Dynamic Thermal Management of Processors Using Thermoelectric Coolers

By Sriram Jayakumar

Submitted in partial fulfillment of the requirements of the degree of Bachelor of Science with Honors in Computer Engineering in the School of Engineering at Brown University.

April 2013

Prepared under the direction of:

Professor Sherief Reda, Advisor

_____ Date _____

Professor Sherief Reda, Advisor

_____ Date _____

Professor Ruth Iris Bahar, Reader

_____ Date _____

Professor Jennifer Franck, Honors Chair

Acknowledgements

I want to thank my advisor, Professor Sherief Reda. I have known Professor Reda throughout all my four years at Brown University. In particular, I want to thank him for teaching me to demand high quality work from myself, when it is required. My parents, sister, and family have provided motivation, support, and love. Without them, this thesis would not have been possible. Finally, all the members of the SCALE lab at Brown University have helped me along the way, and I appreciate it.

Contents

This page is intentionally left blank.

Abstract

The goal of this thesis is to design and implement a microprocessor thermal management system that uses thermoelectric coolers (TECs). Starting in 2006, microprocessors moved from having a single core to containing multiple cores. TECs fit into this trend because they are reliable solid-state devices, have small form factors, and can potentially fit over individual cores of microprocessors, allowing for targeted hotspot cooling. Hotspots thermally constrain performance in modern CPUs. Even when not used for hotspot cooling, TECs can still provide performance benefits by allowing the CPU to operate at a higher frequency. While running an application, if the processor hits the maximum allowable temperature, the TECs can cool the chip down and allow further increases in processor frequency and performance. The disadvantage of TECs is that they provide active cooling and consume power in order to pump heat.

This thesis takes an experimental approach, and constructs an actual cooling system with TECs and a modern 22nm-process quad-core processor. The thesis develops a controller for the TECs that optimizes performance while meeting given temperature and power constraints. As an example of the controller's performance benefits, when running the gcc benchmark while maintaining a temperature constraint of 45 C and a 40 W power budget, the TEC system controller decreases runtime by 8% compared to a controller that only scales processor frequency to meet the constraints. However, the TEC controller consumes 64% more energy. A drawback of the system overall is that it only cools up to 30 W of CPU power because of the limitations of the specific TECs used. Also, since the TEC performance benefits do not come for free, this thesis provides a quantitative analysis of the power-performance tradeoff.

In contrast to prior work that uses simulations, or tests systems with static sources of power, this work tests the implemented TEC system with real benchmarks which have varying power consumption over time. This is the central contribution of the thesis. The transient behavior of the system while it is being controlled is observed.

1. Introduction

A *thermoelectric cooler (TEC)* is a solid-state cooling device. Applying an electric current to the cooler's two terminals pumps heat from one side of the device to the other. Figure 1 shows a block diagram of a TEC, and Figure 2 shows a real TEC. A *thermoelectric generator (TEG)* is the cooler's twin. Both are built from the same underlying technology. However, a generator performs the opposite function and converts the heat flowing from the hot side of the device to the cold side to electrical energy.

An exciting application of thermoelectric generators is generating power during space voyages. For example, the Mars Curiosity rover uses thermoelectric generators combined with a plutonium heat source to produce about 110 W of power, at an efficiency of around 7%. While technologies with moving parts would provide higher efficiencies, they are typically not as reliable. Curiosity and previous space missions used TEGs despite the low efficiency because the TEGs are unlikely to fail during operation. It may take many years for a failure to occur [1], [2]. Thermoelectric generators have potential in producing power from waste heat, for instance, the waste heat from an automobile engine [3, Sec. Abstract]. The pattern of using thermoelectrics in niche applications is common. TECs cool laser diodes, telecommunications equipment, and picnic coolers [4]. The latter application comes about because TECs can have small form factors, and, as mentioned previously, have no moving parts.

The focus here is on using TECs for thermal management of electronic microprocessors. TECs are a refrigeration technology. In a standard two-phase liquid-vapor refrigerator, a substance cycles through the system, transferring heat from the insides of the compartment to the surroundings. The details of the compressor, condenser, expansion valve, and other components are beyond the scope of this discussion [5], [6]. The bottom line is that a TEC also pumps heat, but is a solid-state device.

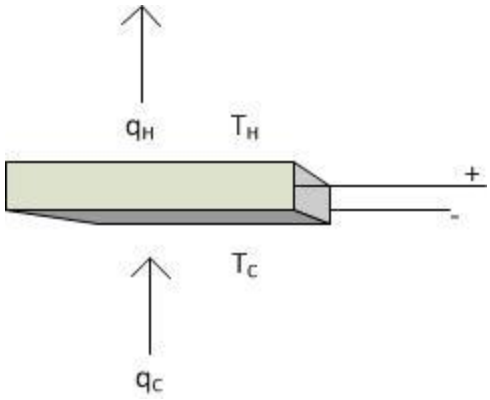

Figure 1 TEC Block Diagram. This block diagram of a TEC shows the heat drawn from the cold side and the heat pumped out the hot side. T_c and T_h represent the cold and hot side temperatures. A positive electric current should be applied from the + terminal to the – terminal.

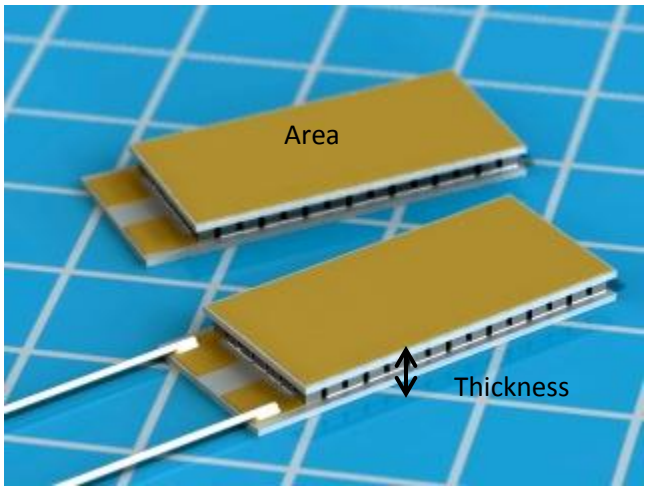

Figure 2 TEC Picture. The picture shows a TEC from the manufacturer used in the experiments conducted in this thesis [7]. The individual thermoelectric couples inside the device are visible. The area of the TEC refers to the area of the plates. Thickness is also shown.

As mentioned in the Abstract, part of the motivation for TECs is that microprocessors have moved from having a single core that can execute instructions to containing multiple cores that can execute instructions in parallel. Adding cores is a viable strategy because it provides additional performance without producing excessive power and requiring unwieldy cooling solutions. Adding cores also is a tractable design problem. As transistor sizes shrink and provide more useable die area, seeing the number of cores increase will be common. A consequence of having multiple cores is that chips are more likely to have localized temperature hotspots. If only one out of four cores is active on a chip, the

die area apportioned to that core will be hotter. The solid-state nature of the TECs and their hotspot cooling ability makes TECs more attractive than other cooling methods like liquid cooling, for instance.

TECs are also an interesting area of investigation because the performance of the TECs themselves appears to be improving, especially with thin-film TECs. Although TECs are not currently a cost-effective or necessary technology for microprocessor thermal management, in the future they might become viable. Therefore, it is worthwhile understanding how such a TEC thermal management system behaves.

The two key contributions of this thesis are listed below.

- The tradeoff between TEC power consumption and processor performance is quantitatively characterized on a real TEC setup that uses bulk TECs to cool a modern quad-core processor. For processors with frequency scaling, the TEC can enable increases in frequency while still meeting a junction temperature constraint. However, the particular TEC used in this thesis is not efficient from the perspectives of an energy or energy-delay metric.
- The work here develops a controller for the TEC system that varies CPU frequency setting and TEC current to meet CPU temperature and total system power constraints. When tested on actual processor workloads while meeting temperature constraints, the controller, in one instance, gives an 8% decrease in runtime compared to a controller that only uses frequency scaling.

Regarding thesis structure, the *Introduction* chapter leads the reader through thermal modeling and the physics of TECs. The *Related Work* chapter elaborates on recent work on TEC modeling, system construction, and control. The third chapter, *Modeling*, discusses possible setups for the TEC system and finishes with the development and experimental verification of an accurate steady-state model for the TECs. The final chapter of new content, *Dynamic Thermal Management*, has two parts. The first part characterizes the cooling capability of the TEC system by comparing it to a system without TECs (termed

the *"standard system"*). This first part also discusses how TECs couple with *DVFS (dynamic voltage and frequency scaling)* to provide increases in performance, though this comes at the cost of power. DVFS lets the processor trade off performance, power, and temperature by providing various clock frequency settings to operate at.

The second part of the fourth chapter culminates the thesis. It constructs a simple feedback controller that optimizes processor performance while meeting temperature and power constraints. The basic procedure of the controller is to use the TEC to lower the junction temperature below the temperature constraint, and then to increase CPU frequency. If the TEC were unavailable, increasing frequency would push temperatures over the limit. The controller provides significant improvements in runtime compared to a controller that can only use DVFS to meet the constraints. The controller typically only allows, at worst, brief 5 C temperature constraint violations.

Thermal Modeling

Conduction and radiation are the primary forms of heat transfer. Convection is considered a combined mode of heat transfer [8, p. 45]. This section considers conduction and convection.

The concept of thermal resistance rests upon empirically known physical laws. Regarding conduction, consider a slab of homogenous material, illustrated in Figure 3, with a cross-sectional area A and length L.

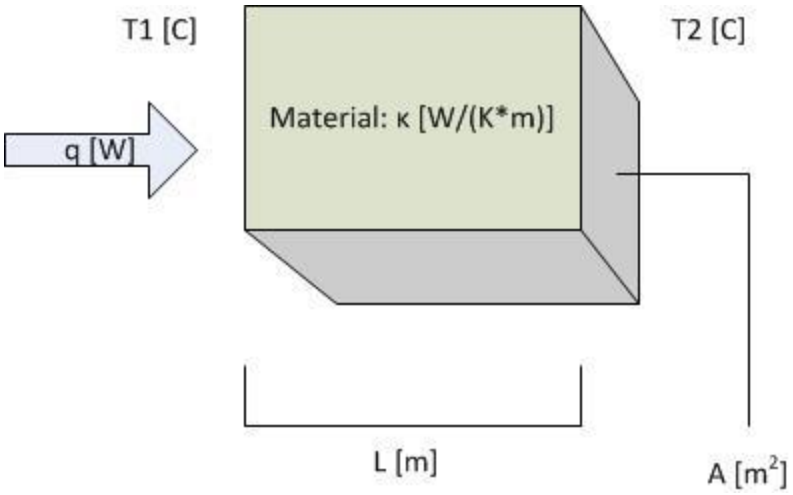

Figure 3 Fourier's Law

Let both sides of the material be at fixed temperatures, T_1 and T_2 respectively. Assuming unidirectional heat flow, that L is relatively small compared to A, and steady state,

$$q = -\left(\frac{\kappa * A}{L}\right) * (T_2 - T_1)$$

This is Fourier's Law. q is the rate of heat flow, in W, from side 1 to side 2. Suppose T_2 is at a higher temperature than T_1. Then, heat will be flowing from side 2 to side 1, so the minus sign is necessary. In differential form, the law becomes $q_x = -\kappa * A * \frac{dT}{dx}$ [9, p. 1].

