

Cooling Solutions for Processor Infrared Thermography

Ehsan K.Ardestani, Francisco-Javier Mesa-Martínez, and Jose Renau
Dept. of Computer Engineering, University of California Santa Cruz
<http://masc.cse.ucsc.edu>

Abstract

Temperature is a key parameter due to its impact on timing, energy, and reliability. A setup to measure temperature in runtime with high spatial and temporal resolution would help to study the thermal behavior of processors. Currently, Infrared Thermography infrastructures has been developed to measure the temperature in real-time. Since the infrared opaque metal heat sinks need to be replaced with an infrared transparent heat sink in these setups, oil based cooling solutions have been proposed. However, oil is not a representative of a metal heat sink because measurement with oil based cooling can change the thermal behavior of the processor. In this paper, we discuss a representative oil based cooling solution, and show that it has the same thermal response as a metal heat sink.

I. INTRODUCTION

Temperature has become a first order design constraint for modern high performance processors. Due to the importance of energy and thermal factors, infrastructures have been developed to accommodate models for thermal behavior of the processors. Different methods are introduced to calculate and measure temperature. This is traditionally addressed in three ways: using thermal and power simulation infrastructures; hybrid approaches mixing thermal simulation and direct measurements of architectural state via performance counters; or via direct measurements using on-chip thermal sensor(s) such as thermal diodes.

Validation of these methods demands accurate measurement of real-time response from the processor. As for performance, designers obtain this data using performance monitoring structures such as performance counters. However, this is not the case for temperature. Unlike performance statistics, modern processors lack structures to gather power and thermal metrics. The amount of on-die temperature sensors are not enough for an accurate characterization. Adding sufficient temperature sensors to obtain the needed level of resolution would consume a significant amount of die area.

Recent development of direct temperature measurement through the chips using infrared thermography [1], [2] has provided an infrastructure for validation of the thermal models for processor behavior. However, there are open questions to these measurement infrastructures that needs to be answered. One of the challenges regarding infrared thermography setups is designing a representative cooling solution to replace the conventional IR opaque metal heat sinks.

This paper proposes a solution to the need for representative IR transparent heat sink for IR thermography infrastructures. We discuss how to design an IR transparent heat sink using composite materials and mineral oil. Then we evaluate the performance and thermal characteristics³ of a representative heat sink. Finally, we discuss how to adjust the cooling solution for different power consumptions.

II. THERMAL MEASUREMENT INFRASTRUCTURE

Previous works [1], [2] have developed IR infrastructures to directly measure temperature through the chip. An infrared (IR) camera is used to measure the temperature of transistor junctions. A detailed thermal map is obtained with the infrared camera. Our setup has a resolution of 1024x1024 pixels with sampling rates of over 100Hz. IR camera operates on the 3 – 5 μ m wavelength (MWIR), a range of light where silicon is partially transparent. Silicon has a fairly uniform 55% transmittance from 1.5 μ m to 6 μ m. As a result, the IR camera can measure the temperature through the chip under test. Figure 1 shows a picture with the major components of the measuring setup in [2].

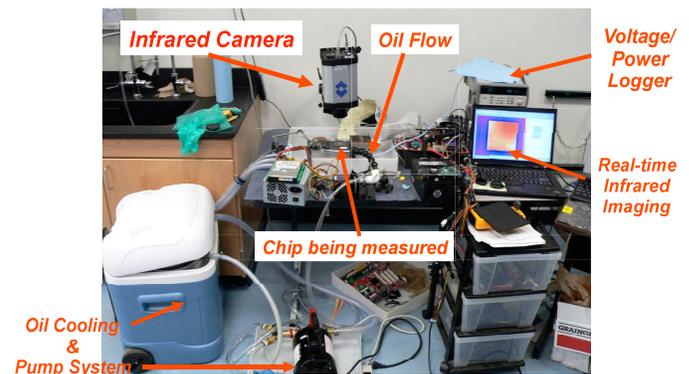


Figure 1. Infrared thermography setup.

