An Efficient Transient Thermal Simulation Methodology for Power Management IC Designs
Karthik Srinivasan, Stephen Pan, Zhigang Feng, Norman Chang, Tim Pawlak
ANSYS Inc.,
2645 Zanker Road, San Jose, CA-95134, USA
{karthik.srinivasan,stephen.pan,zhigang.feng,norman.chang,tim.pawlak}@ansys.com

Abstract
Power Management devices are becoming ubiquitous in every electronic system for achieving energy efficiency with constrained power/thermal budget. Multi-Function and Multi-Channel PMICs are becoming common design trend to support diverse voltage/power requirements of complex SoCs. In this paper, we present an approach to perform a full chip level thermal analysis with the capability to perform a detailed sub-modeling for electro-thermal analysis with Finite Element method and perform thermal-aware EM and stress analysis. The approach includes transient thermal, thermal-aware EM and stress analyses includes the generation of thermal-aware chip power maps, conversion of converged thermal profiles in Power Devices to thermal loadings and detailed sub-modeling of on-chip structures for transient thermal, thermal-aware EM and thermal-induced stress analyses.

Keywords
Power management, transient thermal, Finite-Element, thermal induced stress, thermal-aware EM, Sub-modeling

Nomenclature
SoC - System on Chip, PMIC - Power Management Integrated Circuit, EM - ElectroMigration

1. Introduction
A typical mobile PMIC [1] containing multiple channels of buck and boost converters as key components is shown in Figure 1. The major power dissipation in the multi-functional PMICs is due to these DC-DC converters supplying wide range of voltage and power levels. Along with these buck converters, there can be several additional peripheral components like ADC/DAC, USB controllers embedded in a single system. Although these are isolated from each other all the way to package and PCB, they are in very close proximity so it is critical to consider thermal interactions between these components to perform a reliable thermal modeling and analysis of these devices.

2. Power Modeling of Power Devices
At the heart of these DC-DC converters (a.k.a switching regulators) are the high-side and low-side switches with output connected to external LC low-pass filter as shown in Figure 2.

The switches are made of large MOS devices with hundreds or thousands of transistors in parallel. The layout of switching FET device in buck converter is shown in Figure 3. Having minimal routing resistance and optimal current distribution among the transistor fingers ensures minimal power loss. The switching regulators typically drive several Amperes of current and hence self-heating effects are not negligible. The power is also dependent on external loading of different switching regulators.
For determining overall thermal distribution, we can assume each regulator is driving its maximum current carrying capacity. The power conversion loss, which manifests as heat source can be modeled with the following equation.

\[ P_{loss} = P_{conduction} + P_{switching} + P_{gate\_drive} \]

Where \( P_{loss} \) is the power loss models only the major power loss mechanism which is conduction loss \( P_{conduction} \), switching \( P_{switching} \) and gate drive \( P_{gate\_drive} \) loss.

\[ P_{conduction} = I_{max}^2 \times (R_{ds\_ON} + R_{interconnect}) \times D \]

The conduction loss \( P_{conduction} \) depends on \( I_{max} \) is the maximum load current that buck converter can drive at highest efficiency. \( R_{ds\_ON} \) is the ON resistance of the Power MOSFET, \( R_{interconnect} \) is the interconnect resistance and \( D \) is the duty ratio of the switch. The power loss on switching regulator also includes switching and gate-drive loss which can be modeled by the following equations.

\[ P_{switching} = F_{sw} \times V_{ds} \times I_{max} \times \frac{T_{on} + T_{off}}{2} \]

\[ P_{gate\_drive} = F_{sw} \times V_{gs} \times Q_{g} \]

Where \( F_{sw} \) is the switching frequency of the regulator, \( T_{on}/off \) are duration when both high and low-side switches are ON at the same time due to non-ideal slew of the gate drivers. The Gate-drive loss is similar to switching power in a typical static CMOS logic which depends on \( V_{gs} \), switching frequency, and Gate charge represented as \( Q_{g} \).