- T_1 is the temperature of side 1, in C. T_2 is the temperature of side 2, in C.
- A is the cross-sectional area of the material, in m^2. (Consider facing the material in the direction of heat flow.)
- L is the length of the material, in m.
- κ is the *thermal conductivity* of the material, in $\frac{W}{K*m}$. The thermal conductivity is assumed to be independent of temperature [10, p. 657]. The thermal conductivity of silicon at 300 K (around room temperature) is 156 W/(K*m). The value reported is for a single crystal of silicon; silicon wafers used to make microprocessors are crystalline silicon, so applying the value to the case of

the wafer is justifiable [11, p. A1062]. The thermal conductivity of copper at room temperature ranges between 339 and 391 W/(m*K) [12]. Values are listed below in Table 1.[1] Nickel is listed because the Intel integrated heat spreader is nickel-plated copper [13, p. 18]. Aluminum and silver are other commonly encountered metals. In particular, the thermal interface material contains silver.

Material	**Thermal Conductivity W/(K*m)**
Silver	419
Copper	(339,391)
Aluminum and its alloys	(117,234)
Silicon	156
Nickel and its alloys	(15,62)
AS5 Thermal Interface Material [14, p. 2]	8.9

Table 1 Thermal Conductivities for Various Materials. Parentheses indicate a range of values.

Electronics are often cooled by forcing air or a liquid over the relevant devices. Hence, convection, which describes the situation of a fluid flowing over a body, is important. Newton's law of cooling is $q = -h * A * (T_f - T_b)$ [8, p. 47].

- T_f is the temperature of the fluid in C. T_b is the temperature of the body in C.
- q is the rate of heat flow in W from the body to the fluid.
- h is the heat transfer coefficient, in $\frac{W}{K*m^2}$. h will typically be higher for liquids.
- A is the area of the body contacted by the fluid, in m^2.

Both Fourier's law and Newton's law of cooling are of the form $I = \frac{\Delta V}{R}$, which is similar to Ohm's law. The rate of heat flow is like electrical current, in the sense that both flow through objects. Likewise, temperature is similar to voltage because both are measured across quantities [15, p. 39]. Therefore, it

[1] The values listed in the reference source's [12] tables are in (Btu/hr)/(ft*F). There are 0.2931 W per Btu/hr. There are 9/5 F/C. There are 3.281 ft/m. So, convert to W/(m*K). The values are reproduced in [9, p. 12].

is useful to define the notion of thermal resistance. The rate of heat flow from point 1 to point 2 is $q = \frac{\Delta T}{R}$, where q is in W, the temperature difference $\Delta T = T_1 - T_2$ is in K, and R is in K/W. The reciprocal of thermal resistance R is the thermal conductance K, in W/K. In the case of Fourier's law, $K = \frac{\kappa * A}{L}$. In the case of Newton's law of cooling, $K = h * A$. (For the form of the equations given in this thesis, the minus signs are eliminated in Fourier's law and Newton's law of cooling to fit them to the definition of thermal conductance. Eliminate the minus signs by pulling them into the temperature differences.)

Heat capacity is a significant material property. It describes the amount of energy required to change the temperature of a substance. Heat capacity is a function of temperature, and increases with rising temperature [9, p. 4]. For a mass, heat capacity is expressed in J/K. For materials, typically either *molar heat capacity* J/(mol*K) or *specific heat capacity* J/(kg*K) is given. *Volumetric heat capacity* J/(cm^3*K) is also encountered. In particular, given the volume of some substance, to determine heat capacity from molar heat capacity, the density, in g/cm^3, and atomic weight, in g/mol, is required. In Table 2, some specific heat data is listed for materials encountered in this project [16, pp. 373–375], [17]. In Table 3, densities are listed as well [16, p. 55], [18, Ch. 4.5]. Regarding the Bi_2Te_3 materials, the materials may not exactly match what is used in the TECs for the experiments in this thesis. But, the material properties listed here allow an estimate of heat capacity for the TECs.

Material	**Specific Heat J/(g*K)**	**Condition**
Copper	0.38	25 C. See the footnote.[2]
Silicon	0.702	25 C
Aluminum	0.899	25 C
p-$Bi_{0.52}Sb_{1.48}Te_3$	0.1870	25 C
n-$Bi_2Te_{2.88}Se_{0.12}$	0.1562	25 C

Table 2 Material Specific Heats

[2] The CRC Handbook gives specific heat in cal/(g*K). There are 4.18 J/cal. Shtern et. al. give molar heat capacities J/(mol*K) for the Bi_2Te_3 materials. Using the molecular formulas and atomic weights of the individual atoms, the molecular weight of the p-type material is computed as 671.67 g/mol, and 794.9 g/mol for the n-type.

Material	Density g/cm^3	Condition
Copper	8.93	Solid
Silicon	2.33	Solid
Aluminum	2.7	Solid
Bi_2Te_3	7.86	Solid. See the footnote.[3]

Table 3 Material Densities

Finally, the basic physical components in a typical microprocessor temperature control system are an integrated heat spreader, heat sink, and fan. The integrated heat spreader sits on top of the silicon die of the microprocessor and distributes heat to a larger area. The heat sink contacts the heat spreader, and provides a large surface area for air from the fan to blow over. Most heat sinks are finned, or play similar tricks, to increase surface area. At a microscopic level, surfaces are bumpy and are not completely smooth. Pressing two surfaces, like those of the heat spreader and heat sink, together will not necessarily create good thermal contact because of the bumps. Therefore, some thermal interface material is usually spread between contacting surfaces to fill in the microscopic bumps and provide a good interface [9].

Thermoelectric Cooler Physics

Thermocouples demonstrate thermoelectric phenomena at work. Although the classic is the mercury thermometer, thermocouples are another temperature measuring device. Thermocouples can be used to record the temperature of electronic devices, like discrete power transistors. A thermocouple consists of two bars of metal. Each bar is a different type of metal. For example, a pair of metals that is often encountered is copper-constantan. Constantan is a copper-nickel alloy [9, p. 186]. Suppose the metal bars join at points A and B, and that one of the bars is cut to insert a voltage meter. If A is held at one temperature and B is held at another, a voltage difference will arise at the cut. This voltage is

[3] Wagner [18] references J. Drabble, Progress in Semiconductors, vol. 7. John Wiley & Sons, Inc., New York, 1963.

proportional to the temperature difference: $V = \alpha \Delta T$ [9, p. 169]. This overall effect is known as the *Seebeck effect*. The units of alpha are V/K. Some sign convention regarding the polarity of the voltage will apply. Note that the *Seebeck coefficient* alpha may depend on temperature.

Two other thermoelectric effects are the *Peltier effect* and the *Thomson effect*. The Peltier effect applies to the situation of two different materials meeting at a junction. When an electric current passes through the materials, the current pumps heat from one side to the other: $q = \pi * \mathrm{I}$. The units of the *Peltier coefficient* π are W/A. The Peltier coefficient, like the Seebeck coefficient, could depend on temperature [19, p. 2]. Thermoelectric coolers employ the Peltier effect to their advantage, as shown in Figure 4. Typical thermoelectric coolers use semiconductor materials, like Bismuth-Tellurides. The charge carriers in p-type semiconductor materials are holes, while the charge carriers in n-type materials are electrons. In the figure below, when current flows in the positive direction, the holes in the p-type material move up. Also, the electrons in the n-type material move up. The movement of the charge carriers transports heat from the cold side of the couple, which is the bottom in this case, to the hot side [3, Sec. Abstract]. TECs use semiconductor materials rather than metals because the Seebeck coefficients for semiconductor materials are much better [9, p. 177]. The TEC's figure-of-merit, discussed later in this section, depends on the square of the Seebeck coefficient. The Seebeck coefficient is relevant in this discussion of the Peltier effect because the Seebeck and Peltier effects are related.

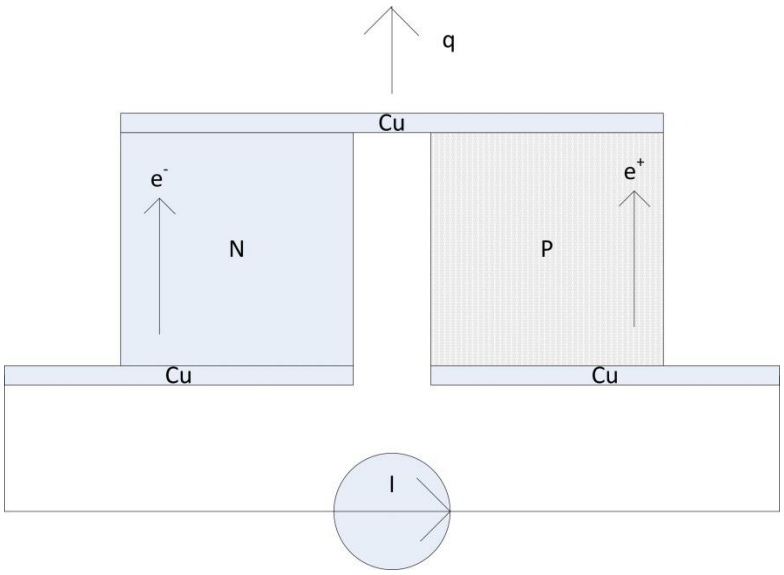

Figure 4 Thermoelectric Couple. The copper straps connect the P and N materials and provide a path for current to flow. The positive and negative charge carriers e^+ and e^- are shown. The figure is adapted from [3].

The *Thomson effect* describes how a single material with a temperature gradient absorbs (or rejects) heat when a current flows through it: $q = \tau I \frac{dT}{dx}$. τ is the *Thomson coefficient*. Certain relationships between the 3 effects also exists: $\pi = \alpha T$ and $\tau = T \frac{d\alpha}{dT}$.

A TEC consists of many thermoelectric couples, which are electrically in series and thermally in parallel. A single electric current passes through all of them, and they all pump heat from the cold side to the hot side. Figure 5 shows a cross section of a TEC [9, p. 178]. The TEC is built of thermal insulation, electrical insulation, copper, and semiconductor materials. In the case of the TECs used in this thesis, the copper straps are soldered to the thermoelements, which are made of Bi_2Te_3 materials, and the mounting surface is a ceramic.