Once the IR opaque metal heat sink is removed, an IR transparent heat sink is needed to keep the processor within a safe temperature range. The original IR transparent heat sink proposed in [2] used mineral oil flowing directly over the silicon substrate. The oil-based cooling solution directly applied an oil layer over the die. This paper proposes an alternative IR transparent cooling solution that also uses mineral oil. Mineral oil is a good coolant with an elevated transparency in the infrared spectrum, high specific heat, and rather high thermal conductivity and low viscosity, and chemical safety. The oil being used is designed for infrared spectrography and delivers excellent infrared pictures. Although oil has many advantages, a problem is that it does not have the same thermal response as a metal heat sink [3].

III. REPRESENTATIVE COOLING SOLUTION

Since the metal heat sink is replaced with oil, the validity of the IR thermal measurement setup needs to be addressed. Firstly, oil has a different heat capacitance and thermal resistance than a metal heat sink. The metal heat sink has a different vertical and lateral resistance than oil. In addition, there is a Thermal Interface Material (TIM) between the silicon and the heat sink. Non-uniform thermal resistance raises possible issues with the IR measurement setup [3]. As a result, it could gener-

ate different thermal responses for transient power pulses and even steady-state power response. Secondly, a uni-directional oil flow can result in a variable cooling efficiency across different sides of the die. This can reach to 4°C in the worst case.

We show the steady state and transient response of the oil heat sink, and point out the differences compared to a metal heat sink.

III.1. Equivalent Oil Cooling Solution

To make the oil heat sink more representative of the original metal heat sink, we propose a two fold solution. First, we apply an infrared transparent composite, e.g. Sapphire Window (SW), on top of the die to compensate for the TIM and resistance changes. A sapphire window increases the thermal capacitance and improves lateral heat spreading. Copper has a $400 \frac{W}{mK}$ thermal conductivity while sapphire only has $45 \frac{W}{mK}$. We adjust the cooling system so that the oil plus the sapphire window faithfully represent the metal heat sink. Figure 2 shows the schematic of the system with the sapphire window.

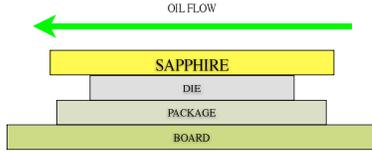


Figure 2. Proposed Infrared transparent heat sink.

Second, we adjust the oil flow to match the cooling performance of the equivalent metal heat sink solution. Once we have a sapphire window, we control the oil volume flow as previously used by [2]. Nevertheless, there are physical limits or lower bounds beyond which the oil flow stops behaving like a laminar flow. To safely avoid non easy to model oil flow artifacts, we fix the oil flow speed to $10 \frac{m}{s}$, and restrict the minimum oil thickness to 1mm to keep the flow laminar.

The other missing factor is the TIM. Typical TIMs have thermal conductivity between 1 and $4 \frac{W}{mK}$. This thermal resistance is placed between the chip and the metal heat sink. For the oil solution, we use oil as an IR transparent TIM. Liquids are effective TIMs but not commercially used because of their short lifetimes.

III.2. Steady State Response

Thermal resistance of the cooling solution determines the steady state response. This is also referred to as the performance of cooling solution, since it determines the final temperature of silicon given a particular power delivered to the chip. The overall resistance is:

$$R_{overall} = R_{Si} + R_{TIM_{oil}} + R_{SW} + R_{conv_{oil}} \quad (1)$$

which should match the overall thermal resistance of the Metal Heat Sink:

$$R_{overall} = R_{Si} + R_{TIM_{MHS}} + R_{MHS} + R_{conv_{air}} \quad (2)$$

Given the equations should match the same $R_{overall}$, we can use different TIM_{oil} liquids ($R_{TIM_{oil}}$), we can adjust the SW

thickness to linearly increase/decrease the vertical thermal conductivity (R_{SW}), and we can control the oil thickness/speed ($R_{conv_{oil}}$).

