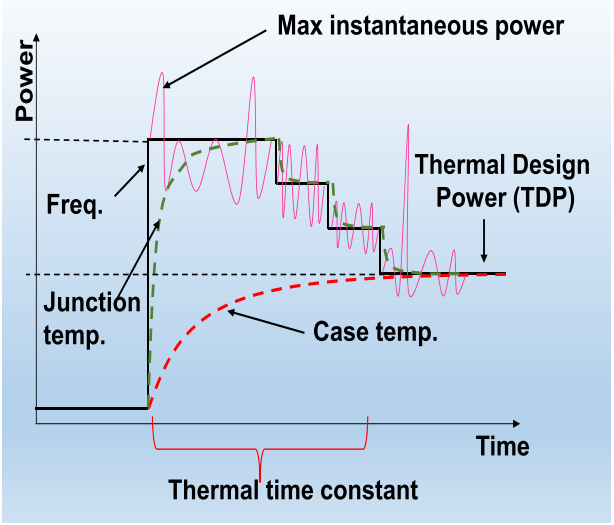


Fig. 3. Turbo power profile – After idle period a burst of high power is allowed for a thermally significant time, stabilizing to steady state thermal design power. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article.)