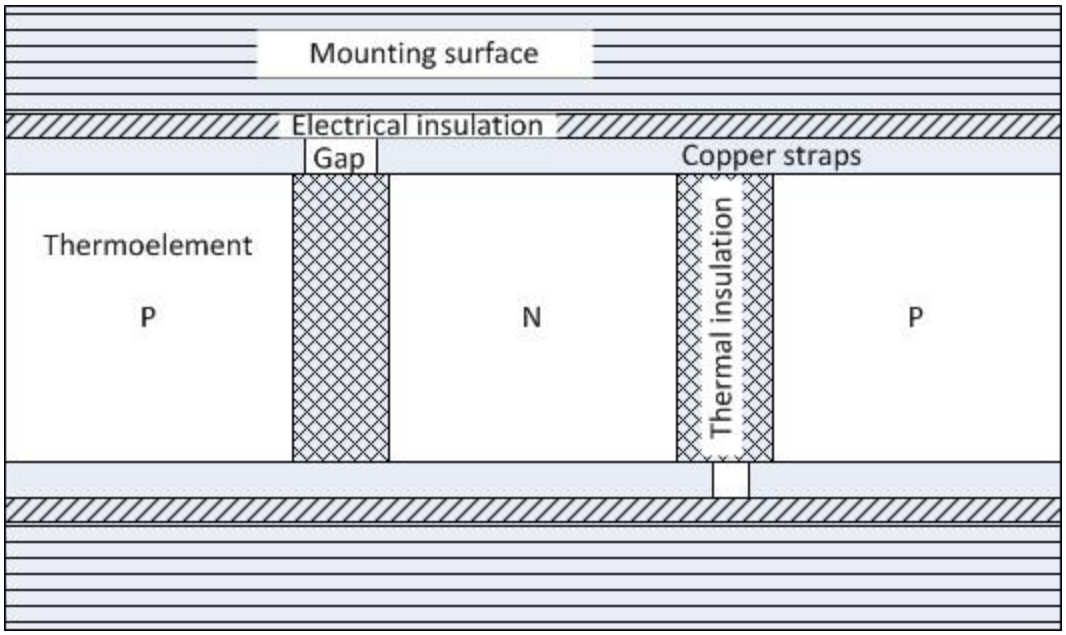

Figure 5 TEC Cross-Section for a Bulk TEC. The figure is adapted from [9]. At the end of this section, bulk TECs are contrasted with thin-film TECs.

The basic TEC variables and parameters are:

$S \coloneqq TEC\ Seebeck\ coefficient, in\ V/K$
$T_c \coloneqq TEC\ cold\ side\ temperature, in\ K$
$T_h \coloneqq TEC\ hot\ side\ temperature, in\ K$
$I \coloneqq TEC\ current, in\ A$
$\Delta T \coloneqq T_h - T_c$
$K \coloneqq TEC\ thermal\ conductance, in\ W/K$
$R_e \coloneqq TEC\ electrical\ resistance, in\ Ohms$

The heat drawn by the cooler at the cold side, q_C (in W), is derived by doing an energy balance and applying the physics of the thermoelectric effects above. Equation 1 is a starting point for this thesis. Note that this equation is sometimes given in terms of the individual thermoelectric couple parameters and the number of couples [20].

$$q_c = ST_cI - K\Delta T - \frac{1}{2}R_eI^2$$

Equation 1 TEC Cold-Side Heat Drawn

The power consumed by the TEC, in Watts, is

$$P_{TEC} = S\Delta TI + I^2R_e$$

Equation 2 TEC Power Consumption

The voltage across the TEC is a combination of the Seebeck effect and electrical resistance ($S\Delta T + IR_e$); power is current times voltage. For any refrigerator it is useful to define its efficiency (*coefficient of performance*) η. In this case, $\eta = \frac{q_c}{P_{TEC}} = \frac{ST_cI - K\Delta T - \frac{1}{2}R_eI^2}{S\Delta TI + I^2R_e}$. For a refrigeration cycle, the best possible efficiency is the Carnot efficiency, which is $\frac{T_c}{T_h - T_c}$ [3, p. 703]. For the case of the TEC, this occurs if the TEC thermal conductance K and TEC electrical resistance R_e are both 0. The heat exiting the hot side of the TEC is $q_h = q_c + P_{TEC} = ST_hI - K\Delta T + \frac{1}{2}R_eI^2$.

Some useful values that describe the TEC at a higher level are: q_{max}, the maximum heat the TEC can remove from the cold side in W, which occurs when ΔT is 0; ΔT_{max}, the maximum hot-side-cold-side temperature difference attainable in K, which occurs when no heat is removed at the cold side; and, I_{max} , the current in A required to obtain ΔT_{max}. V_{max} is the voltage at the TEC terminals when ΔT_{max} is obtained. Manufacturers provide these values for TECs at a fixed hot side temperature. The following equations can be derived by manipulating (differentiation, algebra) Equation 1 [21, p. 4].

$\Delta T_{max} = \frac{ZT_c^2}{2}$, where $Z = \frac{S^2}{R_eK}$ and is the *figure-of-merit* of the TEC. As Z gets better, TEC performance improves.

$$I_{max} = \frac{ST_c}{R_e}$$

$$q_{max} = \frac{S^2T_c^2}{2R_e}$$

In terms of applying TECs to microprocessors, the trend has been towards thin-films. Bulk TECs are constructed by arranging an array of p-n thermoelements. The thickness of the thermoelement is along the direction in which heat is pumped; see Figure 4 and Figure 5 above. The thickness of a bulk thermoelement might be 0.2 mm, and the entire cooler might be 1 mm thick [21, p. 3]. On the other hand, TECs manufactured using thin-film technologies can be much thinner. For example, in 2009, [22] reported the fabrication of thin-film Bi_2Te_3 superlattice-based thermoelectrics into a microprocessor package for cooling purposes. Superlattice-based TECs are a specific type of thin-film TEC. In the case of the paper mentioned, by thin-film, the authors mean that the TEC itself is 100 um thick, while the thin film materials are 5-8 um thick. As a standard of comparison, the minimum diameter of a strand of human hair is roughly 20 um.

Paraphrasing Snyder et. al. [21], thin-film based TECs have three main advantages over bulk TECs. Hotspot cooling requires TECs with high maximum heat flux, which is heat-pumped per unit area. This is because, by definition, hotspots are small, localized areas of high heat flux. In the ideal case, there would be one TEC per potential hotspot. The single TEC would engage when its hotspot requires cooling, which implies that the TEC by itself needs to manage the high heat flux. The TEC maximum heat flux is inversely proportional to thermoelement thickness, so thin-films are advantageous. Second, it turns out that the time response of a thermoelement is proportional to thickness squared. Having a thinner TEC will mean a faster TEC, in the sense that it takes less time to reach steady state. Third, thin-films enable more efficient TECs. For a given ΔT, the TEC optimal COP occurs at some fraction of I_{max}. I_{max} is the current at which the TEC pumps Q_{max} W of heat. At this fraction of I_{max}, only a fraction of Q_{max} will be pumped by the TEC. Therefore, Q_{max} should be extra large, such that even a fraction of Q_{max} pumps the required amount of heat from the processor. Q_{max} can be made large by having a TEC with high heat-flux, since Q_{max} will be the area (see Figure 2 on page 7) all the thermoelements take up times the maximum heat-flux.

2. Related Work

The *Modeling* chapter of this report in part considers a steady state model for the TECs, and develops an expression for the CPU junction temperature as a function of TEC current and CPU power consumption. Zhang et. al. (2010) develop an identical model, and analyze it using the TEC parameters for two commercially available TECs [23]. This thesis differs by additionally analytically considering the optimal TEC current to operate at. Using their model, the authors plot: junction temperature as a function of TEC current for a fixed CPU power consumption; TEC coefficient-of-performance as a function of TEC current for a fixed CPU power consumption; and, system thermal resistance when the TECs are engaged. This thesis uses experimental data to make similar plots. Zhang et. al. also consider whether liquid or air cooling is better when paired with TECs. As expected, liquid cooling is more effective, since it can draw away more heat.

Since this thesis constructs a TEC system, and also briefly touches on hotspot cooling, it is worth mentioning Chowdhury et. al. again [22]. Chowdhury et. al. were mentioned in the *Introduction* regarding thin-film superlattice-based TECs. The authors chose to have the TECs directly contact the silicon die to provide hotspot cooling. However, rather than growing the thin film TECs on the silicon die, the choice was made to grow the thin films on the underside of the integrated heat spreader. This prevents the thin-film process from having to be integrated into the silicon processing. The challenges faced were mostly a result of thermal and electrical contact resistances, which degraded the performance of the TEC. In terms of cooling capability, the TEC had a heat flux of almost 1300 W/cm^2. The package was tested using a heater. If the heater maintained a temperature of 116 C when the TEC was off, the TEC was able to produce a temperature drop of 7.3 C at 3 A of current.

An additional nice characteristic of thin-film TECs is that processor temperatures drop just by integrating the TEC into the package, even when the TEC is off. The reason is that, compared to the standard

package without a TEC, the thin-film TEC replaces some thermal interface material which has a higher thermal resistance than the TEC. This is termed the passive cooling effect of the TEC, in contrast to the active cooling that it performs when a current is run through it.

Unlike with thin-film TECs, embedding a bulk TEC in between the die and spreader will increase junction temperatures when the TEC is off, because the TEC has a big thermal resistance, which will probably be at least 5 K/W. [24] solves this problem, for the case of embedding the TEC in between the spreader and fan, by attaching a second heatsink to the spreader to cool it back down to an appropriate level. However, the paper considers the issue in the somewhat different context of a thermoelectric generator that converts CPU waste-heat.

Similar to Chowdhury et. al., Alley et. al. fabricate TECs on the underside of the spreader for a dual-core microprocessor [25]. In their experiments, the authors load the CPU and increase junction temperatures by running software. If the dual-core CPU is running at 78 C initially, their setup can achieve core junction temperature drops of 5 to 6 C while having the TECs consume around 2.5 W of power. Unlike this thesis, the paper doesn't consider processor frequency as a variable. As a side note, one nice feature of the paper's experiments is that a thermal chamber is used to maintain fixed ambient temperatures.

Chapter 4 builds a controller for the TEC system. [26], which develops a dynamic model for the TEC that captures its behavior over time, is thus relevant. The model's underlying differential equation accounts for heat capacity, thermal conduction, Joule heating, and the Thomson effect, and is linearized to make it easy to work with. Subsequently, the paper designs a proportional-integral controller for the TEC, and implements the controller with an analog circuit, using components like op amps. The controller tolerates variations in TEC parameters, so can withstand manufacturing variation in the TEC. The paper also analyzes how the controller responds to changes in the temperature set point, and to random

variation in the control signal. To verify the controller, a test setup with a fan-cooled TEC on top of a heater is constructed. If the controller is set to maintain a cold-side TEC temperature of 0 C, and the heater steps between 0, 5, and 10 W over time, the cold-side temperature deviates from the set point by +/- 0.1 C. The disturbance value of 0.1 C is certainly specific to the authors' experimental setup, but it does provide insight into how well a TEC controller can perform.

Chaparro et. al. (2009) consider the control-algorithms aspect of thin-film thermoelectrics using simulation [27]. The microprocessor simulator executes instructions, and the power and thermal simulators predict CPU power consumption and CPU junction temperatures. The paper simulates a 16-core architecture, which is most relevant to servers and datacenters. The simulation is set up such that each individual core can have multiple TECs on top of it.

