Ambient variation-tolerant and inter components aware thermal management for mobile system on chips

Francesco Paterna\textsuperscript{1}, Joe Zanotelli\textsuperscript{2} and Tajana Simunic Rosing\textsuperscript{1}

\textsuperscript{1}Dept. of Computer Science and Engineering University of California, San Diego email: \{fpaterna\texttildelow tajana\}@eng.ucsd.edu

\textsuperscript{2}Qualcomm Technologies, Inc. email:josephz@qti.qualcomm.com

Abstract—In this work we measure and study two key aspects of the thermal behavior of smartphones: 1) thermal interaction between the components on the printed circuit board and 2) the influence of phone's ambient temperature which is subject to large variations. The measurements on the smartphone running typical workloads show that the heat generated by the communication subsystem and the high temperatures on the back cover of the phone can increase the SoC temperature by as much as 17\textdegree C. None of the run-time thermal management studies presented to date considered this interaction, as there was no model available. We design a thermal model that captures this thermal dependency and a policy able to avoid thermal emergencies while minimizing the impact on performance.

I. INTRODUCTION AND RELATED WORK

Smartphones are composed by several ICs components (e.g. SoC, DRAM, RX/TX) and devices (e.g. camera, display) which work simultaneously and generate heat. The absence of active cooling devices along large ambient condition variations creates a complex thermal behavior. More specifically, significant thermal interactions occur between SoC, communication module, audio codec, DC-DC converter, battery and display as a function of the changing ambient conditions which are consistent on the back cover of the phone.

Recent publications model the thermal behavior of the multiprocessor systems on chips (MPSoCs) while assuming a stable ambient temperature next to it and disregarding the rest of the system\cite{[1][2][4][5][6][7][8]}. A few studies to date attempted to go beyond just SoC thermal modeling. In [5] a smartphone and a tablet are disassembled and the thermal properties, such as thermal conductivity and heat capacity of the package including cover, display and battery are measured. The thermal study of the entire working phone has not been conducted. In [6], power and energy analysis of the different components of a smartphone such as CPU, memory and display under different usage conditions such as voice call and message texting, is performed. Similar analysis is done in [3]. Here the authors also take into account the power dissipated by the voltage regulator and DC-DC converter. Their main goal is to provide an infrastructure to evaluate the battery lifetime.

In this paper we first analyze the effect of the ambient condition variations and inter-component dependency on the temperature of the SoC and then we build an efficient and accurate thermal model. We finally use our model to devise a policy aware of the undesired effects on the SoC to adjust the clock frequency with the aim of meeting thermal constraints while minimizing the impact on performance. In our experiments, we adopt one of the recent commercial smartphones and use an industrial profiler to collect performance and thermal statistics. We also used thermistors connected to a digital analyzer to measure the temperature on the back cover of the phone under different conditions such as phone lying on a desk and phone held in a hand. We use a set of benchmarks characterized by 2D/3D rendering and arithmetic computations to exercise the SoC. As inter-component dependency, we focus on the RX/TX since it results a significant heat source along with the SoC. The contribution of this paper is threefold and can be summarized as follows.

As first contribution, we show through experiments on our target phone that the combined effect of ambient condition variations and thermal interaction between SoC and RX/TX can increase the SoC temperature by as much as +17\textdegree C. If thermal increase is avoided by slowing down computing, the resulting performance loss is as much as 27%.

As second contribution, we propose a thermal model of the smartphone systems. Our model is composed by two sets of functions. The first one is able to estimate the thermal increment on the back cover of the phone over time in two different conditions: phone lying on a desk and phone held in hand. We suppose to be able to catch the changing of the condition by exploiting the face detector sensor mounted on the front cover of the phone: if the user is looking at the screen, it is more likely he is holding the phone in his hand. The other set of functions is able to predict the SoC temperature after a period as long as 10 sec by exploiting the estimated back cover temperature and the knowledge about the RX/TX activity, the clock frequency and the current SoC temperature read by the sensor. We validate our model against real measurements obtained from our target phone over different contexts in terms of ambient conditions and RX/TX activity. Our results show that the average error of the model is below 1\textdegree C.

