Application Report

How to Boost Your CPU, GPU, and SoC Performance Through Thermal Accuracy

David Vaseliou

ABSTRACT

Whether on a smartphone, Personal Computer (PC), or an everyday laptop notebook, it is important to provide maximum processor performance. Technology is advancing quickly every day, and the need for faster processors is becoming even more important. The processing power of your device is typically determined by the Central Processing Unit (CPU) and Graphic Processing Unit (GPU). A major limiting factor of CPUs and GPUs are their thermal design point.

Table of Contents

1 Introduction......2
2 What is a CPU, GPU, and SoC?......2
3 Temperature vs. Performance......2
3.1 Benefits of Faster Processors......2
3.2 Overclocking and Underclocking (Clock Throttling)......3
3.3 Cooling Systems......3
4 Benchmarking......3
4.1 Benchmarking Data......3
5 How to get Accurate CPU/GPU Temperature......7
5.1 Using Local Temperature Sensor......7
5.2 Using Remote Temperature Sensor......8
6 Summary......8

List of Figures

Figure 4-1. Smartphone: Benchmark Scores vs Temperature......4
Figure 4-2. Gaming PC: CPU Clock Rate vs Temperature......4
Figure 4-3. Gaming PC: CPU Benchmark vs Clock Rates......5
Figure 4-4. Gaming PC: CPU Benchmark vs Temperature......5
Figure 4-5. Gaming PC: GPU Clock Rate vs Temperature......6
Figure 4-6. Gaming PC: GPU Benchmark Score vs Clock Rate......6
Figure 4-7. Gaming PC: GPU Benchmark Score vs Temperature......7
Figure 5-1. Local Temperature Sensor Placements......8
Figure 5-2. Remote Temperature Sensor Application Schematic......8

Trademarks

All other trademarks are the property of their respective owners.

1 Introduction

CPUs and GPUs can get extremely hot and going above their thermal design point can cause irreversible damage. To avoid these devices from overheating cooling systems such as heat sinks, fans, and sometimes expensive liquid cooling systems, are integrated into the system. Along with these cooling systems, thermal safety margins are implemented to ensure the processing units do not exceed their limits. When the temperature of a CPU or GPU approaches the thermal limit, the clock speed will throttle to protect the CPU or GPU, degrading its performance. Due to inaccuracies of the temperature measurements, the safety margins must be larger and the processor will start to slow its performance earlier than necessary. Because of this, getting more accurate temperature data is crucial to optimizing a processors performance, which maximizes the overall user experience.

2 What is a CPU, GPU, and SoC?

A CPU is the main chip on the motherboard or main board of a PC, smartphone, or other electronic device. Each CPU has three main components: an Arithmetic Logic Unit (ALU), a Control Unit (CU), and memory unit. Today, it is common to find CPUs that have multiple cores, typically between 2-16, each having their own ALU, CU, and memory. The CPU is given instructions from a program or application and executes it. These instructions can be tasks such as basic arithmetic computations, numeric comparisons, or memory movement and storage. How fast these instructions can be processed is limited by the number of cores and clock rate of the CPU. The clock rate of a CPU is measured in GHz and is the number of instructions a CPU can execute per second. Therefore the faster the clock rate the faster the CPU can process instructions and the greater the performance of the CPU.

A GPU is a processing unit that is specifically designed to handle graphics. Similar to a CPU, a GPU consists of multiple cores that have ALUs, CUs, and memory units. The main difference is that GPUs are architected to have hundreds to thousands of cores. This makes them specially designed for handling parallel throughput computing. How fast a GPU can process data is also limited by its clock rate.

A system on chip (SoC) is an integrated circuit that consists of multiple electrical components on a single platform. The CPU and GPU are usually integrated into a SoCs. SoCs are much smaller than multi-chip designs found in PCs or notebooks and consume less power, but are typically much slower. Because of their smaller size and power consumption they are generally used in smartphones, tablets, and other mobile devices.

3 Temperature vs. Performance

To understand why CPUs and GPUs overheat we have to look at these devices at the transistor level. There can be up to 100 million transistors per square millimeter on a processing unit. Each transistor acts as a switch to allow or stop current from flowing through it. When transistors switch, they go through a more resistive state. Any current passing through a resistance will generate I^2R losses which generates heat. At higher clock rates the transistors goes through the more resistive state more often which creates more heat. This self-heating can damage the processor if not monitored accurately.