Several controllers are discussed in the context of maintaining a temperature constraint. Although this thesis restricts itself to considering TECs and DVFS, Chaparro et. al. also consider thread migration as a strategy. An example of thread migration is moving a high-power thread from a hot core to a colder core, or perhaps a core that has TECs on top of it. All the cores don't have to have TECs on them, since going all-out might not be cost effective. The inputs for the controllers developed are CPU power, TEC power, and the core temperatures. For evaluating the performance of the controllers if there are power constraints, the authors use an ED^2 metric (energy-delay squared). The metric considers both energy consumption and runtime, and should be as low as possible.

The interesting controllers are highlighted here. For a simple algorithm, an on-off controller which always engages the TECs at a fixed current is used. The more complicated algorithm suggested is a PID (proportional-integral-derivative) controller. These two algorithms are then combined with DVFS and thread migration. In a performance maximizing scenario, when the TECs are turned on the DVFS controller increases frequency. In the simulations, although performance improves by up to 13%

compared to using DVFS by itself without TECs, the ED^2 metric goes up by at least 25%. While the TECs definitively improve performance, even for the case of thin-films their power consumption makes them unattractive from the perspective of ED^2. The authors also consider a scenario in which the controller tries to optimize for ED^2, but in this case performance barely improves.

Finally, the paper fine tunes the TEC control algorithm. It has the controller account for the fact that engaging a TEC on top of a neighboring core will have some cooling effect on the core under consideration. It also builds the TEC models into the controller, making the controller model-aware.

This thesis takes the perspective of using TECs to improve performance. Reda and Paterna utilize TECs to address the problem of dark silicon, which occurs when all of the cores on a multicore processor are not used [28]. Dark silicon might arise either because applications are not parallel, or because hotspots prevent all cores from being utilized. While the previous paper again examined improving performance, Biswas et. al. consider using TECs to alleviate the load on global cooling in datacenters [29]. The basic idea is that the TEC located on a hotspot can be engaged while the other TECs remain off. In contrast, any non-targeted cooling system cools all portions of the die equally. It cools some areas of the die unnecessarily and consumes extra power in the process. The paper suggests the ambient temperature of the datacenter can be allowed to rise if TECs are used, which will save on air-conditioning costs. [21] takes the viewpoint of reliability. TECs can decrease junction temperatures and improve processor lifetime.

3. Modeling

Introduction

This chapter first details the TEC experimental setup. At least a cursory understanding of the experimental setup is a prerequisite for reading the next chapter about the system characterization and controller implementation.

Later, this chapter considers two alternative system setups. A significant disadvantage of the basic TEC setup is that the thermal resistance of the TECs is high. For a given CPU power level, when the TECs are off, junction temperatures are significantly higher compared to a system with no TECs in it, which as mentioned before, is termed the *"standard system"*. To address this problem, it seems reasonable to put some thermally conductive material in parallel with the TECs to bring down temperatures when the TECs are off. However, this design is ineffective, and having the TECs by themselves is better. A second design choice is whether the TECs should go directly on the die or on top of the integrated heat spreader. While placing the TECs on the die provides hotspot cooling capabilities, it results in the system having a higher thermal resistance, at least for the specific setups considered here. Finally, this chapter develops a steady-state thermal model for the experimental setup with the TECs. Given system parameters, the model describes how junction temperature varies as a function of CPU power and TEC current. Experimental data shows that the model is accurate. With the steady-state model in hand, the thermal resistance of the TEC system can be analyzed.

Experimental Setup

A diagram of the basic system setup is below in Figure 6. Some alternative setups will be considered later.

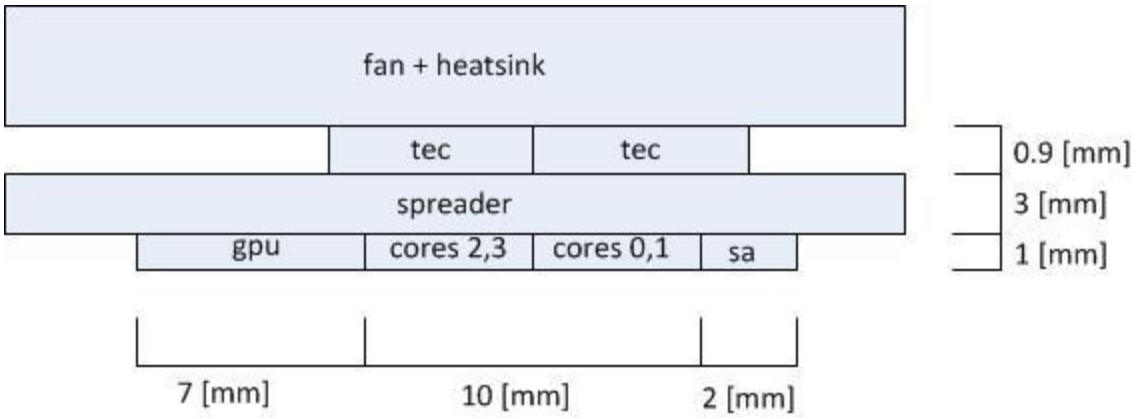

Figure 6 System Diagram

The central component is the CPU. This particular chip, the Intel i5-3450S (Ivy Bridge), has an integrated GPU and is a quad-core processor. The distribution of area between cores, GPU, and system agent (memory controller etc.) is based upon a die map, which is shown in Figure 7 below. Nominally, the maximum frequency is 2.8 GHz, although the chip does have a turbo boost feature which can bring the frequency up to 3.5 GHz. It has DVFS steps ranging from 1.6 to 2.8 GHz, in 0.1 GHz steps. Interestingly, each core can run independently at a different frequency, although the entire CPU can only run at a single voltage [30, Vol. 1, p 50]. Each core of the CPU has a digital thermal sensor which records the core's temperature. When purchased, the processor comes with the copper heat spreader installed. The heat spreader provides mechanical protection for the die and distributes the heat generated by the CPU. Adjacent layers in the above diagram have thermal interface material between them to provide good thermal contact.

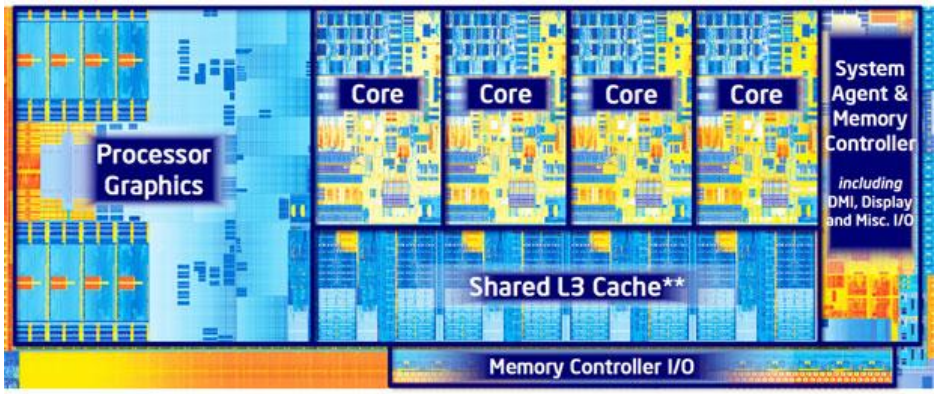

Figure 7 Intel Ivy Bridge Die Map. The image is from http://hothardware.com/Reviews/Intel-Core-i73770K-Ivy-Bridge-Processor-Review/ [31].

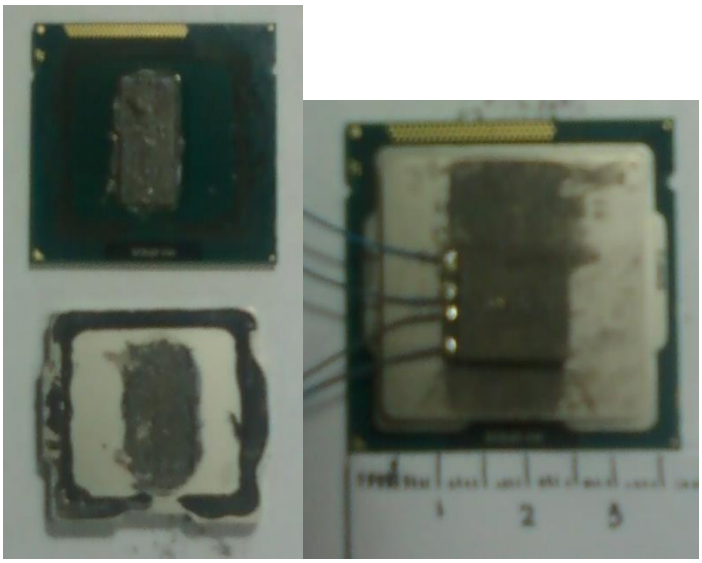

Figure 8 The Intel i5-3450S Silicon Die and the TECs. The photo on the left shows the processor with the integrated heat spreader removed. The silicon itself isn't visible because it is covered by thermal interface material. The photo on the right shows the TECs on top of the spreader and centered over the cores.

The TECs are wider than the 8 mm width of the die so stretch across both the cores and the L3 caches. Figure 8 shows the die, and the TECs on top of the spreader, just like in the system setup. Each of the two TECs is controlled by its own power supply. The reason for having two independently controlled TECs, each one over a pair of cores, is hotspot cooling. With two TECs, if only one pair of cores gets hot, then only one TEC needs to be engaged. Unfortunately, with this setup, hotspot cooling isn't observable.

The reason is that the heat spreader distributes the cooling effects of the TECs evenly across the die. I_{max} for each TEC is 5.3 A, ΔT_{max} is 67 K, and Q_{max} is 18.7 W.

All of the experiments consist of running SPEC2006 benchmarks. To be clear, the purpose of running the benchmarks is not to measure the performance of the CPU. Instead, the goal is to make the CPU generate power. A mix of floating point and integer benchmarks are used. The data collected for each benchmark run is listed below:

- *CPU core temperatures*, in C. The thermal sensor for each core reports a temperature, in a whole number of degrees. The sensors are sampled at a particular rate, say 1 s. Often, it is useful to subsequently average the collected data over time, on a per-core basis. Averaging can also be done over cores. For example, core 1 might have an average temperature over time of 30 C, core 2 40 C. The average over both cores is 35 C. Whenever junction temperature is mentioned, the readings from the digital thermal sensors are implied.
- *Fan speed*, in rpm. This is recorded to verify that the fan runs at roughly constant speed. The intent is to have the fan be a constant, rather than variable, thermal resistance.
- *Multimeter voltage*, in mV. The multimeter records the voltage across a shunt resistor, which is embedded on the path from the power supply to the CPU power motherboard connection. This gives a measure of current running through the supply to the CPU. This particular supply line is at 12 V. Thus, an estimate of CPU power consumption is possible.
- *TEC current and voltage*. As current is supplied to a TEC, it develops some voltage. This voltage is recorded.
- *DVFS setting*. The core frequencies are recorded. Typically, all the cores are set at the same frequency.