As third contribution we devise an ambient variation-tolerant and inter components aware thermal management technique able to meet thermal constraints set up for the SoC while minimizing the impact on performance. We prove through experiments on our target phone that our approach has adaptive properties with respect to the ambient condition variations and the inter-component dependency.

The rest of this paper is organized as follows. In Section II we detail our experimental setup. In Section III we evaluate the impact of the ambient condition variations and the inter-component dependency on the SoC temperature. In Section IV we present our thermal model and our policy. In Section V we provide experimental results about the model.
validation and the policy effectiveness. Section VI concludes this paper.

II. EXPERIMENTAL SETUP

This paper’s key contributions hinge on experiments performed on a smartphone to show and quantify the effects of the ambient conditions and the activity of the RX/TX module on the SoC performance. Thus, we start by describing the phone we used and the experimental setup.

A. Target phone characteristics

Our target phone is Snapdragon™ smartphone by Qualcomm® [11]. Its printed circuit board (PCB) has components on both sides. On the top of the PCB there are the MSM8660 SoC and the Qualcomm® QTR8615 RX/TX communication module. The key components on the SoC are two ARM15 [13] CPUs and Adreno™ GPU [12].

We collect detailed power, performance and thermal data using the Trepn profiler [10]. Profiler’s sampling resolution is 100ms. We collected power measurements of the SoC’s two ARM15 application CPUs, a GPU and DRAM as well as the clock frequency of the CPUs and the GPU. We set the phone in the airplane mode and activate the WiFi which has always been connected to our network. The phone is first placed on a desk during the first phase of experiments and then is handheld. We also performed another set of experiments by placing the back of the phone 10 cm above a 400W lamp to increase the temperature surrounding the phone and hence mimic what happens if the phone is handheld over long periods. We measure the temperature on the back of the phone by using two thermistors [14] and the analyzer USB-6216 by National Instruments [15]. We capture the traces by adopting LabView [16]. The locations on the back cover of the two thermistors were in correspondence of the SoC and the RX/TX respectively as these components are mounted on the PCB inside the phone. We average the values provided by the two thermistors.

B. Benchmarks

As benchmark for the experiments, we use the 2D and 3D rendering samples and the Math samples found within the 0xBenchmark [17] suite. The 2D set is composed by 7 benchmarks detailed as follows.

1. **GL Cube.** A sample program from the Android API Demo that uses OpenGL ES to render a rotating Rubik’s Cube.
2. **GL Teapot.** It uses OpenGL ES to render a rotating Utah Teapot.
3. **NeHe Lesson08.** A rotating 3D cube with textured applied and alpha blending enabled.
4. **NeHe Lesson16.** A rotating 3D cube with textured applied and GLFog enabled.

The Math set is composed by 7 benchmarks detailed as follows.

1. **Linpack.** It makes use of a general numerical linear algebra operation to test the device’s ability to perform floating point operations (MFLOPS).
2. **Scimark2.** It includes six scientific calculations (Composite, FTT, SOR, MonteCarlo, SparseMatMult and LU) for testing the device’s ability to perform floating point operations (MFLOPS).

We ran these three sets of benchmarks in different conditions with respect to the temperature on the back of the phone and the WiFi activity. To exercises the WiFi, we launched a downloading from Gmail of a 400Mbyte file via the ONE Browser app [18]. We used this size to make sure that the downloading lasts as long as the benchmark executions. We were not able to activate the 3g in our phone.

III. THERMAL EFFECTS ON THE SMARTPHONE SYSTEM

In this section, we analyze through real measurements the impact of the ambient condition variations and the WiFi activity on the SoC temperature and the consequent performance loss.

