Power Envelope Analyses of Chiplet Module and System-on-Chip

Eric Ouyang, Xiao Gu, Yonghyuk Jeong
JCET Group
880 N. McCarthy Blvd., Suite 250
Milpitas, California, 95035, USA
Ph: 510-979-8383
Email: eric.ouyang@jcetglobal.com

Abstract

A novel power management method was developed to predict the reliability of silicon, and the method is applicable to both multichip and System-on-Chip packages. The System-on-Chip here refers to a single die which has many functional blocks and each functional block can be treated as an individual die. With power envelope analysis, the risk values of die powers are determined. An effective risk value considering all the powers of dies may be defined to determine the overall impact. As circuit design engineers select and define the power magnitudes of dies, the distances to the threshold planes on the power envelope plots are calculated to determine the thermal reliability. The method also allows reliability engineers to define a weight scale value to reflect the ruling of each die, and this is very important for practical design and assembly because different dies may have different levels of reliability concerns. An advanced histogram method was used to compare the impact of power magnitudes of dies on the thermal reliability. We have implemented the approach to study the thermal reliability of a chiplet module and a System-on-Chip. In conclusion, this paper implements a power envelope analysis to help determine the allowed and optimized powers of the dies on a chiplet module or the powers of the functional blocks on an SoC package.

Key words

Chiplet, System-on-Chip, Power Envelope, Thermal Reliability

I. Introduction

A chiplet is an integrated circuit (IC) that has been specifically designed to work with other similar chiplets to form a larger and more complex package. The chiplet modules, which divide a complex design such as a high-end processor or a networking chip into several small dies instead of one large monolithic die, are a good alternative, especially for the data-center products because of the demanding computing performance and the consideration of the cost of design and manufacturing. By using two or more chips, designers can use different nodes, which may be older nodes coupled with leading-edge nodes, to save cost and achieve demanding computing performance. Advanced packaging solutions, together with chiplet technology, enable higher levels of integration and improve overall system performance. In addition to the cost of design and manufacturing, the benefits of chiplets also include the improvement of die yield, shorter time to market, wider supply chain, and possibly lower junction temperatures of dies, compared to that of using monolithic die or a System-on-Chip (SoC) solution [1]. For chiplet module and SoC die, the study of thermal performance is usually focused on the temperature distribution and the magnitudes of junction temperatures of chiplet dies or the functional blocks on SoC die, and the impact of power magnitude or distribution on thermal reliability does not seem to be straightforward or is not the first parameter to be considered [2, 3]. However, from the perspective of circuit design, we may need to know the safety range of powers to design the package from the very beginning of the design cycle. Therefore, in this paper, we will present a power envelope analysis to evaluate the thermal reliabilities of the chiplet module and SOC die. The impact of the powers of dies on the chiplet module and the powers of the functional blocks on an SoC die on thermal reliability is studied.

II. Methodology of Power Envelope Analysis

The thermal test environmental conditions for multi-chip packages are specified by the JCEDEC 51-31, and we may describe the relationship of powers and temperatures of the dies on a multichip package with Equation 1 [4], wherein θ

is the thermal resistance, P is the power, T is the temperature, and T_a is the ambient temperature. If the temperatures of the dies are the maximum allowed junction temperatures, the allowed power magnitudes of dies are bounded by Equation 2. If the thermal resistance matrix θ is treated as a characteristic parameter of the multichip package, and the maximum allowed junction temperatures of dies are fixed, not dependent on other variables or factors, we may focus on the manipulation of the powers of dies for thermal management purposes. We will define this approach as the power envelope analysis, and this approach is shifting our attention from the optimization of die temperatures to the optimization of die powers [5].