Errors

Experimental error can come from the temperature sensors. The Intel manuals don't list how accurate the sensors are. In terms of other instrumentation precision, the multimeter and power supply give relatively precise readings. However, as discussed later on, the power measured is not the power consumed by the CPU alone, but also includes the power consumed by the motherboard regulators. Regulators are not 100% efficient, so the measured power will be higher than the actual CPU power. CPU to CPU manufacturing variation is also not accounted for, since all the experiments in this thesis are performed with one physical CPU.

Equipment

- Hardware
    - Agilent A34410A digital multimeter. This is used to measure the voltage across the shunt resistor mentioned below.
    - Agilent A34330A 1 mV/A shunt. This allows measurement of the current drawn by the CPU from the power supply.
    - 2 Tektronix 4205 20V, 5A power supplies. These control the current to the TECs.
    - 2 RMT 1MDL06-052-03AN TECs. The TECs are 6 mm x 12 mm, and 0.9 mm thick. The TECs fit over the entire width of the die since 12 mm is bigger than 8 mm, which is the width of the die.
    - CoolerMaster V8 180 W fan.
    - ASRock Z77 Extreme4 Motherboard
    - Ultra X4 CPU Power Supply
    - 2 GB RAM
    - 250 GB hard disk

- Intel i5-3450s microprocessor. The die itself is roughly 8 mm x 19 mm, and 1 mm thick. This is a quad-core processor.
- Arctic Silver 5 thermal interface material.

- Software
  - Linux 2.6.39, Ubuntu 9.04.
  - SPEC2006 benchmarks.
  - Data collection software. This runs while the benchmarks are running. It collects temperature sensor, multimeter, fan speed, and power supply data. The data collection software has low overhead in terms of processor utilization.

TEC Parameters

The basic TEC parameters are the Seebeck coefficient S, the thermal conductance K, and the electrical resistance R_e. On the other hand, typical datasheets for TECs provide a different set of experimentally measured parameters. For example, the 1MDL06-052-03 TEC used in the experiments presented here has the following parameters.

Q_{max}	18.7 W
ΔT_{max}	67 K
I_{max}	5.3 A
V_{max}	6.3 V

These parameters are given at vacuum at 300 K for the TEC hot side temperature.

Expressions for Q_{max}, I_{max}, ΔT_{max}, and V_{max} can be found by taking the equation for the heat removed at the cold side of the TEC ($q_c = ST_cI - K\Delta T - \frac{1}{2}R_eI^2$) and maximizing either q_c or ΔT. Note that there are 4 expressions, but only three unknowns, S, K, and R_e. Therefore, choose three out of the four equations to solve. Here, the following equations are chosen.

$$I_{max} = \frac{S * T_c}{R_e}$$

$$\Delta T_{max} = \frac{S^2 T_c^2}{2KR_e}$$

$$V_{max} = \frac{S^3 * T_c^2}{2KR_e} + S * T_c$$

We also know that $\Delta T_{max} = T_h - T_c$. Substituting the manufacturer datasheet parameters and measurement conditions, we get the following values for a single TEC:

$K = 0.19 \frac{\mathrm{W}}{\mathrm{K}}$; the corresponding thermal resistance is 5.26 K/W

$$R_e = 0.92\ \Omega$$

$$S = 0.021\ \frac{V}{K}$$

Comparing to values computed in other papers, these values are in the typical range [23, p. 564]. These values will be useful in numerically testing the thermal model developed later. Finally, it's worthwhile putting these parameters in the context of what is desired out of a TEC. In a cooling application, a high TEC efficiency is required. Q_{max} should be large, but for a low level of input power.

Steady State TEC Equations

A simple model of the thermal system that this report deals with is in Figure 9 (a). A source of heat (in this case a CPU that generates P W) sits below a TEC, which in turn lies below some thermal resistance R K/W. T_a is the ambient temperature, which is typically 293 K (20 C). T_h and T_c are, respectively, the hot side and cold side temperatures, in K, of the TEC. R models the thermal resistance of the fan. This model is approximate because it neglects the thermal resistances of the copper heat spreader, silicon die, and

thermal interface material. These thermal resistances are relatively small compared to the fan and TEC thermal resistances.

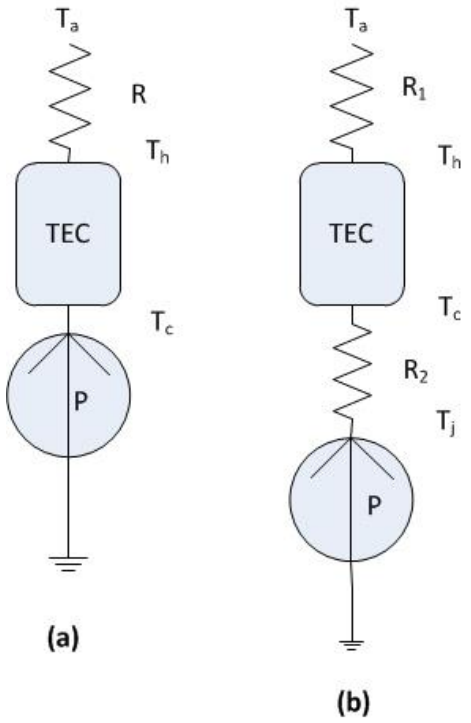

Figure 9 Thermal Models

A more accurate model is Figure 9 (b), which includes a thermal resistance R_1 above the TEC, and a thermal resistance R_2 below the TEC. If the TECs are on top of the copper heat spreader, , as in Figure 6 on page 24, R_1 represents the fan thermal resistance, and R_2 accounts for the thermal resistance of the silicon die and the heat spreader. Tj is the junction temperature of the silicon, in K. If the TECs are directly on top of the die, as seen with some of the setups discussed in the *Related Work* chapter, R_1 models the thermal resistance of the fan and heat spreader, and R_2 accounts for the thermal resistance of the silicon die. Reference Figure 19 on page 48 for a diagram of a setup with the TECs on die.

Overall, neither of these models accounts for thermal contact resistances. Therefore, in the experimental setup, a thermal interface material is always spread between adjacent surfaces to provide good thermal contact.

First consider the case of diagram (a) in Figure 9, with a single TEC sitting above the heat source. Examining the case of 2 TECs, each with its own power supply, will come later. The latter case is relevant because the experimental setup uses 2 TECs. Regardless, the goal is to derive an expression for T_c as a function of R, TEC current, TEC parameters, P, and T_a. Some basic definitions are listed below:

$S \coloneqq TEC\ Seebeck\ coefficient, in \frac{V}{K}$
$T_c \coloneqq TEC\ cold\ side\ temperature, in\ K$
$T_h \coloneqq TEC\ hot\ side\ temperature, in\ K$
$T_a \coloneqq ambient\ temperature, in\ K$
$I \coloneqq TEC\ current, in\ A$
$\Delta T \coloneqq T_h - T_c$
$K \coloneqq TEC\ thermal\ conductance, in \frac{W}{K}$
$R_e \coloneqq TEC\ electrical\ resistance, in\ Ohms$
$P = the\ power\ produced\ by\ the\ microprocessor, in\ W$
$R = system\ thermal\ resistance, \frac{K}{W}$

In both figures (a) and (b), the following two equations hold. Note that the temperatures really do have to be in Kelvin. The objective is to solve for the cold side temperature T_c. In figure (b), if T_j is desired, add $P*R_2$ to T_c. If a numerical model of the system were implemented, the following two equations would be sufficient. Performing Kirchoff's current law at each node would generate a system of equations to solve. However, the goal here is to investigate an analytical model. For figure (b), use R_1 instead of R. In the following discussion, some of the detailed algebra is purposely skipped over for the sake of brevity. The intermediate steps can be worked by hand.

1. $T_h = T_a + R\left[ST_hI - K(T_h - T_c) + \frac{1}{2}I^2R_e\right]$
2. $P = ST_cI - K(T_h - T_c) - \frac{1}{2}I^2R_e$

Rearranging (1) to isolate T_h, we get

$$T_h * (1 - RSI + KR) = T_a + KRT_c + \frac{1}{2}RR_eI^2$$

Substituting this expression for T_h into (2) and rearranging to solve for T_c, we get

$$P + \frac{K\left(T_a + \frac{1}{2}RR_eI^2\right)}{1 - RSI + KR} + \frac{1}{2}R_eI^2 = T_c(SI + K - \frac{RK^2}{1 - RSI + KR})$$

Doing some further rearranging by putting the right hand side under a common denominator,

$$T_c = P * \frac{1 - RSI + KR}{SI - RS^2I^2 + K} + \frac{KT_a + KRR_eI^2 + \frac{1}{2}R_eI^2 - \frac{1}{2}RR_eSI^3}{SI - RS^2I^2 + K}$$

Equation 3 TEC Cold Side Temperature

As a check on this expression, evaluate it when I is 0. The thermal resistance should simply be (1/K+R). Additionally, the expression was checked against the output of a computer algebra system, Maple. The results match. As a function of I, T_c is a rational function. It has a third order polynomial over a second order polynomial. Unfortunately, without knowing the specific values of parameters, it is difficult to evaluate the behavior of the equation. However, some interesting aspects of the equation to explore are $\frac{dT_c}{dP}$, and $\frac{dT_c}{dI} = 0$. The former is important because it shows how T_c grows as function of CPU power for a fixed current. The latter is useful because, empirically, for a fixed P, the cold side temperature as a function current is U-shaped and has an optimal current at which T_c is minimal.

Trivially, $\frac{dT_c}{dP} = \frac{1-RSI+KR}{SI-RS^2I^2+K}$. Determining an expression for the optimal current I is more algebra. The basic idea is to differentiate using the quotient rule. Then, there is some expansion of products and rearranging to be done. Ultimately, the following characteristic polynomial is obtained.

$$\frac{1}{2}R^2R_eS^3I^4 - RR_eS^2I^3 + \left(-PR^2S^3 - \frac{1}{2}KRR_eS + \frac{1}{2}R_eS\right)I^2$$
$$+ (2PRS^2 + 2KPR^2S^2 + 2KRS^2T_a + 2K^2RR_e + KR_e)I + (-PS - 2KPRS - KST_a)$$
$$= 0$$

Equation 4 TEC Optimal Current

An analytical expression for the roots of a fourth order polynomial does exist, but it is unwieldy. The best approach is to keep the above equation in mind, and given specific values for the parameters, solve for the roots numerically. As a check on the expression, I chose particular realistic parameter values and compared the minimum computed via the root-finding method (two roots were complex conjugates, one was negative, and one was positive, making the real solution easy to pick) against a plot of T_c using Equation 3. The minimums matched up.