In the first experiment, we use the rendering benchmarks. We have a 2D and a 3D set of execution samples. The 2D set takes roughly 2 min while the 3D 2.5 min. Each experiment is characterized by the execution of the 2D or the 3D set multiple times in sequence. Each experiment is repeated four times in order to create four different contexts specified as follows. 1) Back cover temperature 30°C; 2) Back cover temperature 30°C while WiFi is downloading; 3) Back cover temperature 45°C; 4) Back cover temperature 45°C while WiFi is downloading. Please note 45°C is usually the thermal constraint on the back cover of the phone since it is the maximum temperature the human skin can tolerate [21]. We report the SoC temperature and the performance evaluated in frame-per-second in Tables I, II in which each row refers to the execution of one 2D/3D set over the sequence. For both the 2D and the 3D analysis, the WiFi makes the SoC temperature increase up to 8°C and the performance drop by as much as 22%. When the back cover temperature reaches 45°C, the SoC temperature increases by as much as 17°C.
For the second experiment, we use the Math suite. We repeatedly execute the set of computational samples for 13 times. Each experiment has been done for three frequencies \{1180MHz, 864, 540\} for the four different contexts described in the first experiment.

We adjust the clock frequency of the CPUs by using the Android Tuner app [20]. We capture the SoC temperature at 0.1sec granularity. Figure 1 shows the temperature over time. First row: back cover is 30°C. Second row: cover is 45°C. First column: no downloading. Second column: downloading. When downloading the maximum SoC temperature increases up to 9°C. We analyzed these data to build a thermal model of the SoC temperature with respect to the back cover temperature, the CPU frequency and the WiFi activity. This study is reported in the next section.

For the third experiment, we analyze the back cover temperature increase over time. Figure 2 shows the back temperature increase over 36 minute executions. When the phone is held in a hand, the temperature quickly reaches the human skin temperature 36°C, after then the increase is linear. We report the trendline which shows that the increment is as much as 0.001°C/sec. The usage of the WiFi introduces an overall increase by +1°C. A similar behavior happens when the phone is lying on a desk. The increment is as much as 0.002°C/sec.

<table>
<thead>
<tr>
<th>Temp (°C)</th>
<th>OpenGL Cube</th>
<th>OpenGL Blending</th>
<th>OpenGL Fog</th>
<th>Flying Teapot</th>
</tr>
</thead>
<tbody>
<tr>
<td>38</td>
<td>60.04</td>
<td>64.24</td>
<td>64.28</td>
<td>60.60</td>
</tr>
<tr>
<td>43</td>
<td>60.08</td>
<td>64.18</td>
<td>64.09</td>
<td>60.59</td>
</tr>
<tr>
<td>44</td>
<td>60.11</td>
<td>64.12</td>
<td>63.87</td>
<td>60.60</td>
</tr>
<tr>
<td>45</td>
<td>60.11</td>
<td>64.22</td>
<td>64.23</td>
<td>60.60</td>
</tr>
<tr>
<td>38</td>
<td>60.14</td>
<td>64.04</td>
<td>63.80</td>
<td>49.44</td>
</tr>
<tr>
<td>47</td>
<td>60.08</td>
<td>64.28</td>
<td>63.67</td>
<td>47.61</td>
</tr>
<tr>
<td>53</td>
<td>60.05</td>
<td>63.79</td>
<td>63.85</td>
<td>49.44</td>
</tr>
<tr>
<td>54</td>
<td>60.12</td>
<td>64.12</td>
<td>64.00</td>
<td>51.58</td>
</tr>
<tr>
<td>38</td>
<td>59.97</td>
<td>60.13</td>
<td>60.15</td>
<td>60.12</td>
</tr>
<tr>
<td>55</td>
<td>64.29</td>
<td>64.09</td>
<td>64.12</td>
<td>64.16</td>
</tr>
<tr>
<td>59</td>
<td>63.92</td>
<td>64.15</td>
<td>64.25</td>
<td>64.11</td>
</tr>
<tr>
<td>59</td>
<td>53.41</td>
<td>55.55</td>
<td>55.54</td>
<td>55.55</td>
</tr>
<tr>
<td>38</td>
<td>60.11</td>
<td>61.12</td>
<td>60.1</td>
<td>60.14</td>
</tr>
<tr>
<td>57</td>
<td>64.1</td>
<td>63.94</td>
<td>64.28</td>
<td>63.95</td>
</tr>
<tr>
<td>61</td>
<td>63.96</td>
<td>64.11</td>
<td>63.99</td>
<td>64.13</td>
</tr>
<tr>
<td>55</td>
<td>51.27</td>
<td>51.28</td>
<td>49.45</td>
<td>49.45</td>
</tr>
</tbody>
</table>