3.1 Benefits of Faster Processors

A better performing processor yields many advantages in both smartphones and PCs. A faster processor can, reduce latency, process information faster, and allow the device to run sophisticated applications. With the integration of augmented reality, artificial intelligence software, and high-resolution graphics into smartphones, a faster processor is crucial for a fast and smooth user experience. Common tasks such as loading webpages, watching videos, and playing games will also be faster and more enjoyable with a faster processor. Similar benefits can be seen with better performance on a PC. PCs though are also typically used for research, evaluation, and development as well. This means that running applications faster can reduce development time and increase productivity.

3.2 Overclocking and Underclocking (Clock Throttling)

3.2.1 Overclocking

Overclocking is the practice of running the clock rate of a processor faster than what the system was designed. This limit is typically determined by the maximum temperature ratings with appropriate safety margins. While overclocking can speed up the processor, it will also generate more heat. In many cases extra cooling and thermal management is necessary when overclocking a processor. While this may sound dangerous to the processor, overclocking is a common practice. With proper thermal management and accurate temperature monitoring, overclocking can be done safely.

3.2.2 Underclocking (Clock Throttling)

Similarly, underclocking or "clock throttling" is the practice of running the clock rate of a processor slower than it is designed. This is used to reduce system power consumption when higher performance is not needed, and used as a safety measure to reduce heat generated by the processor when higher performance is needed. Most processors will automatically underclock when the temperature approaches its thermal limit with safety margins. In this case, the processor reduces its performance to stay within its thermal limits. More accurate temperature measurements allow a system to get closer to its thermal design limit, thus reducing the amount of underclocking and optimizing performance.

3.3 Cooling Systems

Due to the self-heating nature of processors, cooling systems such as heat sinks, fans, or liquid cooling systems are usually necessary. While these systems allow the processor to run faster, they can take up a lot of space, use a lot of power, and can be very loud. As the temperature of the processor increases, the fan speed or liquid pump increases. This will increase the power consumption of the system as well as generate more audible noise to the displeasure of the user. More accurate temperature measurements can be used to keep the processor within a desired temperature range with minimal power consumption and audible noise from the cooling system.

4 Benchmarking

Benchmarking is the practice of comparing performance metrics to a consistent standard or baseline. In regards to benchmarking a CPU or GPU, a consistent task or workload is given to the processor and different metrics are measured. These metrics are then compared to a standard baseline and a score given based on how well the processor performed. Some common metrics that are measured for processors include clock rate, latency, execution time, and throughput. Frames per second (FPS) is also a common metric specific to benchmarking a GPU.

4.1 Benchmarking Data

To show the relationship between clock rate, temperature, and performance, a series of benchmark tests were performed on both a smartphone and gaming PC. During these tests, the clock rates of the devices were limited at different rates and the temperatures and benchmark scores were measured.

4.1.1 Smartphones

The below data was collected on a smartphone device. The performance of the smartphone was limited at different levels and a benchmark test was performed. Figure 4-1 shows that, as the GPU and CPU performance increases, the temperature of the smart phone also increases because of the increased clock rates. This shows that, as the processor approaches the thermal design point, throttling the device at a lower temperature can affect the benchmark score by around 2% per °C.