Finally, consider the case of two TECs, mentioned previously. Assume the cold and hot side temperatures are the same for both of the TECs. This assumption is an approximation, considering that during experiments with the real CPU, each TEC has a slightly different voltage across it. Also assume that the current I through each TEC is identical and independently controlled, and that the TEC parameters are identical. In this case,

$$P = 2\left(ST_cI - K\Delta T - \frac{1}{2}I^2R_e\right)$$

Similarly, the heat pumped at the hot side of the TECs is

$$q_h = P + (2I^2R_e + 2IS\Delta T) = 2ST_hI - 2K\Delta T + 2\left(\frac{1}{2}I^2R_e\right)$$

Following an idea from Gray, make the following transformations [19, p. 45]. Let $I_t = 2I, K_t = 2K, R_{et} = \frac{1}{2}R_e$. Then $P = ST_cI_t - K_t\Delta T - \frac{1}{2}R_{et}I_t^2$. And, similarly, $T_h = T_a + R(ST_hI_t - K_t\Delta T + \frac{1}{2}R_{et}I_t^2)$. Having

made these transformation for I, K, and R_e, we end back at equations (1) and (2) mentioned above. Hence, all the analysis done thus far applies.

These models are shown to accurately predict the experimental behavior of the system in the Experimental Verification section, which comes after the discussion of system thermal resistance.

System Thermal Resistance

All of the components of the standard system – the fan, the thermal interface material, the copper heat spreader, and the silicon – can be lumped into a single thermal resistance. In the TEC system, when the TECs are off, having the TECs in the system will increase the thermal resistance above the level of the standard system. Thus, the thermal resistance of the standard system and the TEC system can be compared. To estimate these thermal resistances,

- Run a variety of benchmarks on the systems. For the standard system and the TEC system, gcc and povray were run over the 1.6 to 2.8 GHz range. Fan speed is fixed at a specific rpm.
- Plot temperature, averaged over both time and cores, against CPU power. The trend should be linear since thermal resistance is a linear element.

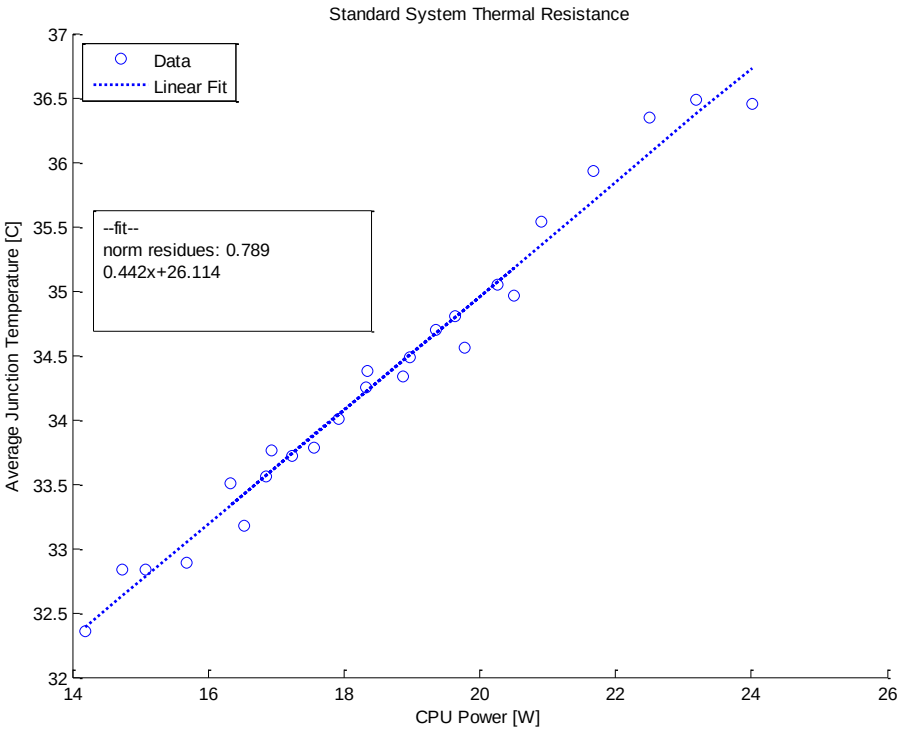

Figure 10 Standard System Thermal Resistance

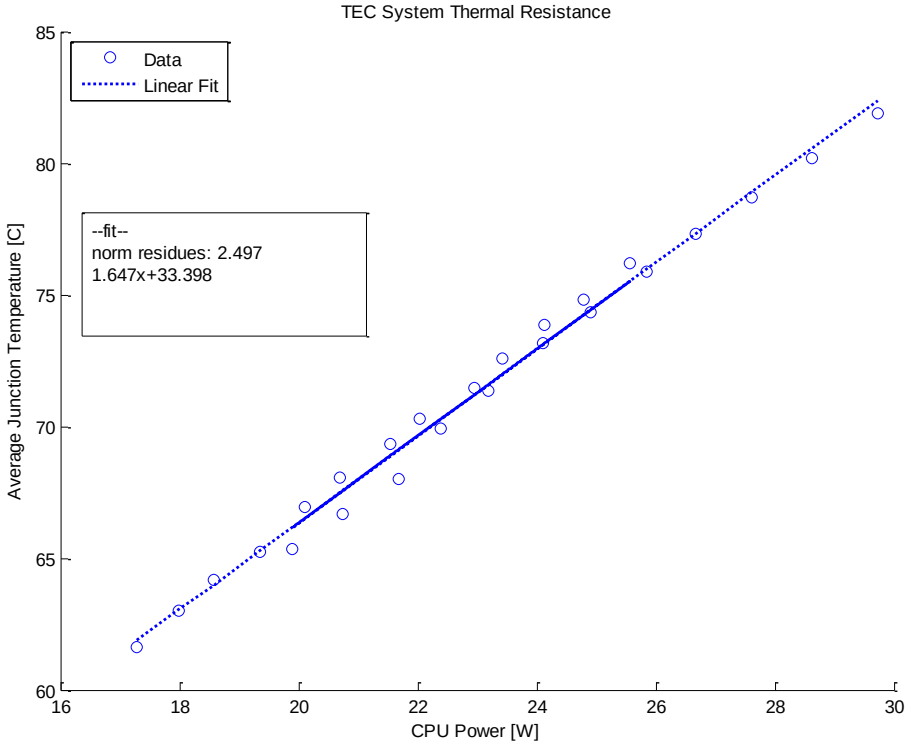

Figure 11 TEC System Thermal Resistance at 0 A TEC Current

As shown in Figure 10 and Figure 11, the thermal resistance of the standard system is around 0.44 K/W, while the thermal resistance of the TEC system (with no TEC current) is 1.6 K/W. As a result, the TEC system temperatures are 2 to 2.5 times higher in this 15 to 30 W range of CPU power. The digital thermal sensor readings that report CPU core temperatures only have 2 significant figures (e.g. 35 C). Even though the multimeter has greater precision, the thermal sensor precision makes the thermal resistances have 2 significant figures as well. On the other hand, one of the nice features of the Intel digital thermal sensors is that they are located at the hottest portions of the die [13, pp. 53–57].

The data shows nice linear trends. The y-intercept of each line should represent room temperature, which is roughly 24 C in the lab based upon the thermostat reading. When the CPU produces no power, the junction temperature should simply be 24 C. The data suggest room temperature is either 26 C or 33 C, which is off the mark. This demonstrates that the accuracy of the digital thermal sensors is not very good, particularly at lower temperatures.

Another discrepancy between the two plots is that the range of CPU power differs. For the TEC system experiments, CPU power is consistently 2 W higher. One potential reason is that leakage power grows with temperature; the TEC system runs at least 30 C hotter. However, this 2 W difference will have a minimal impact. Even if the CPU power consumption were 2 W higher for all data points in the standard system plot, temperatures would only shift up by 1 C. After all, the thermal resistance of the standard system is only 0.44 K/W.

In the standard system, the fan accounts for most of the thermal resistance. The copper heat spreader is about 3 cm x 3 cm, and is about 0.26 cm thick, based upon a micrometer measurement. The thermal resistance is $\frac{2.6*10^{-3}\ m}{9*10^{-4}[m^2]*391\frac{W}{m*K}} = 0.007\frac{K}{W}$. The silicon die is around 8 mm x 19 mm, based upon hand measurements, and maybe 1 mm thick [32, p. 4]. The thermal resistance is $\frac{1*10^{-3}\ m}{152*10^{-6}\ m^2*156\frac{W}{m*K}} = 0.04\frac{K}{W}$.

The thermal interface material can be neglected since it is spread in very thin layers, perhaps 25 um (1 mil) thick (although even in such thin layers it might exceed the thermal resistance of the copper heat spreader). So, in our measurements, subtracting off the die and spreader thermal resistances, the fan thermal resistance is around 0.4 K/W. In the TEC system, on the other hand, the thermoelectric coolers account for the bulk of the thermal resistance. Assuming they account for all of the extra thermal resistance in the TEC system, their thermal resistance is around 1.2 K/W (1.6-0.4). This doesn't match with the TEC datasheet values, which put the thermal resistance of 2 TECs in parallel at 2.6 K/W.

These thermal resistance values are only approximations. First, the method of measuring CPU power by measuring the power of the 8-pin 12V motherboard supply is inaccurate[4]. The 12V supply powers the regulators for the CPU, not the CPU directly, and these regulators (VRMs) might only be 80% efficient at best. Voltage regulator efficiency can vary as a function of load. The thermal resistances reported above are underestimates, since the actual CPU power is really less than reported. This partly explains why the experimentally determined TEC thermal resistance of 1.2 K/W is less than the datasheet value of 2.6 K/W.

The variation of regulator efficiency as a function of load can explain why the plots predict room temperature inaccurately. Consider a real thermal resistance of 1 K/W. At 0 W real cpu power, temperature should be 20 C, at 1 W 21 C, at 2 W 22 C. If the regulator is 50% efficient at all loads, then the readings would be 2 W measured power at 21 C, 4 W at 22 C. The y-intercept would still be at 20 C. If the regulator is 50% efficient at 1 W CPU power, and 25% efficient at 2 W, then the readings would be 2 W measured power at 21 C, and 8 W at 22 C. Based on the data, the y-intercept is no longer 20 C, but is instead 20.67 C.

[4] The motherboard manual shows the motherboard connectors, and mentions that the motherboard uses an 8+4 phase power supply design [33, pp. 6–8].

There are other problems with the methodology as well. Not all the CPU power goes up through the spreader. Some exits through the bottom of the motherboard. Failing to account for this effect will lower measured thermal resistances. Radiation should be a negligible effect. Also, treating the CPU power in bulk is inaccurate. For example, only 2 out of the 4 cores of the CPU generate power during the experiments, and the portions of the spreader directly above these cores will be somewhat hotter.