$$\begin{bmatrix} \theta 11 & \cdots & \theta 1n \\ \vdots & \ddots & \vdots \\ \theta m1 & \cdots & \theta mn \end{bmatrix} \begin{bmatrix} P1 \\ \vdots \\ Pn \end{bmatrix} = \begin{bmatrix} T1 - Ta \\ \vdots \\ Tn - Ta \end{bmatrix} \quad (1)$$

$$\begin{bmatrix} \theta 11 & \cdots & \theta 1n \\ \vdots & \ddots & \vdots \\ \theta m1 & \cdots & \theta mn \end{bmatrix} \begin{bmatrix} P1 \\ \vdots \\ Pn \end{bmatrix} \leq \begin{bmatrix} Tj1 - Ta \\ \vdots \\ Tjn - Ta \end{bmatrix} \quad (2)$$

Mathematically, there are infinite sets of die powers to meet the requirement of Equation 2, but in real applications, the magnitudes and the ranges of die powers dies are usually determined, for example, by their electrical designs, the conditions of implementations, or the purposes of applications. To demonstrate the usage of power envelope analysis, we are using an example as shown below to do the power envelope analysis. The first matrix at the left side is the thermal resistance matrix, and [p1; p2; p3] are the allowed power magnitudes, and the [95; 95; 95] are the differences between the junction temperatures of the dies and the ambient temperature. For this case, we will assume the maximum allowed junction temperatures are all 125C for all three dies and the ambient temperature is 30C. However, please note that if the dies are made with different fab processes and have different material properties, they may have different maximum allowed temperatures.

$$\begin{bmatrix} 0.2881853, & 0.173129524, & 0.186470 \\ 0.1796696, & 0.385964571, & 0.208115 \\ 0.1739064, & 0.188766857, & 0.391953 \end{bmatrix} \begin{bmatrix} P1 \\ P2 \\ P3 \end{bmatrix} \leq \begin{bmatrix} 95 \\ 95 \\ 95 \end{bmatrix} \quad (3)$$

For a 3 by 3 thermal resistance matrix, three power envelope planes were obtained. When a set of power magnitudes of dies are selected, the distances of the selected power magnitudes to the power envelope planes can be calculated. For this particular case, if the selected powers are [350; 100; 270] Watts, shown as the red dot in Figure 1, the power values are definitely over the allowed magnitudes, and the dot is above the power envelope planes. The distances of red dot to three envelope planes can be calculated as well and the data is shown in Figure 2, and we will define these data as risk values as they indicate the power reliability of the dies. The positive risk values illustrate the risks to be concerned thermally. If the selected powers are [80; 100; 125] Watts, as shown in Figure 3, the red dot is under the power envelope surfaces and it is invisible, thus the power magnitudes of dies are safe and they do not have risks. The lower the risk values, the safer the dies are thermally. The distances of dot to three envelope planes are calculated and the data is shown in Figure 4. With the proposed methodology, the power reliability of dies can be evaluated clearly and the methodology is useful when we are doing the design of circuitry, thermal management, and reliability prediction. In the following, we will apply this methodology to a chiplet module and an SoC die. We will elaborate on the implementation of power envelope analyses for the chiplet or SoC having more than 3 dies or 3 functional blocks.