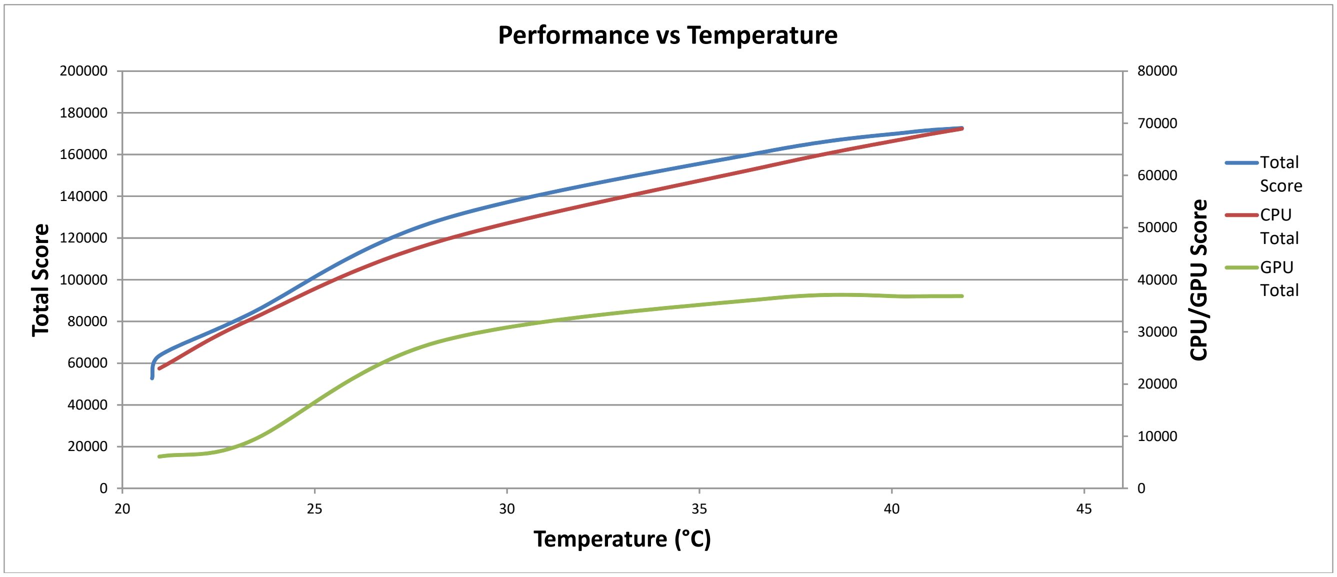

Figure 4-1. Smartphone: Benchmark Scores vs Temperature

4.1.2 Gaming PCs

The below data was collected on a gaming PC. Two different benchmarks were used to exercise the CPU and GPU. The benchmark tests were limited at different clock rates and the temperature and benchmark scores were measured.

4.1.2.1 Gaming PC CPU Performance

Figure 4-2 shows that the temperature of the CPU increases when the clock rates increase.

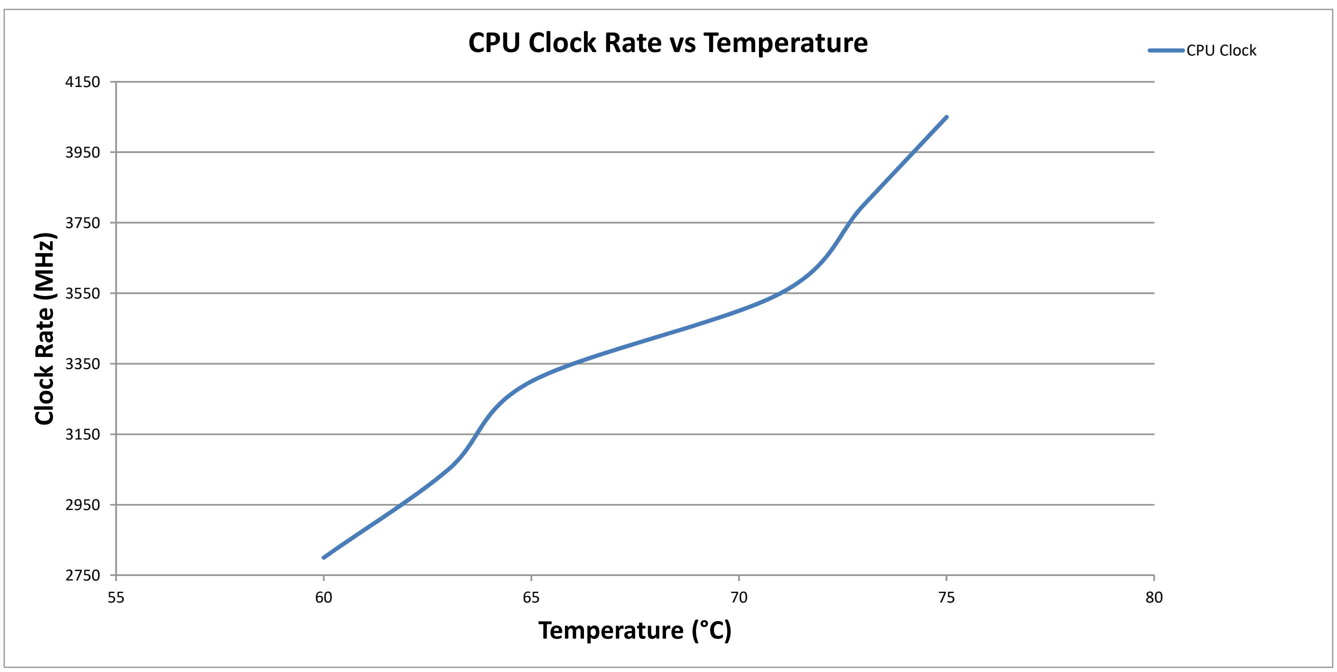

Figure 4-2. Gaming PC: CPU Clock Rate vs Temperature

Figure 4-3 shows that the benchmark performance of the CPU increases when the clock rates increase.