III.3. Transient Response

Given the time constant proportional to the product of overall capacitance and resistance, we have:

$$\tau_{overall} \propto R_{overall} (C_{Si} + C_{SW} + C_{oil}) \quad (3)$$

$$\tau_{overall} \propto R_{overall} (C_{Si} + C_{MHS}) \quad (4)$$

To keep the transient response the same, equations 4 and 3 should be equal. To do so, we can adjust the oil thickness (C_{oil}) and the SW thickness (C_{SW}). As Table 1 shows, sapphire has 14% less thermal capacitance ($\frac{J}{mK \cdot mm^3}$) which means that a small thickness adjustment is enough. In the following section, we empirically show that our oil and sapphire window solution provides a faithful representative cooling solution compared to the original metal heat sink with air as the coolant.

Material	R ($\frac{W}{m \cdot K}$)	C ($\frac{J}{mK \cdot mm^3}$)
Oil	-	1419
Silicon	120	918
Sapphire	40	2977
Copper	401	3441
Aluminium	250	2435

TABLE 1: Material properties. R and C stand for Resistance and Capacitance.

IV. EVALUATION SETUP PARAMETERS

For our experiments we use a testchip with a $484mm^2$ die area implemented on a BGA GL771 package as shown in Figure 3. The power consumption for each block of the chip can be independently controlled. In our experiments, we power up block P with different power densities. The area of the block is $4.84 mm^2$. A thermal diode in each block senses temperature with sub-millisecond thermal responses. With this testchip, we can evaluate the proposed cooling solution against different metal heatsinks and for different power densities.

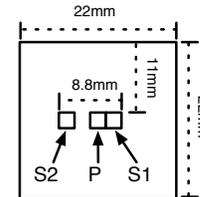


Figure 3. Testchip floorplan.

In order to satisfy the requirements, we use an oil-sapphire combination. A 3mm thick sapphire window with a 50mm diameter is placed on top of the chip. This window is then cooled by a flow of mineral oil, which is also transparent to IR and removes the heat. The mineral oil is temperature-controlled by a heat exchange maintaining the oil between 15°C to 20°C. This

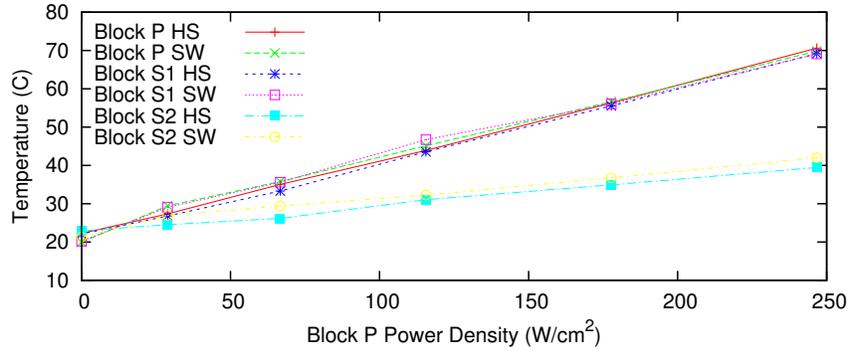


Figure 4. Temperatures for blocks P, S1, and S2 with different constant power in block P. HS and SW stand for AMD Mobile heat sink and sapphire window respectively.

allows the oil to remove heat from the sapphire window optimally, while maintaining a laminar flow with a speed of around $10 \frac{m}{s}$. We use a mineral oil with a specific heat of $1.63 \frac{J}{gK}$.

V. EVALUATION

V.1. Steady State Response Analysis

Different cooling solutions can have different thermal capacitance and thermal resistance. For steady-state analysis, only the thermal resistance affects the cooling efficiency of the system. If we apply a constant power, the thermal capacitance does not have any effect. The electrical capacitors do not have any effect when a constant voltage is applied. Similarly, the thermal capacitance does not affect the temperature for a constant power.

Figure 4 shows the temperature for blocks P, S1, and S2 in the test chip when different steady-state power consumptions are applied to block P. The oil flows from top to bottom. When Block P gets powered from 0 to $250 \frac{W}{cm^2}$, we measure a consistent linear increase in temperature for the heat sink (HS) and the sapphire window (SW). This implies that the overall vertical thermal resistance of the oil and the metal heat sink are very similar.

To validate the lateral thermal resistance, we measure blocks S1 and S2. S1 is adjacent to block P, and therefore its temperature is very close to P when using either the heat sink or the sapphire solution. However, the temperature difference between blocks P and S2, placed $5mm$ away from each other, increases as the power density increases. This is because the lateral thermal resistance creates thermal gradients on the die. We observe that both heat sink and sapphire have a consistent slope.