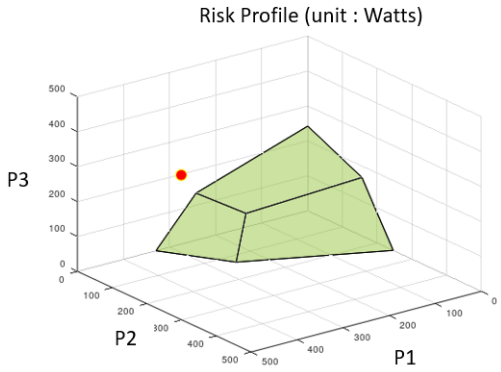

Figure 1. One example of power envelope plot

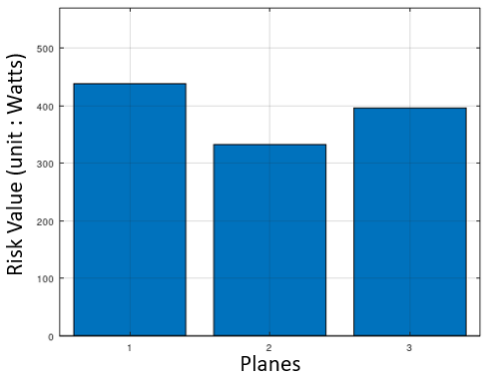

Figure 2. One example of risk values

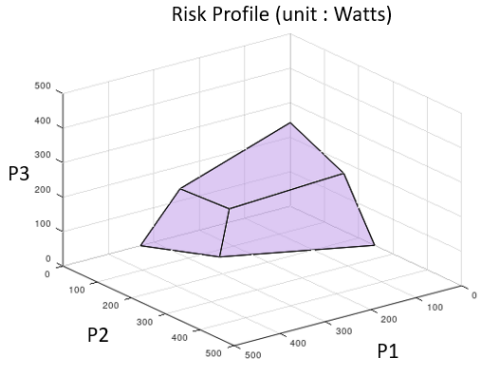

Figure 3. One example of power envelope plot

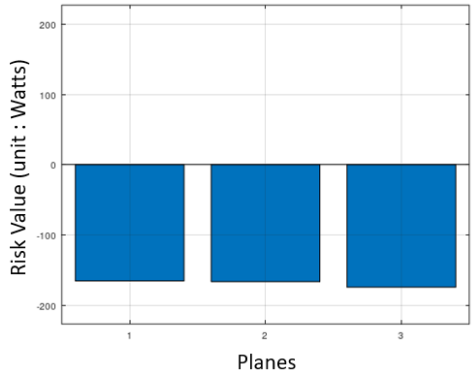

Figure 4. One example of risk values

III. Power Envelope Analysis of a Chiplet Module

In this section, we will demonstrate the use of power envelope analysis to optimize the power magnitudes of the chips on a chiplet module. A 2.5D interposer having 5 dies will be used. The same analytical approach can be used for different numbers of dies and for other similar configurations of IC packages such as multichip modules (MCM). Figure 5 illustrates a configuration of a 2.5D interposer chiplet module with an outline dimension of 38 mm by 29 mm. The size of the ASIC die is 20 by 27 mm, and there are four SRAM chips, and each SRAM chip size is 8 by 12 mm.

In order to demonstrate the application of power envelope analysis to evaluate the risks of the chiplet module, the chiplet module was assumed to be placed inside a simplified 1U server rack. The norm applied powers on the ASIC die and on each SRAM are 300 watts and 32 Watts respectively. The outside ambient temperature is set at 30C, and the flow rate of this forced convection 1U server rack is 20 cubic feet per minute. The in-plane heat sink size is about 107 mm x 78 mm, and there are 33 aluminum fins with a height of 20mm in the out-of-plane direction. The above selection of the configuration and the parameters is to mimic the real-world applications of a system server. Figure 6 shows an example of the temperature plot of the thermal model. The surface temperature of the heat sink fin is not plotted so we can see the temperature distribution of internal dies. A commercial electronics cooling simulation software, which is finite volume-based, was used. A "Multi Grid" simulation solver with at least 500 iterations was used. The smallest mesh size is about 0.3mm and the total number of mesh is over 9 million. The simulated temperatures were monitored to make sure they were converged, not changing more than 0.5C as a function of iterations or mesh quality.

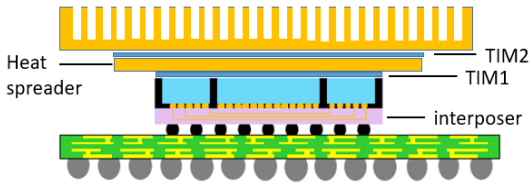

Figure 5. Configuration of a 2.5D interposer chiplet module.