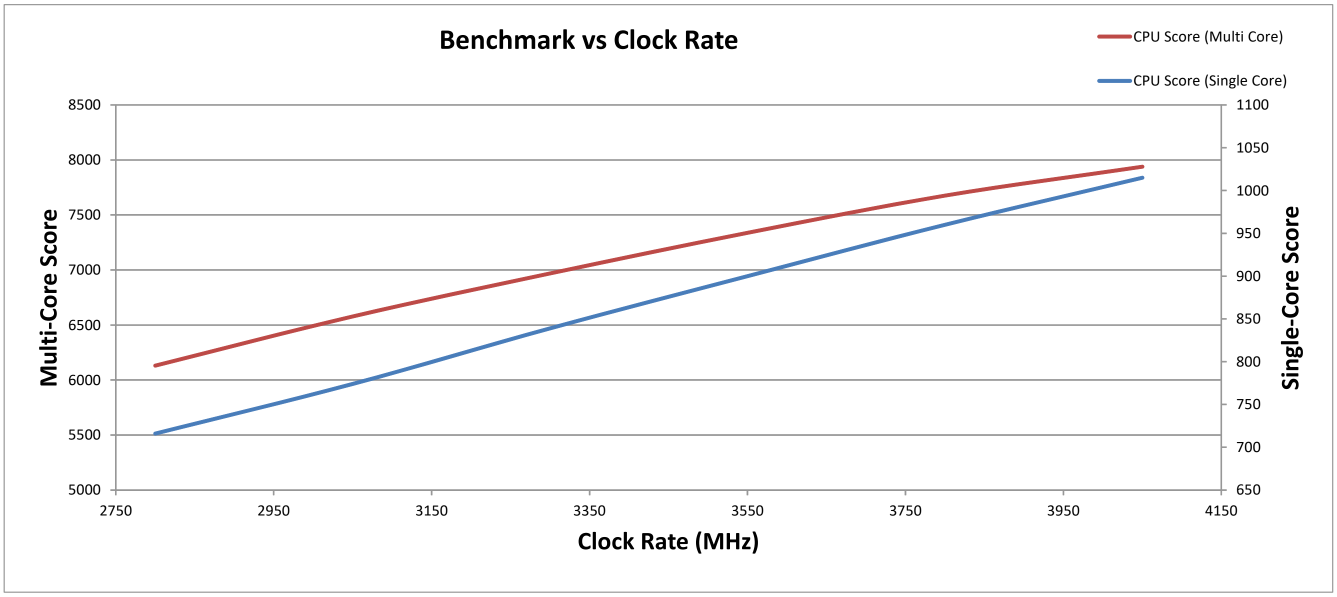

Figure 4-3. Gaming PC: CPU Benchmark vs Clock Rates

Lastly, Figure 4-4 shows that the temperature of the CPU increases when the clock rates the benchmark scores increase.