This simple experiment also validates the behavior of the sapphire window as an alternative cooling solution to a metal heat sink with similar vertical and lateral thermal resistances.

V.2. Oil Flow Direction

To compensate for the different cooling efficiency across the die due to the direction of the oil flow, the IR setup performs an additional image correction over each captured frame. In the worst case, we observe a maximum temperature gradient of $4^\circ C$ between opposite sides of the test chip. This corresponds to approximately $0.2^\circ C$ correction for each mm that the oil flows over a hot block.

To further reduce the gradients due to oil flow, it is possi-

ble to add a diamond heat sink on top of the sapphire window which increases the lateral resistance. Alternatively, we explore a software correction mechanism.

If all the blocks are uniformly heated, applying the $\frac{0.2^\circ C}{mm}$ correction is a simple and effective alternative. However, real chips do not display such uniform temperature across their dies. Ideally, a model describing the fluid dynamics of the oil should be used to perform the oil flow correction. However, this solution is too compute-intensive especially considering the $4^\circ C$ worst case. Instead, we have a quick approximation estimating the oil flow correction. For every mm that the oil flows over a block, we adjust the correction by $0.2 * \frac{BlockTemp - 45^\circ C}{10^\circ C}$. We never let the correction be negative. This is a simple algorithm with linear cost that provides a fast and effective solution.

Block	Top-Bottom	Left-Right	Right-Left	Bottom-Top
P	64.9 (65.3)	64.9 (65.3)	64.9 (65.3)	64.9 (65.3)
S1	63.9 (63.9)	64.7 (65.4)	63.6 (63.6)	63.5 (63.5)
S2	48.2 (48.2)	48.1 (48.1)	47.8 (48.6)	48.2 (48.2)

TABLE 2: Oil flow direction impact. Uncorrected value in parenthesis.

To evaluate the accuracy of the correction and the impact of the oil flow, we repeat the same experiment as in Figure 4. This time we apply the oil from four possible directions when block P is powered with $7.8W$. The values in parenthesis are the uncorrected values obtained when the oil flow correction algorithm is not applied. The only blocks affected by the oil flow direction are S1 and S2 when we have a horizontal flow. Without correction, the maximum error is $1.8^\circ C$ (S1 with Left-Right flow). After the correction, the error is reduced to $0.9^\circ C$.

V.3. Transient Response Analysis

As reported by [3], oil flow has a lower thermal capacitance and it has a faster response showing clearer thermal phases. The previous section has shown that a sapphire window reduces the thermal resistance difference between oil and the heat sink. Sapphire also affects the thermal capacitance because it has double the specific heat of copper ($0.75 \frac{J}{g*K}$ vs $0.385 \frac{J}{g*K}$) but approximately half the density. The overall material properties are shown in Table 1. As a result, copper and sapphire have an equivalent thermal capacitance. In conclusion, sapphire window has a more attenuated thermal response than Oil.

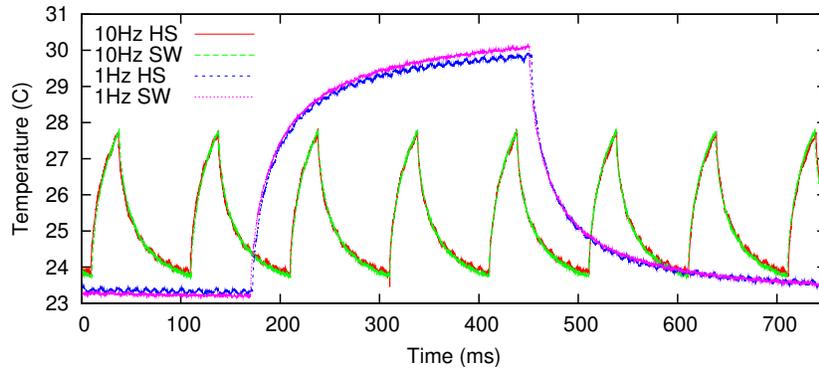


Figure 5. Thermal transient response for a test chip when a 25 ms power pulse is applied periodically every 100 ms (10Hz). HS and SW stand for Heat Sink and Sapphire Window respectively.