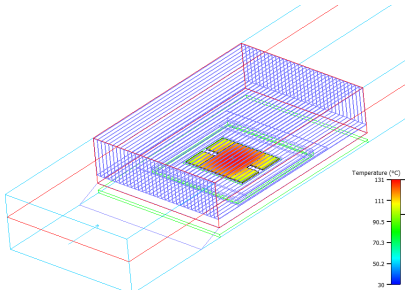

Figure 6. Thermal model of a chiplet module with 5 dies.

For a chiplet module having 5 dies, 5 simulation legs are needed to calculate the thermal resistance matrix. Table 1 is our selected power magnitudes of the dies for simulation. Leg 1 shows the norm powers of dies while other legs' powers are randomly selected within the operational range. With the selected powers of dies, the junction temperatures of the dies are calculated and Table 2 shows the data of junction temperatures. With the simulated junction temperatures and known power magnitudes, the thermal resistor matrix is obtained as shown in Table 3.

Table 1. Power magnitudes of dies

Powers (Watts)	Leg 1	Leg 2	Leg 3	Leg 4	Leg 5
Die	350	200	250	150	180
SRAM1	20	15	25	32	15
SRAM2	20	10	5	28	15
SRAM3	20	8	10	25	15
SRAM4	20	5	15	8	15

Table 2. Simulated junction temperatures

TJ (°C)	Leg 1	Leg 2	Leg 3	Leg 4	Leg 5
TJ-Die	136.34	89.661	105.66	84.899	87.705
TJ-SRAM1	116.22	79.219	94.355	83.321	78.604
TJ-SRAM2	115.83	77.129	88.192	80.612	78.431
TJ-SRAM3	115.78	77.405	90.692	81.051	78.355
TJ-SRAM4	115.4	75.749	90.088	75.221	78.181

Table 3. Thermal resistance matrix

0.26727	0.17258	0.20596	0.08406	0.17712
0.19468	0.44824	0.13514	0.20151	0.11923
0.19323	0.13187	0.43586	0.14504	0.1972
0.1937	0.22564	0.17423	0.34718	0.15225
0.19235	0.13076	0.26734	0.03838	0.46737

One important goal of the power envelope analysis is to verify the accuracy of the implementation of the thermal resistance matrix. We have selected a few other power magnitudes of the dies, and use the thermal resistance matrix to calculate the corresponding junction temperatures to compare with those obtained from the full thermal simulation. The errors of the calculated junction temperatures vs. simulated junction temperatures are around 0.1%, which is amazingly accurate. Following the same procedures as explained in the previous section, the risk values of selected powers to the power envelope planes can be calculated. Figure 7 shows the risk values for a set of powers, which is [320; 18; 10; 15; 15] Watts, of the five dies. An effective risk value of the selected powers can be calculated using Equation 4, wherein the *C1* to *C5* are weight scales, and *r1* to *r5* are the risk values to power envelope planes [6]. For this case, we assume that the weighting scales, *C1* to *C5*, are all equal to 1. When the power magnitudes of the dies are over the allowed limit, a positive sign of the value is used while the negative sign is used when the power magnitudes are within the safe region and are not over the allowed limits.

$$R_{eff} = \pm\sqrt{c1 \times (r1)^2 + c2 \times (r2)^2 + \cdots + c5 \times (r5)^2} \quad (4)$$

There are various implementations of effective risk values to evaluate the thermal reliability of the chiplet device. One implementation is to compare the effective risk values of the dies among different sets of power magnitudes. For example, the 5 dies of this chiplet may have 3 sets of operating powers, first set [310; 23; 17; 10; 19] Watts, second set [320; 18; 10; 15; 15] Watts, and third set [295; 25; 26; 18; 25] Watts. The effective risk values of the die powers may be calculated to compare their efficiency, and Figure 8 is an example. The effective risk values are negative, which means that these three sets of powers are all safe to use. Set 3, with its lower effective risk value, is probably safer compared to the other two sets, and the reason could be its lower die powers, at least by looking at the power magnitude of its first die with 295 Watts. At the time of preparing this paper, we believe there are various ways of implementing risk values to optimize the power distribution of the dies, and there is a need for more studies.