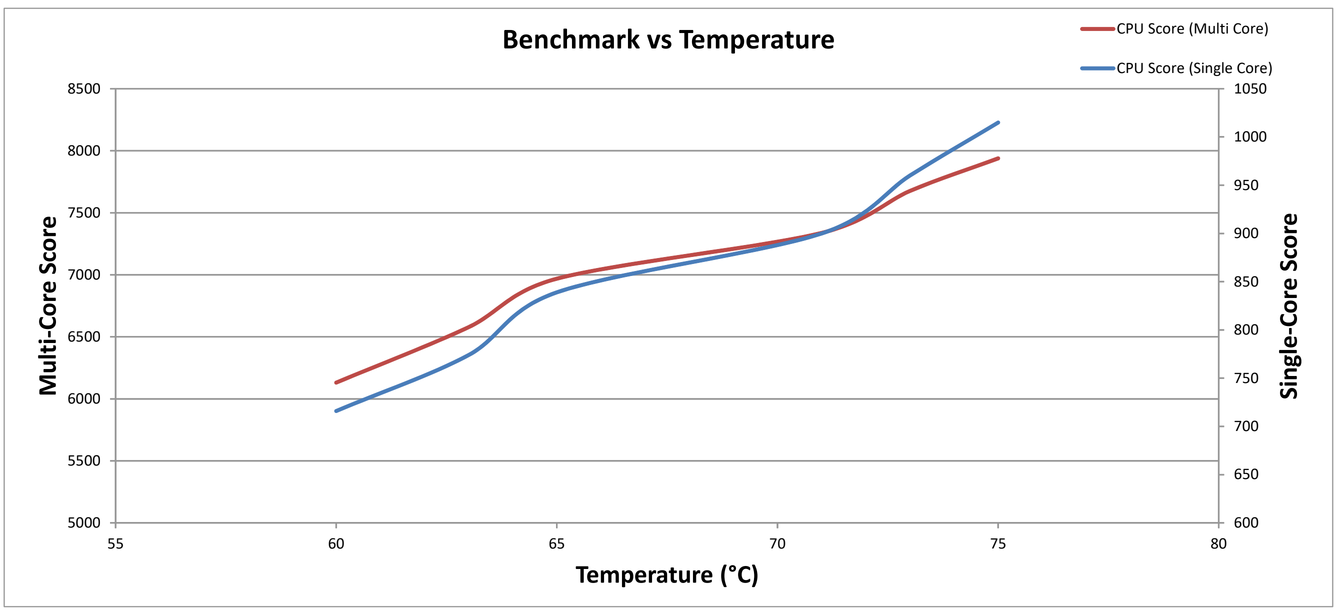

Figure 4-4. Gaming PC: CPU Benchmark vs Temperature

From these figures, a 2 °C smaller safety margin can improve the benchmark score by about 4.0%.

4.1.2.2 Gaming PC GPU Performance

Figure 4-5 shows that the temperature of the GPU increases when the clock rates increase.

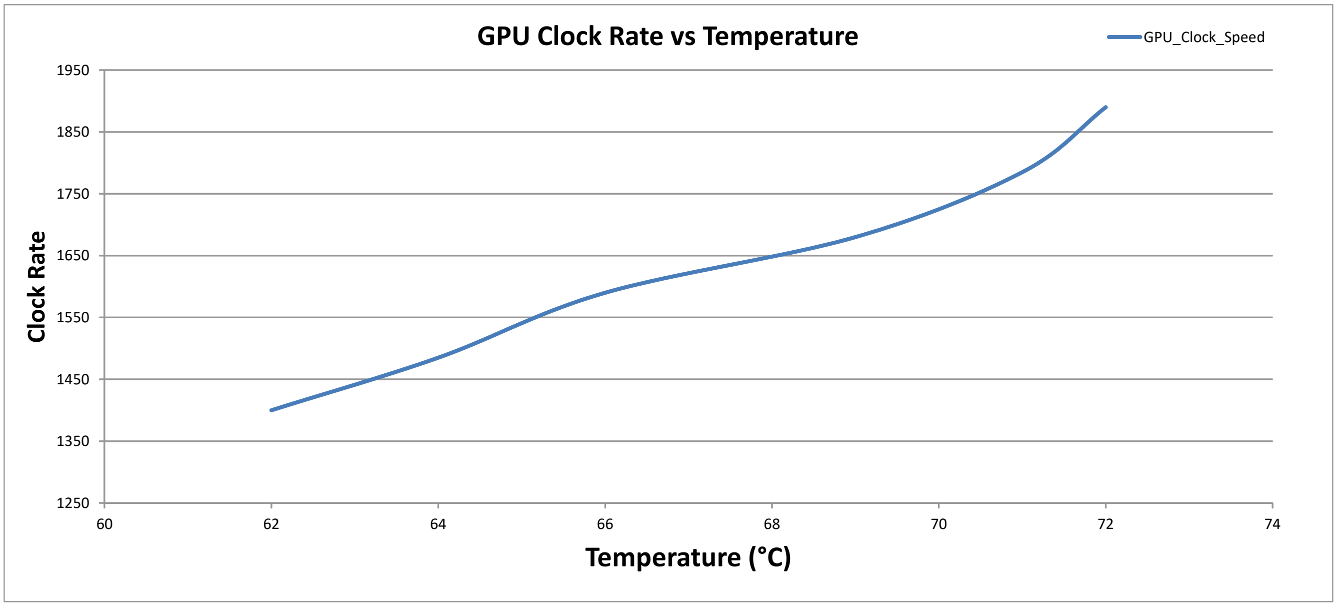

Figure 4-5. Gaming PC: GPU Clock Rate vs Temperature

Figure 4-6 shows that the benchmark performance of the GPU increases when the clock rates increase.