Figure 12 shows how the TEC system thermal resistance varies as a function of TEC current. This data was collected by running gcc and povray from 1.6 to 2.8 GHz, as before, but at various TEC currents. In the Steady State TEC Equations section of the Modeling chapter, an expression for the rate at which the cold-side TEC temperature varies was derived: $\frac{dT_c}{dP} = \frac{1-RSI+KR}{SI-RS^2I^2+K}$. The junction temperature $T_j = R_2 * P + T_c$ since T_j and T_c are simply separated by some thermal resistance; see Figure 9 in the Modeling chapter. Therefore the TEC system thermal resistance is $\frac{dT_j}{dP} = \frac{dT_c}{dP} + R_2$. This thermal resistance $\frac{dT_j}{dP}$ is a function of TEC current, so the changing thermal resistances (the slopes of the lines) in the plot make sense. For a fixed TEC current, the thermal resistance is constant: R, S, and K are system parameters.

The data shows that for the case of this TEC system, the thermal resistance decreases as a function of current. However, the main trend is that as current increases, the y-intercept decreases, though there are diminishing returns at higher currents. For instance, the lines for 4 A and 5 A TEC current essentially overlap, so the 5 A line was omitted from the plot. Since the 4 A line is roughly the best the system can do, it can be used to compute at what CPU power the TEC system will no longer be able to cool below the standard system. Setting the thermal resistance of the standard system equal to the TEC system thermal resistance at 4 A: $1.3P + 2.5 = 0.4P + 26$. Solving, P is roughly 26 W, after which point the TEC system will have higher temperatures than the standard system, even when a current is run through the TECs.

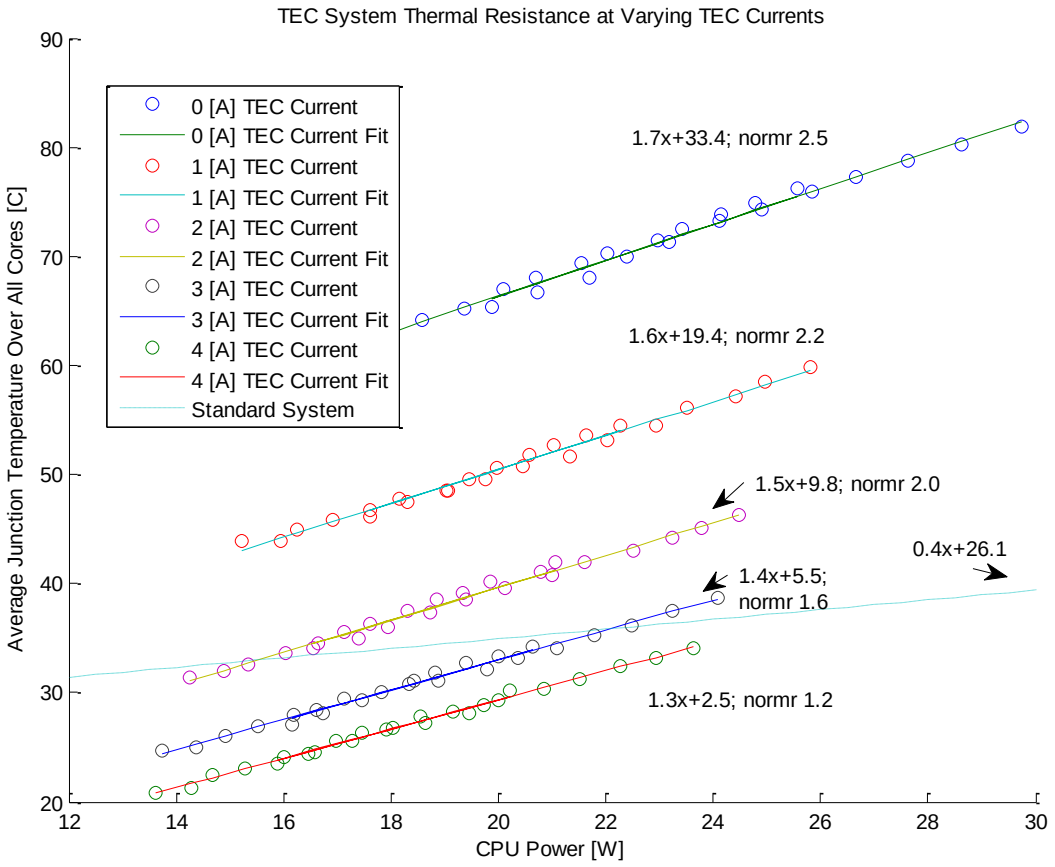

Figure 12 TEC System Thermal Resistance at Varying TEC Current. As current increases, the thermal resistance tends to decrease. However, the more significant trend is that the y-intercept of the lines decreases as TEC current increases. normr is the norm of the residues, and shows how good the linear fitting to the data is. The standard system line is shown to demonstrate the CPU power up till which the TEC system can cool below the standard system.

Experimental Verification

As discussed, a relatively simple analytical model for the TEC system is available. The purpose here is to show that the model fits the experimental data, if the model parameters are chosen appropriately. The model treats all thermal resistances as lumped quantities. The model also requires certain parameters: the thermal resistance R_1 that comes from the fan, the thermal resistance R_2 that comes from the silicon die and heat spreader, the thermal conductance K of the TECs, the electrical resistance R_e of the TECs, and the Seebeck coefficient S of the TECs.

For R_1 and K, the thermal resistances determined based upon the experimental data will be used. This is necessary because, as mentioned previously, the CPU power measured is not the real CPU power, but includes the inefficiency of the voltage regulators. R_1 and K from the experimental data are computed based upon the inaccurate CPU powers, and are therefore consistent with the inaccurate CPU power values. The model must use the CPU power values since they are the only ones available; using the experimental R_1 and K rather than calculated values will provide better results. Since R_2 is relatively small, a calculated value will be used; R_2's value will not have much of an impact anyways. The experimental value for R_1 is 0.4 K/W. The calculated value of K for the TEC, based upon the datasheet is 0.19 W/K, while the experimental data suggests it is 0.41 W/K for a single TEC. However, it turns out that the model fits the data better if a value of 0.3 W/K for a single TEC is chosen.

For R_e and S, the TEC datasheet values will be used. R_e and S are unrelated to thermal resistance, so it is a good idea to use the datasheet values rather than trying to determine them from the data. The data is inaccurate because of the CPU power issue, so will probably badly estimate the real values of S and R_e. However, it is possible to estimate the values from the data. For example, the equation for TEC power, $P_{TEC} = S\Delta TI + R_e I^2$ can be used. Given two separate (I, P_{TEC}) data points, and knowing T_c and T_h ($\Delta T = T_h - T_c$), S and R_e can be solved. Unfortunately, computing T_h and T_c requires using the CPU power values. For example, T_h is roughly $T_{ambient} + R_{fan} * P_{tec+cpu}$.

The parameters used are below. The TEC parameters K, S, and R_e are for one TEC. R and ambient temperature are not TEC parameters. Ambient temperature is set at 30 C rather than the measured room temperature of 24 C because the thermal resistance plots suggest the system sees room temperature as being closer to 30 C.

R	0.4 K/W
K	0.3 W/K
S	0.021 V/K
R_e	0.92 Ohms
T_{amb}	30 C

The analytical and experimental graphs for gcc at 2.8 GHz and 22 W of CPU power are below. Note that the plot is for average temperature, averaged over all cores. This provides better results since the plot employs a TEC static model, which assumes the system is at a steady state.

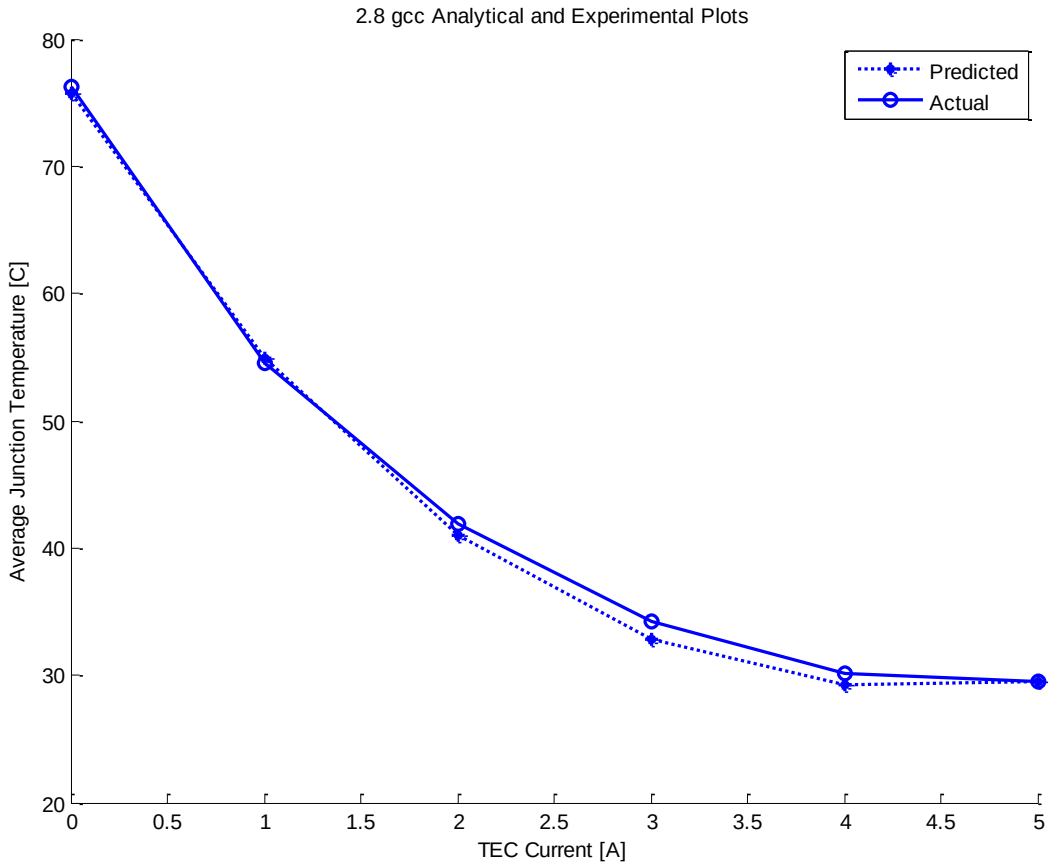

Figure 13 gcc Analytical Model

The analytical plot is relatively accurate. The mean absolute error is 0.6 C. (For other frequencies besides 2.8 GHz, the mean absolute error is similar. The worst mean error is 2.4 C at 1.6 GHz.) For most current

values, the analytical plot underestimates the actual temperature. The analytical solution predicts an optimal current of 4.4 A, and the real optimal current is between 4 and 5 A as well.

The analytical model also fits data from other benchmarks well. For example, the average temperature vs. current plot fits very nicely with the analytical model, when using the same parameters as used with gcc (see Figure 14). The mean error is 1.2 C.