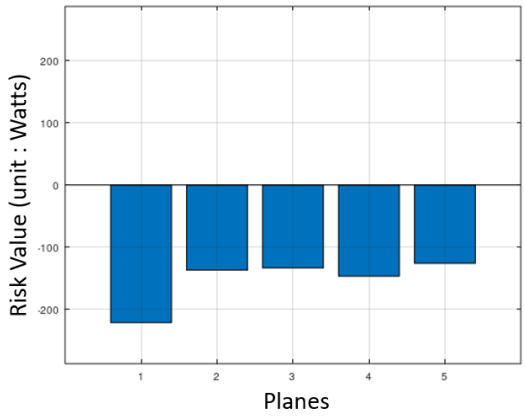

Figure 7. Risk values of a set of die powers on a chiplet module

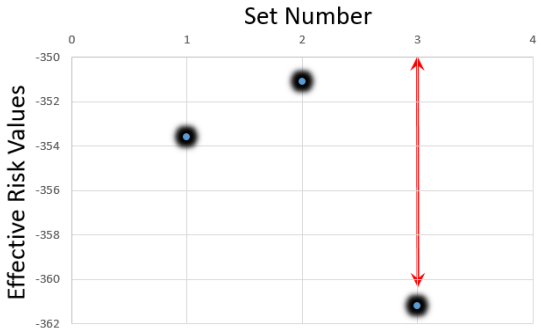

Figure 8. Risk values as a function of power set number

IV. Power Envelope Analysis of a SoC

After the implementation of power envelope analysis for chiplet module, one question was raised. Can the approach be used for a System-on-Chip which is a single die package but the die has several functional blocks. To evaluate the usage of the power envelope approach to study the thermal reliability of a SoC, a single silicon with 7 functional blocks is assumed as shown in Figure 9. These functional blocks are CPU, Fabric, GPU, two DRAM, Neural Engine, and Cache. The CPU, GPU, and Neural engine may have 4, 8, or 16 cores to perform complicated operations. The power magnitudes of these functional blocks may follow for example Table 4 wherein the “mid” is the norm power magnitudes of the blocks. The functional blocks may use higher or lower powers depending on the application situations of the device. For this power envelope analysis of SoC, the 1U server, flow rate, configuration of heat sink, and the ambient temperature are assumed to be same as those used for the analysis of previous chiplet module, and Figure 10 illustrates a figure of our thermal model.

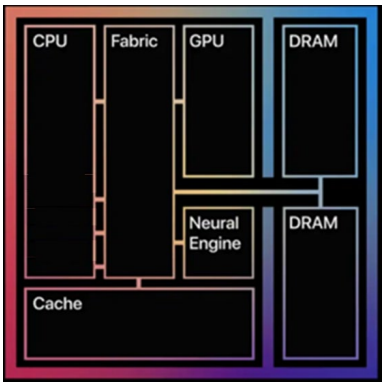

Fig. 9 A SoC die with 7 functional blocks

Table 4. Range of powers of functional blocks

	Max	Min	Mid	Max -30%	Min+30%	Max-50%
CPU	20	5	12.5	14	6.5	10
GPU	15	5	10	10.5	6.5	7.5
Fabric	5	1	3	3.5	1.3	2.5
Neural Engine	5	1	3	3.5	1.3	2.5
DRAM	15	5	10	10.5	6.5	7.5
DRAM	15	5	10	10.5	6.5	7.5
Cache	15	5	10	10.5	6.5	7.5