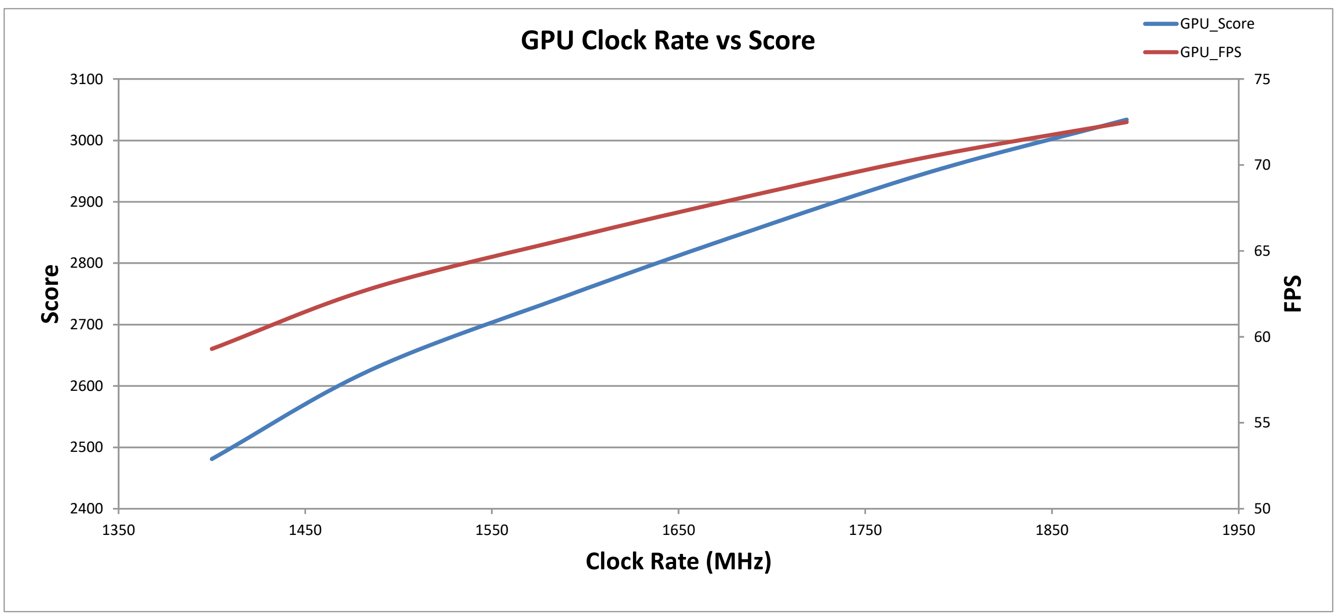

Figure 4-6. Gaming PC: GPU Benchmark Score vs Clock Rate

Lastly, Figure 4-7 shows the temperature of the GPU increases when the clock rates benchmark scores increase.