Table II: SoC temperature while running a sequence of the 3D suite. Performance is reported for each benchmark by frame-per-second (fps).

![Figure 1](image1.png) **Figure 1** Execution of the Math suite for 13 times in a row. Plots show the SoC temperature over time. First row: back cover is 30°C. Second row: cover is 45°C. First column: no downloading. Second column: downloading.

![Figure 2](image2.png) **Figure 2** Analysis of the back cover temperature variation over time for different contexts. Linear trendlines reported: \(y(°C) = m \cdot t(\text{sec}) + q\).
IV. THERMAL MODEL AND MANAGEMENT

A. Model

We define our thermal model as a function able to estimate the SoC temperature after a fixed time by the knowledge about the current SoC temperature supposed read by a sensor, the CPU frequency, the temperature on the back cover and the WiFi activity (i.e. whether it is downloading). In particular the model is able to estimate how the thermal increment per second changes with respect to the factors we mentioned above. The model we propose is based on the observations related to the analysis in the previous section. We also validate its accuracy in the results of the next section.

Since in the practical usage, we are not able to measure the temperature on the back cover, we need a model also for it. We based on the third experiment presented above. The back cover temperature changes very slowly with respect to the SoC temperature and we can take into account this behavior by assuming a linear trend as this well shapes the curves for the all four cases of Figure 2. We can discriminate between phone handheld in which the increment is as much as 0.001°C/sec and phone lying on desk in which the increment is as much as 0.002°C/sec. To know whether the user is holding the phone in his hand, we can base of what the face detector on the front cover captures; in particular if the user is looking at the phone, probably the phone is handheld. Also, when the ambient conditions change we can add a ±5°C in according to whether the phone passes from the desk in a hand or vice versa. Hence, we define the model for the increment of the back cover temperature $\Delta T_a$ in (1) in which $t$ is the time in sec elapsed from the previous observation.

$$\Delta T_a = \begin{cases} 0.001 \cdot t & \text{hand (+5 when switch)} \\ 0.002 \cdot t & \text{desk (-5 when switch)} \end{cases} \quad (1)$$

We can now proceed with the model of the SoC temperature with respect to the CPU frequency. We do not model the GPU since we are not able to change the GPU frequency in our phone. As in the second experiment of the previous section, we collected thermal traces for the SoC at two different back cover temperatures and two different types of WiFi activities for three different CPU frequencies. We report in Figure 3 the four different trends over the CPU frequencies. For the all four cases which differ in terms of WiFi activity and back cover temperature, the trend over the clock frequency can well approximated by exponential curves. The idea could be to understand how the parameters of the exponential function depend from the back cover temperature and the WiFi activity in terms of data transferring type (i.e. download/upload) and signal. We are not able to act on the data transferring rate and signal, hence we only discriminate whether the WiFi is downloading or not. Also we did not consider the upload as we assume that effect is comparable to the download. For each of the four cases, we associate a different function to model the SoC temperature expressed by $y[^\circ C/sec] = l \cdot \exp(h \cdot f [MHz])$. We assume that the coefficient $l$ changes linearly with respect to thermal gradient on the back cover. Instead, we assume that the $h$ coefficient is constant when the WiFi is not working; we average the values obtained at the two back cover temperatures. When the WiFi is downloading we also assume a linear variation of the coefficient $h$ with respect to the thermal gradient on the back cover. The resulting functions are reported in Equation (2). The final SoC temperature is calculated through Equation (3), in which the SoC thermal increment calculated through (2) is multiplied by the slot time $\Delta t$ and then added to the temperature read by the sensor.