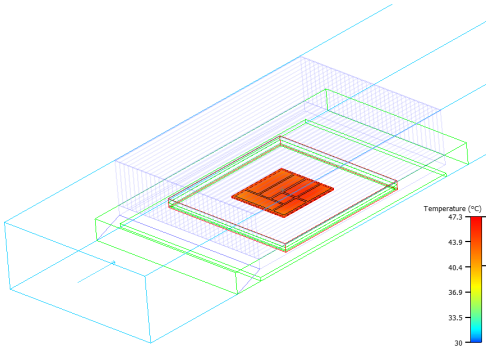

Figure 10. SoC thermal model

To obtain the thermal resistance matrix of this SoC having 7 functional blocks, we may randomly select the powers from Table 4 to form 7 different simulation legs in order to calculate the junction temperatures of functional blocks. The selection of the die powers must be appropriate to ensure matrix Equation 1 is solvable, for example, we cannot use the same die powers or the die powers are only different by a multiplier scale. If we do so, there will not be 7 different equations to solve the 7 variables. With the simulated junction temperatures, the thermal resistance matrix is obtained as shown in Table 5. Figure 11 is the histogram plot of the risk values of the dies with power magnitudes [92.5; 10; 3; 3; 10; 10; 10] Watts. This set of die powers is safe and the risk values are all negative. To verify the accuracy of this thermal resistance matrix, we have tested various power magnitudes of functional blocks within the operational ranges. The errors of junction temperatures obtained from the thermal resistance matrix compared to those obtained from full thermal simulation are under 1%.

Table 5. Thermal resistance matrix

0.488134	0.218859	0.298016	0.200667	0.187859	0.174731482	0.213939
0.256033	0.330839	0.479537	0.422	0.199089	0.159027659	0.266571
0.166391	0.549046	0.246798	0.282667	0.394296	0.219436226	0.207432
0.188533	0.268748	0.247541	0.646	0.271498	0.35758277	0.244235
0.163829	0.35457	0.217759	0.235333	0.61782	0.243985486	0.192721
0.168295	0.255522	0.23477	0.323333	0.293772	0.576467183	0.212124
0.128609	0.257602	0.311071	0.394667	0.262602	0.345501027	0.360587

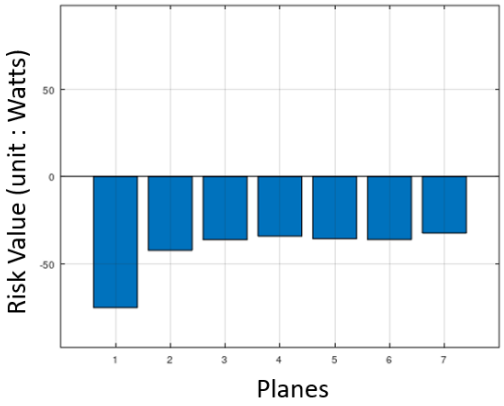

Figure 11. Histogram plot of the risk values of the dies

To perform the power envelope analysis, a thermal simulation tool is needed to consider the powers of functional blocks, environmental conditions, structure and configuration of the device, conditions of the system, and the material properties of all components. After the junction temperatures of the functional blocks are calculated, we can use high-level programming languages, such as Python or Matlab, to calculate the thermal resistance matrix and to plot the results. With the thermal resistance matrix, we can determine whether a selected set of powers of the functional blocks on the SoC is reliable. The computational time of using high-level languages to calculate the risk values is usually within a few seconds making the approach very efficient.

In addition to the advantage of quick determination of the thermal reliability of selected powers, one of the important applications of the power envelope analysis is to evaluate the budget of die powers to determine the most economical one in electricity consumption. Figure 12 illustrates one example of comparing six sets of selected die powers. The selected die powers are [52.5; 10; 3; 3; 10; 10 10] Watts, [32.5; 10; 3; 3; 10; 10; 10] Watts, [12.5; 10; 3; 3; 10; 10 10] Watts, [52.5; 25; 3; 3; 10; 10; 10] Watts, [32.5; 55; 3; 3; 10; 10; 10; 10] Watts, and [42.5; 45; 3; 6; 20; 15; 15] Watts. All these six sets of powers are verified to be safe but we may want to use the third set because it consumes the smallest total power. In real applications, we believe a good power envelope analysis

must also consider the electrical performance, the application conditions, and the thermal reliability of dies.