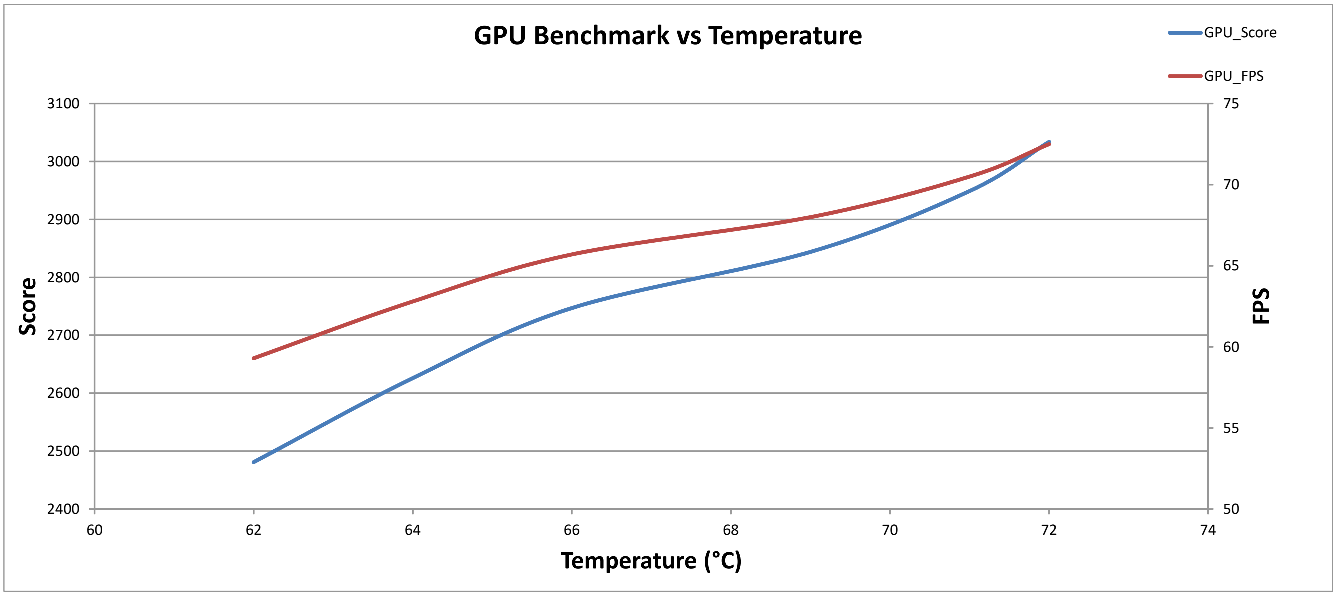

Figure 4-7. Gaming PC: GPU Benchmark Score vs Temperature

From these figures, a 2 °C smaller safety margin can improve the FPS and benchmark score by about 6.0%

5 How to get Accurate CPU/GPU Temperature

There are many different methods to monitor processor temperature. NTC thermistors are commonly used by placing them close to the processor with a good thermal path between the devices. However, these devices can be very inaccurate and can increase the necessary thermal safety margin. Local temperature sensors, however, can be very accurate and integrated in to the system using the same methods as an NTC. Some local temperature sensors are extremely accurate and can be rated up to 0.1 °C accuracy.

Another method is to use an integrated temperature sensor if the processor has one. Similar to NTCs, this method can be extremely inaccurate (as high as ±10 °C). For that reason, many processors have dedicated pins to use remote temperature sensors. These dedicated pins are connected to a thermal diode directly on the die of the processor. The remote temperature sensor measures the temperature of the die directly, therfore the sensor can be extremely accurate.

5.1 Using Local Temperature Sensor

Local temperature sensors measure their own die temperature to determine the temperature in a specific area. Therefore, it is important to understand the dominant temperature conduction paths between the die and the processor. Heat is conducted primarily through two paths: through a Die-attach pad (DAP) attached to the package, or through the package lead pins. To make an effective thermal path to the local temperature sensor from the processor, the devices should be placed as close as possible to the processor through a solid ground plane. Figure 5-1 shows the different locations to place a local temperature sensor to get high accurate temperature measurements.

Location A shows a sensor located in a hole drilled into the center of a heat sink of the CPU or GPU. The heat sink can be clipped to the processor or attached with epoxy, and generally sits on top of the processor. Another location for accurate processor temperature monitoring is in the cavity beneath a socketed processor (location B). Given the sensor is isolated from the airflow, ambient temperature has minimal impact on the sensor reading. Additionally, if the heat sink gets detached from the processor, the sensor indicates an increase in processor temperature. Lastly, location C shows a sensor mounted on the circuit board next to the processor. While this is simple to implement, the correlation between sensor temperature and processor temperature is much weaker.