In our experimental validation, we use the test chip which provides μs sampling capabilities. Figure 5 shows block P temperature when a periodic power pulse is applied to it and the rest of the chip is idle. The same power pulse is applied for a heat sink “HS” and a sapphire window “SW”.

The 25ms power pulse is applied every 100ms (10Hz). We clearly observe that the thermal transients of the heat sink and the sapphire are very close. [3] pointed that an oil cooling solution with this power pulse would have a significant error for fast transients. The measured results show that a sapphire window solves the problem. We also perform 250 ms power pulses to validate the equivalency between the heat sink and the sapphire window for slower transients. Again, the proposed cooling solution closely matches the heat sink solution.

Combining the fast and slow transient response accuracy with the cooling efficiency validation for the vertical and lateral thermal resistances, we conclude that the oil cooling solution with a sapphire window is an appropriate vehicle to capture existing thermal phases.

VI. HEAT SINK DESIGN

Having shown that an oil based cooling solution can be implemented to represent a metal heat sink, the next question is how to adjust the heat sink for different power consumption. As discussed earlier, the parameters that can be played with to change the characteristic of the heat sink are $R_{TIM_{oil}}$, R_{SW} , $R_{conv_{oil}}$, C_{oil} , and C_{SW} . Instead of Sapphire Window (SW), it is possible to use other IR transparent materials like Silicon and diamond. While Silicon would be effective for worse quality heat sinks, diamond would be effective for very efficient heat sinks. Although not necessary for our evaluation, it is possible to have a composite material that integrates layers of diamond, sapphire, and/or silicon.

For performance of the cooling solution, however, only thermal resistance is important. It is only resistance that affects steady state thermal behavior, and determines the temperature. Given the overall thermal resistance, the max power that the heat sink can support can be computed as follows:

$$P = \frac{T_{max} - T_{amb}}{R_{overall}} \quad (5)$$

which T_{max} is the maximum temperature that junctions can safely operate at, T_{amb} is the ambient temperature and $R_{overall}$ is

overall thermal resistance.

VII. CONCLUSIONS

In this paper, we discuss how to design an oil-heatsink representative of metal heatsink to be used in IR thermography infrastructures. We show that composite materials such as a Sapphire window along with mineral oil can build up an efficient cooling solution, with the same steady state and transient response as metal heat sink. We also address the problem of different cooling efficiency due to oil flow direction. We discuss a set of post-processings to make the oil-heatsink applicable. Finally, we discuss how to configure the parameters of the cooling system to adjust for the targeted power consumption, ranging from embedded systems to high performance processors.

ACKNOWLEDGMENTS

We like to thank the reviewers for their feedback on the paper. Special thanks to Sai Ankireddi and Sun Microsystems for their testchip. This work was supported in part by the National Science Foundation under grants 0546819, 0720913, and 0751222; Special Research Grant from the University of California, Santa Cruz; Sun OpenSPARC Center of Excellence at UCSC; gifts from SUN, nVIDIA, Altera, Xilinx, and ChipEDA. Any opinions, findings, and conclusions or recommendations expressed herein are those of the authors and do not necessarily reflect the views of the NSF.

REFERENCES

1. H.F. Hamann, J. Lacey, A. Weger, and J. Wakil, “Spatially-resolved imaging of microprocessor power (SIMP): hotspots in microprocessors,” in *Thermal and Thermo-mechanical Phenomena in Electronics Systems, 2006*. May 2006, pp. 121–125, IEEE Computer Society.
2. F.J. Mesa-Martinez, J. Nayfach-Battilana, and J. Renau, “Power model validation through thermal measurements,” in *ISCA '07: Proceedings of the 34th annual international symposium on Computer architecture*, New York, NY, USA, 2007, pp. 302–311, ACM.
3. W. Huang, K. Skadron, S. Gurumurthi, R.J. Ribando, and M.R. Stan, “Differentiating the roles of ir measurement and simulation for power and temperature-aware design,” in *ISPASS, 2009*, pp. 1–10.