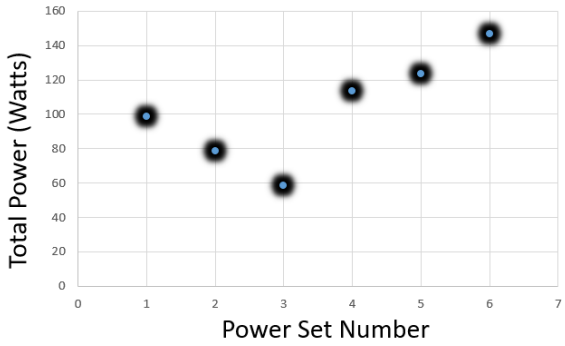

Figure 12. Total powers of dies as a function of set number

V. Cycling Powers and BCI-ROM Method

In addition to the steady state analysis, we have also implemented this power envelope approach on the IC dies having transient and cycling powers. Figure 13 illustrates that the chiplet dies have sinusoidal or pulse power variations. For the transient analyses, the simulation times and computational resources must be considered, and we have used a boundary-condition-independent reduced-order (BCI-ROM) [7-11] method to obtain the temperature responses of dies, as shown in Figure 14. With the known powers and temperatures, we further calculated their root-mean-squares (RMS) values, and used the RMS values to obtain a representative thermal resistance matrix. With the known thermal resistance matrix and a set of chosen powers, the RMS temperatures of the dies can be predicted. Table 6 illustrates the differences of RMS temperatures from the BCI-ROM method and from the thermal resistance matrix. The result is nicely matched.

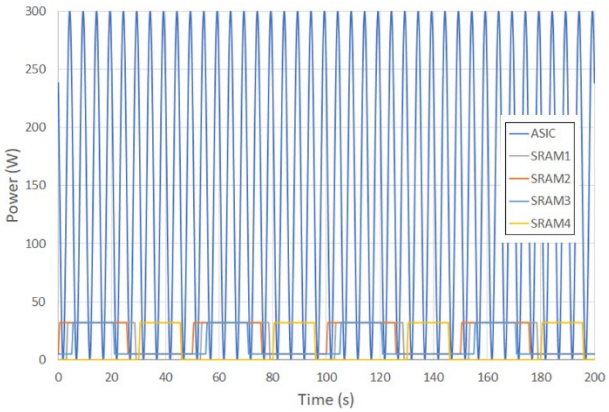

Figure 13. Sinusoidal and pulse power variations for the 5 dies in the chiplet module

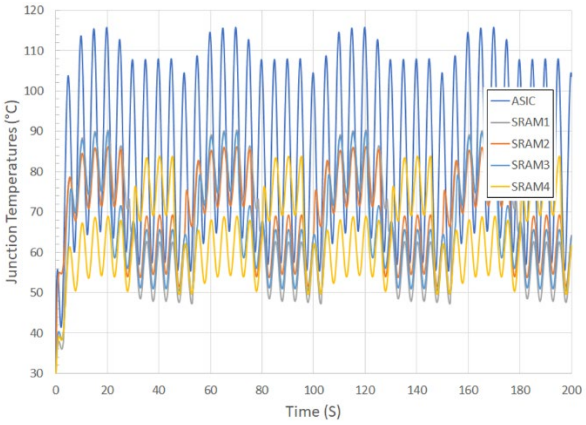

Figure 14. Time-dependent temperatures of 5 dies in the chiplet module.