The switching functional blocks like LED drivers, ADC/DAC and other peripheral circuitry can be considered as blocks of typical SOC as

\[ \text{Power} \sim C_{L} \times V_{dd}^2 \times f \]

If each instance is a heating object, the thermal modeling efforts will be overwhelming to simulate in current FEM technology. The current approach is to lump them into smaller number of heat sources. The transistor or instance power is typically temperature dependent and mainly due to leakage current component. The total power in a device includes the dynamic switching power and the leakage power. Figure 4 shows the proposed flow for Chip Thermal Model (CTM) generation of a PMIC design from the chip power integrity tool [5] with a collection of the power map calculated at several temperature points mainly due to leakage power which is temperature-dependent for the chip. The generation of CTM was from either vectorless or event-driven power scenarios. It is the power on the chip at a specific operating mode averaged over a period of time. The package/PCB data along with Chip Thermal Model is then used in Computation Fluid Dynamics solution [7] to determine the thermal boundary condition of these complex multi-functional PMICs.

The above approach is useful in establishing the base-line temperature and also gets a smeared tile-based temperature distribution across the silicon considering package/PCB [2,3]. However, for determining accurate thermal gradient across the power-management device and thereby further model the interconnect reliability [4] and stress on on-chip interconnect or package layers/bumps, a much finer modeling is required.

3. Thermal Modeling for Power Devices

The flow for performing fine-grained thermal analysis taking detailed chip level data into a large capacity Finite Element solution [6] is described in Figure 5. Chip wire and device self-heating was generated from power analyses in the chip power tool and applied to finite element models for either full chip or sub-regions. The finite element models were solved for thermal profiles on the chip and the wire temperatures extracted to a form ready for back-annotation in the final EM (Electro-Migration) analysis for chip design sign-off purpose.

For complex and large chip designs, the full chip model could be excessively big for direct thermal field solution. However, the Joule heating on wires are tiny and are usually limited in the region of influence to victims, e.g., local thermal responses from an aggressor decays to less than 10% in 5um range when surrounded by the low-K dielectric materials [4].
4. Full chip Thermal Modeling

If the full chip was divided into sub-regions, the model size could be greatly reduced to accelerate the solution process (Figure 6). The resulting wire temperature from different sub-model can easily be assembled for the full chip in back-annotation of thermal-aware EM analysis.

For chip-only modeling, reflective thermal boundary conditions are usually used for either the full model or the sub-models. The assumption in the reflective thermal boundary condition is that no heat fluxes into or out of the edges of the analysis model in the planar directions. For the influences from device heating to wires, the formula-based approaches from foundries are mostly assuming vertical coupling to wires only and that is considered sufficient based on silicon data. The symmetric thermal boundary conditions for the devices near the cutting edges are close to this vertical only status used by foundries, but most part of the wires/devices coupling in the sub-model are realistic in all the directions.

5. Results

In the following section, we discuss the results from Finite Element simulation using the chip level sub-modeling approach described above. The thermal contour of wires above twelve hot devices with views from top and bottom the interconnection layers is shown in Figure 7. Joule heating on wires are also included for thermal coupling effect among wires, but of secondary effects in thermal responses due to strong vertical thermal coupling between wires and devices with high power.

Figure 8 shows normalized temperature distribution among the device layer using the sub-model based Finite Element simulation. Temperature was plotted along an array of devices. Devices are numbered from 1 to 6 in the plot. The temperatures are normalized with ambient temperature.
A similar plot can be generated along the layer stack from the device to the top-most metal layer as shown in Figure 9.

![Figure 9: Temperature distribution from device to the metal interconnect (z-direction)](image)

Since the power management devices are made of switching regulators, transient thermal analysis is required to determine the final temperature of the power FET. Transient responses on chip are affected significantly by the environment of the chip, i.e., package and system. The thermal capacity or capacitance of the chip is normally only less than 5% of the thermally affected environment and temperature cannot rise abruptly even the power is high. Typically, in a thermal test of a chip in package, the time to steady state of temperature on chip is up to 15 min or close to 1000 sec depending on the package sizes and other environment factors. If the thermal capacity of the package and system is not accounted for, chip temperature can instantly reach steady-state which is not likely to happen in a real case. In this study, heat capacity was assumed and assigned to the bottom side of the chip substrate so that the time to steady state, as shown in Figure 10 is around 100 sec for a step change of the device power, which is the average power of the power cycling described in Figure 11. The vertical axis is the device temperature normalized with respect to the ambient temperature of the chip. At steady state, the device temperature is up to about 2.8x of the ambient temperature.

The typical switching frequency of the buck converter is assumed to be 500KHz for these experiments. At 500KHz power cycling of high/low power modes, oscillation of temperatures can be extracted. At high power mode, the power on each device was up to around 20mW which may only happen in PMIC cases. The low power mode is at 1/10 of that from high power mode. About 0.5 normalized temperature in oscillation amplitude was observed in this case. However, the overall trend of temperature rise is governed by the average power in the cycling condition, i.e., 55% of high power level on the device.