$$\Delta T_{SoC}(t) = \begin{cases} (19 + 5 \cdot \Delta T_a) \cdot 10^{-4} \exp(13 \cdot 10^{-4} - f) & \text{no downloading} \\ (15 + 11 \cdot \Delta T_a) \cdot 10^{-4} \exp(20 - 0.8 \cdot \Delta T_a) \cdot 10^{-4} - f) & \text{downloading.} \end{cases} \quad (2)$$

$$T_p(t + \Delta t) = \Delta T_{SoC}(t) \cdot \Delta t + T_c(t) \quad (3)$$
B. Policy

We leverage the models we developed to design a policy that improves performance of the CPU as a function of ambient conditions and RX/TX activity without causing violations of SoC thermal constraints. Our policy is designed for the Math benchmark suite illustrated above. This set of benchmark results no memory bound which means that the execution time of each singular samples linearly increases with respect to the frequency. For instance for Linpack the MFlops are \{14.28, 30.61, 48.24\} at \{540, 846, 1180 MHz\} respectively. We do not report the same information for all the suit for space limitation. In addition the power linearly scales with respect to the frequency too; for the same frequency group the average power results \{0.10, 0.17, 0.23 W\} respectively. We conclude that 1) from the energy point of view there is no preferred frequency 2) from the performance point of view the higher the better 3) from the thermal point of view a maximum allowed frequency may exist and may vary over time. We then devise a technique which increases as much as possible the CPU frequency over time by avoiding thermal constraint violations. Before giving details about the policy we define the adopted variables as follows.

- \( f \): CPU frequency (\( f_{\text{max}} \): max freq. - \( f_{\text{min}} \): min freq.);
- \( Tp \): predicted SoC temperature;
- \( WiFi \): a Boolean value to specify whether the WiFi is downloading;
- \( Ta \): back cover temperature;
- \( Tc \): actual SoC temperature read by the sensor;
- \( gr \): time granularity of the time slot \( \Delta t \);
- \( T^* \): SoC temperature constraint;
- \( \text{model}() \) : function implementing the thermal model.

Figure 4 illustrates the proposed algorithm. Each period of time \( \Delta t \), the current back cover temperature is estimated through (1), the WiFi activity is sampled, the SoC temperature is read from the sensor. The frequency which is the output of the algorithm is set to its maximum value. All this information composes the variable \( x \) which is used to estimate the SoC temperature after the time \( gr \) through (2) and (3). The predicted temperature is compared with the thermal constraint and as long as that value results larger, the frequency is decreased by one step in according to the possible settings. If the predicted temperature is smaller than the threshold or the frequency has reached the minimum possible value, the algorithm exits and the frequency is set to the CPU.

V. Results

A. Model validation

In this first part of results we validate our thermal model presented above. To prove that our model well captures ambient variations and inter component interactions we execute the Math suite continuously for 7500sec. During this time we changed ambient condition by holding the phone in a hand or letting it lying on a desk. We also download a file for some periods and change the CPU frequency. How we divided the time slots is specified in Figure 5. We collect our thermal traces related to the SoC and the back temperature with a granularity by 0.1sec. Our model estimates the SoC temperature at the next instant by avoiding thermal constraint violations. The average error is less than 1°C and the maximum is below 4°C; a summary is reported in Table III.