Table 6. Comparison of the junction temperatures

	RMS Powers	BCI-ROM TJ	R-matrix Tj
Die	**175**	81.501	81.48
SRAM1	**8.25**	57.848	58.21
SRAM2	**25.75**	69.444	69.61
SRAM3	**7.25**	57.742	57.81
SRAM4	**27.75**	70.595	70.48

VI. Conclusions

This paper implements a power envelope approach to determine the thermal reliability of a chiplet module and an SoC package. There are more areas that deserve further researches and studies, and we will list them here: (a) The real loadings or operations of die powers may be transient. The transient behaviors of the dies must be considered in the future for a better implementation of the power envelope analysis. A simple example of cycling powers is given here and we need to consider other forms of transient powers. The BCI-ROM method is to be used extensively to couple with current power envelope analysis. (b) A complicated system may have a combination of many different packages including chiplet, SoC, McM, and other components. The extension of the current power envelope approach to consider all packages and components altogether shall be very valuable. (c) The combined consideration of power management and circuitry application is needed. (d) An industrial collaboration to develop a better power envelope approach may help better control of the powering of the chiplet dies or the functional blocks on an SoC.

Acknowledgment

Authors would like to thank Dr. Voon Hon Wong and Dr. Andras Vass Varnai at Siemens for numerous discussions and collaboration to couple thermal resistance analysis with the BCI-ROM method. We are currently working on a paper to analyze the thermal reliability of IC dies having various

forms of transient powers.

References

[1] Y. Hang, R. Agarwal, E. Ouyang, YH. Jeong, X. Gu, “Thermal Consideration between Monolithic die and 2.5D/3D packages,” OCP Global Summit, November 9-10, 2021.

[2] B. Lall, B. Guenin, and R. Molnar, “Methodology for Thermal Evaluation of Multichip Modules,” IEEE Trans. CPMT, Vol. 18, No. 4, 1995, pp. 758-764.

[3] J. Hong, SB. Park, S. Shao, H. Wang, Y. Niu, VL. Pham, “Design Guideline of 2.5D Package with Emphasis on Warpage Control and Thermal Management,” 68th ECTC, 2018, pp. 682-692.

[4] JT. Chang, A. Gupta, “Estimating the Thermal Interaction between Multiple Side by Side Chips on a Multi-Chip Package,” Electronics Cooling, June 2014, pp. 14-17.

[5] E. Ouyang, X. Gu, YH. Jeong, M. Liu, “Power Envelope analysis for the Thermal Optimization of a Chiplet module,” ASME 2022 International Technical Conference and Exhibition on Packaging and Integration of Electronic and Photonic Microsystems, October 25-27, 2022.

[6] E. Ouyang, X. Gu, YH. Jeong, M. Liu, “Thermal Design of a Chiplet Module using Monolithic Die and 2.5D/3D Packages,” ITHERM 2022, May 31–June 3, 2022.

[7] B. Blackmore, M. Donnelly and M. Alkhenaizi, "Including Electrothermal Effects in Electronics Design with Connected FANTASTIC BCI-ROMs," 2021 37th Semiconductor Thermal Measurement, Modeling & Management Symposium (SEMI-THERM), San Jose, CA, USA, 2021, pp. 78-87.

[8] L. Codecasa, D. D’Amore, P. Maffezzoni, “Parameters for multi-point moment matching reduction of thermal networks,” in Proc. IEEE THERMINIC, pp. 151-154, 2002.

[9] L. Codecasa, V. d'Alessandro, A. Magnani, N. Rinaldi and P. J. Zampardi, "Fast novel thermal analysis simulation tool for integrated circuits (FANTASTIC)," 20th International Workshop on Thermal Investigations of ICs and Systems, Greenwich, UK, 2014, pp. 1-6, doi: 10.1109/THERMINIC.2014.6972507.

[10] L. Codecasa, V. d'Alessandro, A. Magnani and N. Rinaldi, "Matrix reduction tool for creating boundary condition independent dynamic compact thermal models," 2015 21st International Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Paris, France, 2015, pp. 1-5, doi: 10.1109/THERMINIC.2015.7389596.

[11] J. H. J. Janssen and L. Codecasa, "Why matrix reduction is better than objective function based optimization in compact thermal model creation," 2015 21st International Workshop on Thermal Investigations of ICs and Systems (THERMINIC), Paris, France, 2015, pp. 1-6, doi: 10.1109/THERMINIC.2015.7389635.