The plot shows the transient response of the initial one micro-sec for both power cycling and average power conditions. As shown in Figure 11, the power cycling responses follow the average power responses closely and the trend is expected to be the same up to the steady state temperature beyond 100sec. The overshoot of temperature in each power cycling is about 0.3 normalized temperature. The average temperature rise to steady-state from the Finite Element simulation at each geometry is further used to determine the DC EM violations on the interconnects considering the post-thermal effects on DC EM limit. This is particularly important since the thermal gradient of an PMIC can be larger than 30 degrees.

6. Size/Performance Trade-off

Figure 12 shows the comparison of the thermal profiles on a bottom metal layer in the full model of the big FET in buck-converter (0.321x0.124mm with six metal layers) to that of a quarter model. Both the Tmax or the temperature distribution...
are similar, which implies that the symmetric BC has minimal impact on the results. Plots in Figure 13 shows that sub/quarter model is much more efficient than that of the full model, especially in the modeling time. The total solution times using one full model and four quarter models are about the same. The solution times are linear in terms of degrees of freedom (dof), i.e., 20.8M vs. 5.9M in the full/quarter models.

The solution in Figure 13 used Shared-Memory Parallel (SMP) option of High Performance Computing (HPC) in thermal analysis tool with 4 cores. The solution time of the full model is less than one hour. Further speed up of the solution time is possible using the Distributed Computing or MPP option of HPC.

Figure 13: Solution time comparison between quarter of chip and full-chip

![Solution time comparison](image)

Figure 14: Scalability of FEM solution for the thermal analysis tool

![Scalability of FEM solution](image)

Figure 14 is a separate performance tests of a chip model of 24M and 72M dof’s using MPP with 16 cores. The solution time for the case of 72 million unknowns is 3.75 hr which is acceptable to the chip designer and is linearly scalable to the dof’s. The largest case that the FEM-based thermal analysis tool ever solved was up to 2 billion unknowns, limited by the configuration of the servers used. However, for thermal analysis on complex multi-channel PMIC designs the complexity can go beyond this limit. Since the accuracy loss is minimal, the approach suggested section 4 of generating multiple sub-models for different partitions of the chip and then solving them in parallel will help achieve reasonable runtime and optimal memory usage.

7. Thermal-Induced Stress Analysis

It has been demonstrated that either warpage or thermal-induced stress can be analyzed based on the thermal sub-model [8] to reveal localized stress on chip wires or in low-K dielectric material layers. Figure 15 [8] is an example of the equivalent stress on chip wires showing the high stress zones which is useful for further review of chip reliability.
Figure 15. Stress concentration at via/wire intersection due to thermal-induced stress.

Conclusion

This paper describes an automated sub-modeling approach with chip layout details in the finite element model. The flow starts from chip power calculation of self-heating on device and Joule-heating on wires, generates either a full model or sub-models for thermal analysis with thermal boundary conditions from package/system, and back-annotates thermal-aware EM with the resulting wire temperature for EM sign-off. The approach was demonstrated with a full PMIC chip model and with the alternatives using sub-models to cover the full chip area. Either steady-state or transient FE analyses were tested to demonstrate the features in extracting temperature distribution and predicting temperature growth and overshooting at a power cycling condition of a PMIC design. Discussions on model size/performance trade-off are presented. The finite element approach is accurate as it models chip details with minimal assumptions and is flexible in extending the application to thermal-induced stress on wires and in low-K dielectric materials of a chip. Further studies will be on increasing the efficiency in modeling and simulation to make this approach a viable flow in chip design, especially for PMIC with relatively small footprints.

Acknowledgements:

Authors would like to thank Dale Ostergaard, Ying-Shiu Li from ANSYS for their valuable feedbacks during discussions on power management, finite element simulation related topics.

References

1. Natahan Hanagami, Steve Knoth and Marty Merchant, “Car infotainment systems need multi-output power ICs”, EE times article, March 5th, 2013

2. S. Pan, N. Chang, M. Ma, G. Shankaran, and M. Nagulapally, “Thermal Co-analysis of 3D-IC/Packages/System”, DesignCon2013
5. User manuals of ANSYS Redhawk and Totem, R18.0, 2016