![Figure 5 Measured vs. estimated SoC temperature over time.](image)

<table>
<thead>
<tr>
<th>Time (sec)</th>
<th>Back 36°C</th>
<th>Back 36°C + down.</th>
<th>Back 30°C</th>
</tr>
</thead>
<tbody>
<tr>
<td>( t_0 )</td>
<td>756 45 540 45 756 45</td>
<td>1180 48 1180 52 1180 47</td>
<td>1180 48 1180 50 1180 49 1180 49</td>
</tr>
<tr>
<td>( t_1 = 100 )</td>
<td>918 45 1180 52 1180 47 1180 48 1180 44 1180 49</td>
<td>918 47 1180 52 1180 48 1180 44 1180 49</td>
<td></td>
</tr>
<tr>
<td>( t_2 = 500 )</td>
<td>918 45 1180 52 1180 48 1180 44 1180 49</td>
<td>918 47 1180 52 1180 48 1180 44 1180 49</td>
<td></td>
</tr>
<tr>
<td>( t_3 = 600 )</td>
<td>1180 50 540 48 918 47</td>
<td>1180 50 540 48 918 47</td>
<td></td>
</tr>
</tbody>
</table>

Table IV Runtime policy acts to increase the CPU frequency as much as possible while holding the CPU temperature below 50°C. Three different context related to the back cover temperature and the WiFi activity.

B. Policy Effectiveness

In this experiment we prove the effectiveness of our strategy in managing the CPU performance in different contexts of ambient conditions and RX/TX activity. We run in our phone the Math benchmark suite and execute in a row all the samples whose details are reported above. We check the SoC temperature each 100 sec through the \( \text{adb} \) shell [19] on our desktop computer. Indeed our phone is connected to our computer via usb. We implemented the policy algorithm and the models we illustrated above in Matlab. The Matlab routines we implemented calculate the new frequency of the CPU which will try to avoid the thermal violation while impacting as less as possible the performance. The routines, as explained in the policy section, also need the information about the temperature on
the back cover of the phone and the activity status of the WiFi; i.e. whether it is downloading. We adjust the clock frequency of the CPUs by using the Android Tuner app. The possible clock frequencies are: {1180MHz, 1130, 1080, 1020, 972, 918, 864, 810, 756, 702, 648, 594, 540).

We use three different contexts to prove the two properties of our management technique: 1) ambient-variation tolerance and 2) inter component awareness. In the first one the back phone starting temperature is 36°C; our phone was in particular held in a hand. The second is as the first but we also activate a file downloading from Gmail by using ONE Benchmark. The file size is 400 Mbyte and we repeatedly launched the downloading to cover the all length of the experiment. In the third context, the phone was lying on a desk and the back phone starting temperature is 30°C. The SoC thermal constraint is 50°C.

We start to activate our observations when the SoC temperature is stable to 45°C. Table IV shows, for the different contexts, the frequency set by the policy and the consequent SoC temperature after 100 sec running; the policy then decides the new frequency reported at the row below. We only report the first seven traces since the behavior is actually recurring. For the first context we can see that the frequency is set to the maximum value and when the temperature is 49°C to avoid the thermal violation the frequency is decreased to 918MHz. One time slot later, the high performance can be resumed. When downloading the file the behavior is more critical, any time the policy decides to raise up the frequency to the maximum value the temperature will slightly violate the constraint but then the policy is able to reestablish the thermal safety by decreasing the frequency accurately. In third context, in which the phone is lying on desk and then the back temperature is cooler, our policy is able to capture the possibility to hold the maximum frequency for a longer period without violating the thermal constraint.

VI. CONCLUSION

In this paper we face the problem of the thermal modeling and management of mobile system on chips. Using an experimental setup composed by a commercial smartphone running typical graphic and computational benchmarks we show that the temperature variation on the back of the phone and the WiFi activity can increase the SoC temperature up to 17°C. If thermal increase is avoided by slowing down computing, the resulting performance loss is as much as 27%. To face this problem we first elaborate a thermal model and then a policy which is able to contrast the ambient variations and the inter-component dependency by acting on the computational speed of the SoC in an efficient way. We validate our model and prove the proposed technique through experiments on our target phone.

VII. ACKNOWLEDGEMENTS

This work was supported by Qualcomm, SRC (Task ID 2169.001) and NSF variability (g.a. 1029783).

REFERENCES