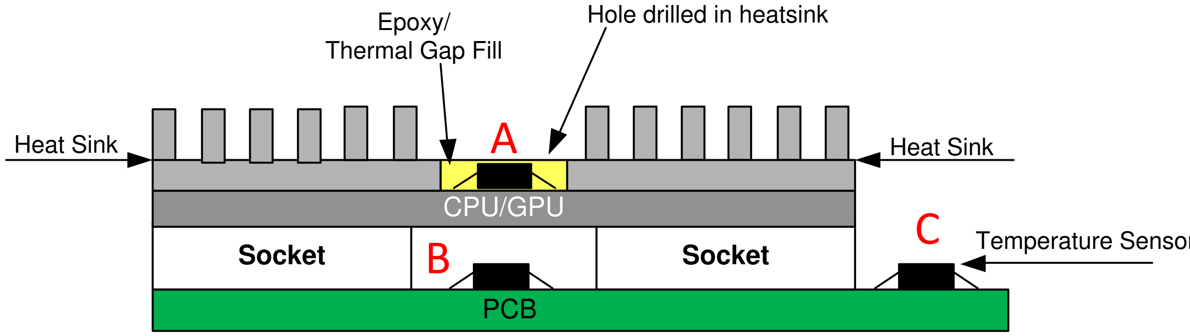

Figure 5-1. Local Temperature Sensor Placements

5.2 Using Remote Temperature Sensor

When the CPU or GPU has thermal diode pins available, a remote temperature sensor can be used. Remote temperature sensors apply different currents and read back the change of voltage on the processor pins. The temperature of the junction can be determined through a change in voltage. When using remote temperature sensors, the electrical path to the processor is important because electrical noise on these signal pins can cause temperature errors. Other sources of errors include ideality factor variation, series resistance, and low beta junctions. Some remote temperature sensors have built-in features to limit the temperature error from these error sources. Figure 5-2 shows a typical application schematic of using a remote temperature sensor with a CPU or GPU.

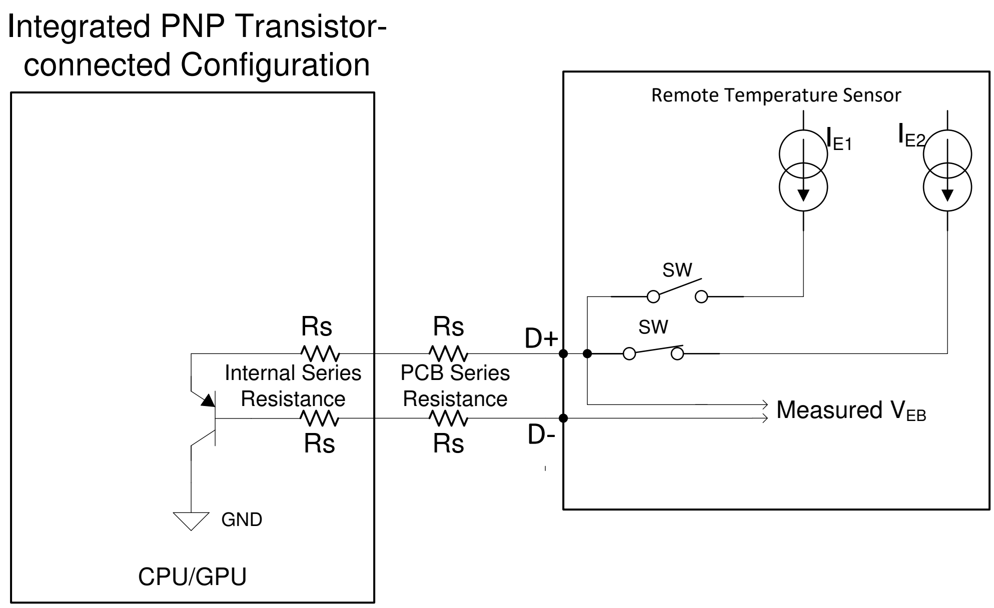

Figure 5-2. Remote Temperature Sensor Application Schematic

6 Summary

In summary, CPUs and GPUs generate a lot of heat during operation. Often large, loud, and expensive cooling systems are added to these devices to keep them within their temperature limits. Due to built-in safety margins under the inaccuracies of the temperature monitoring systems, these devices will throttle the clock rate, which will decrease the CPU or GPU performance and drive cooling systems harder. Accurate temperature monitoring for CPUs and GPUs protect processors from thermal damage, can cut down on cool system costs and power consumption, and optimize device performance. Throttling the processor just 2 °C earlier can lessen the benchmark scores by up to 4% on smartphones and 6% on PCs. Also, using remote temperature sensors or highly-accurate local temperature sensors ICs allows the user to optimize CPUs and GPUs performance and ensure safe operation.