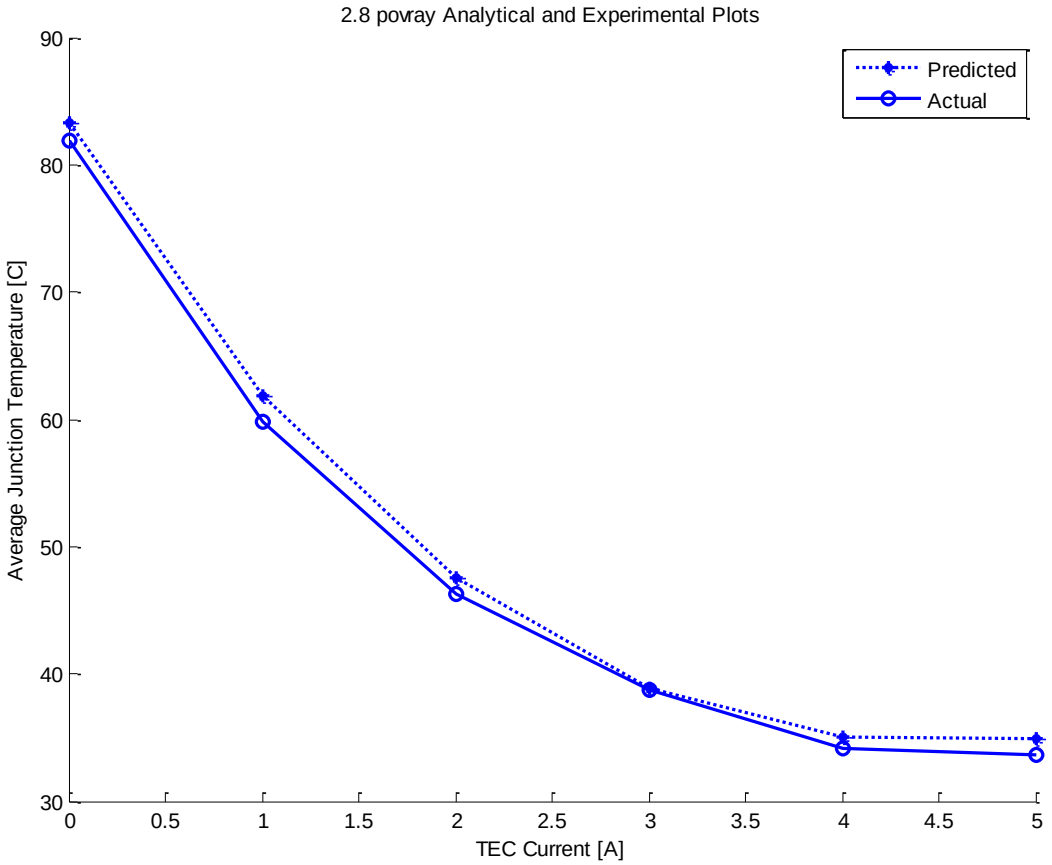

Figure 14 povray Analytical Model

Alternative Setups

A Copper Shunt for Heat to Escape

A significant drawback of the TEC system is that the TECs have a high thermal resistance when the TECs are turned off. The datasheet for the TECs used here gives the thermal resistance for 2 TECs side by side as around 2.6 K/W. At 10 W CPU power, that means the system would run 26 K hotter with the TECs in it, if the TECs were not engaged. A way to mitigate this is to add some copper (or other metal) in parallel with the TECs, to provide another path for heat to escape. This is shown in Figure 15 below.

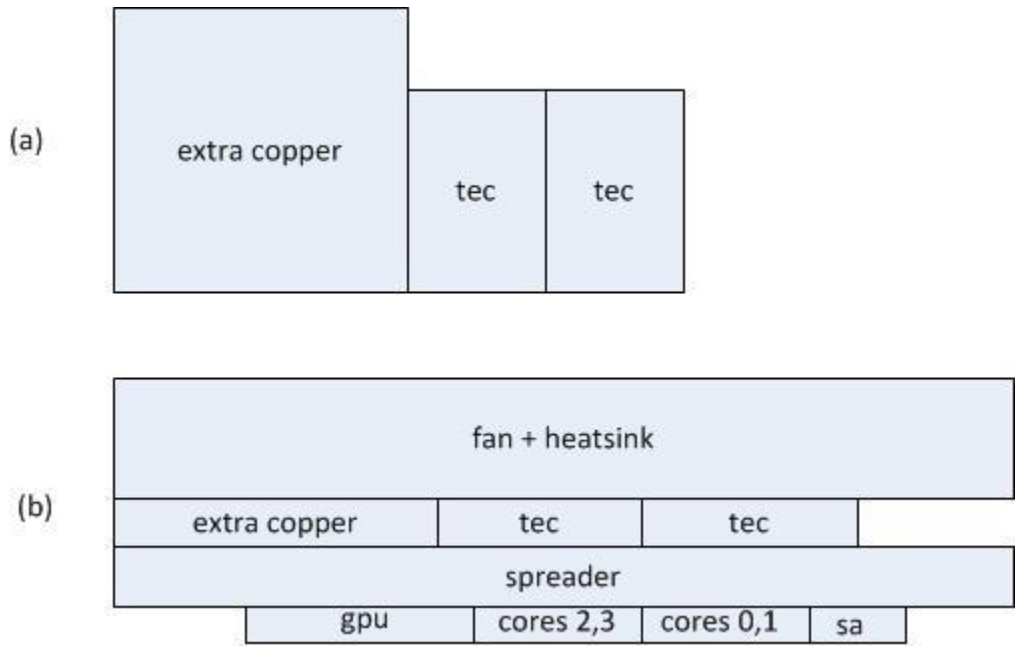

Figure 15 Extra Copper. These figures show what a system might look like with extra copper added, though the figures aren't to scale. In the simulations described in this section, the amount of extra copper present is varied. (a) A top down view of the TECs, in which the spreader is underneath the TECs and extra copper. (b) A side view showing all the components of the system. The TECs are effectively embedded in the spreader.

While this will have a benefit when the TECs are not engaged, TEC performance will go down when the TECs are on. Recall that the heat drawn from the cold side of the thermoelectric cooler is $ST_cI - K\Delta T -$

$\frac{1}{2}R_e I^2$ (see Steady State TEC Equations). When this copper is added, K goes up, making the TEC less effective. The question is whether there is some optimal amount of extra copper to add. The point is not to be concerned with TEC power consumption. Instead, the objective is to choose the amount of copper that allows the system to reach the lowest temperature possible, even if that lowest temperature is reached when the TEC is on.

To check what the optimal amount of copper is, start by adding a 3 cm x 3 cm block of copper which is about 2.54 mm (100 mil) thick next to the TECs. (Add some extra copper, which has the same area as the TECs, below the TECs so the height of the TECs and this extra copper adds up to 2.54 mm as well.) Incrementally decrease the size of the 3 x 3 block, and at each step determine the minimum temperature obtainable; the minimum temperature might be obtained when the TECs are on. TEC current is limited to 4.9 A maximum, since that is roughly the maximum TEC current the datasheet instructs not to exceed. Use a numerical model for the system in Figure 6 (see Steady State TEC Equations on page 29 for more details on this model). The simulation specifics are:

- For simplicity's sake, assume that the two TECs are connected electrically in series. That is, a single TEC current feeds the two TECs. This is different from the experimental setup, in which each TEC has its own power supply. For one TEC, K = 0.19 W/K, R_e = 0.92 Ohms, S = 0.021 V/K. Two TECs connected in series like this can be modeled as a single TEC with K' = 2K, $R_e' = 2R_e$, and S' = 2S. The TEC parameters are based off of the datasheet for TEC device used in all of the thesis' experiments.
- The fan thermal resistance is assumed to be 0.4 K/W. This value is based upon the standard system thermal resistance experiments. See Figure 10 on page 35.
- The CPU generates either 30 or 40 W of power, all of which flows upward through the TECs.

- Assume the material parameters listed in the section Thermal Modeling on page 9. If a range of values is possible, choose the maximum value.
- For physical dimensions, use the dimensions listed in the Equipment section on page 27. This applies particularly to the silicon die dimensions. Other dimensions are important as well. For instance, the TEC dimensions partly determine how much thermal interface material gets spread. The TIM is spread 25.4 um (1 mil) thick.
- Use a lumped model for all of the thermal resistances.

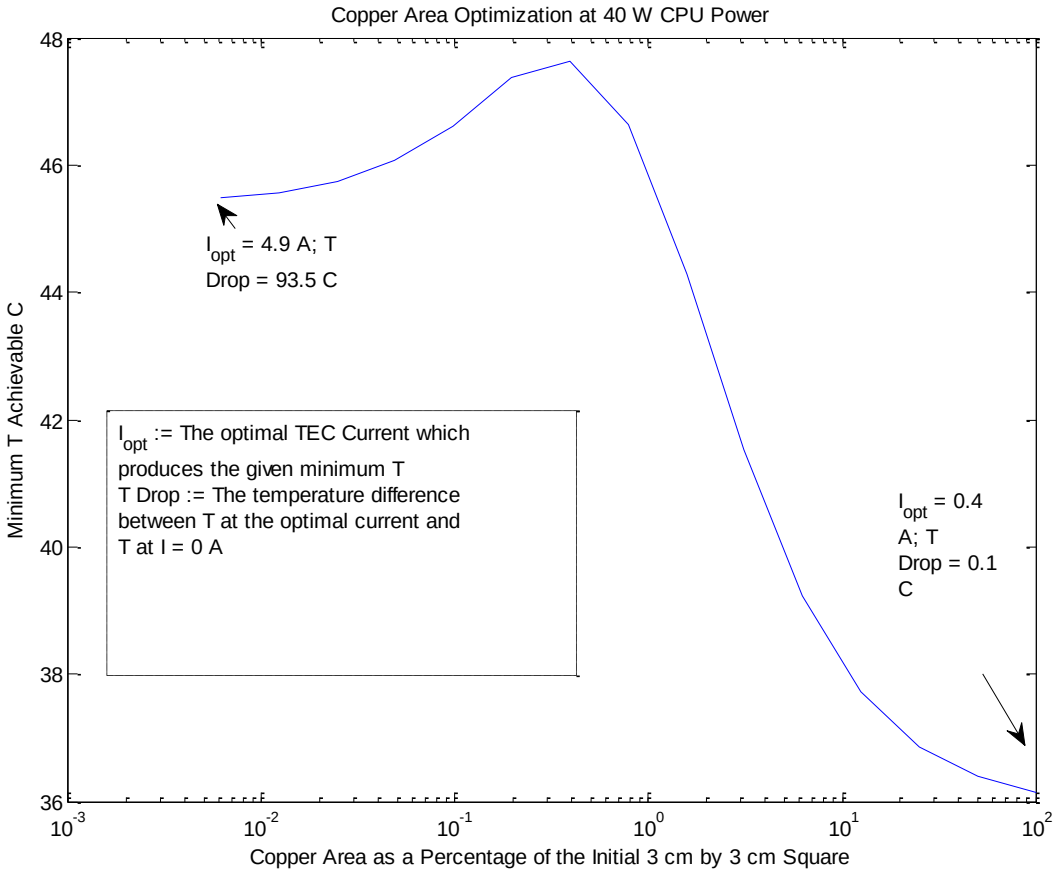

Figure 16 Copper Area Optimization at 40 W CPU Power

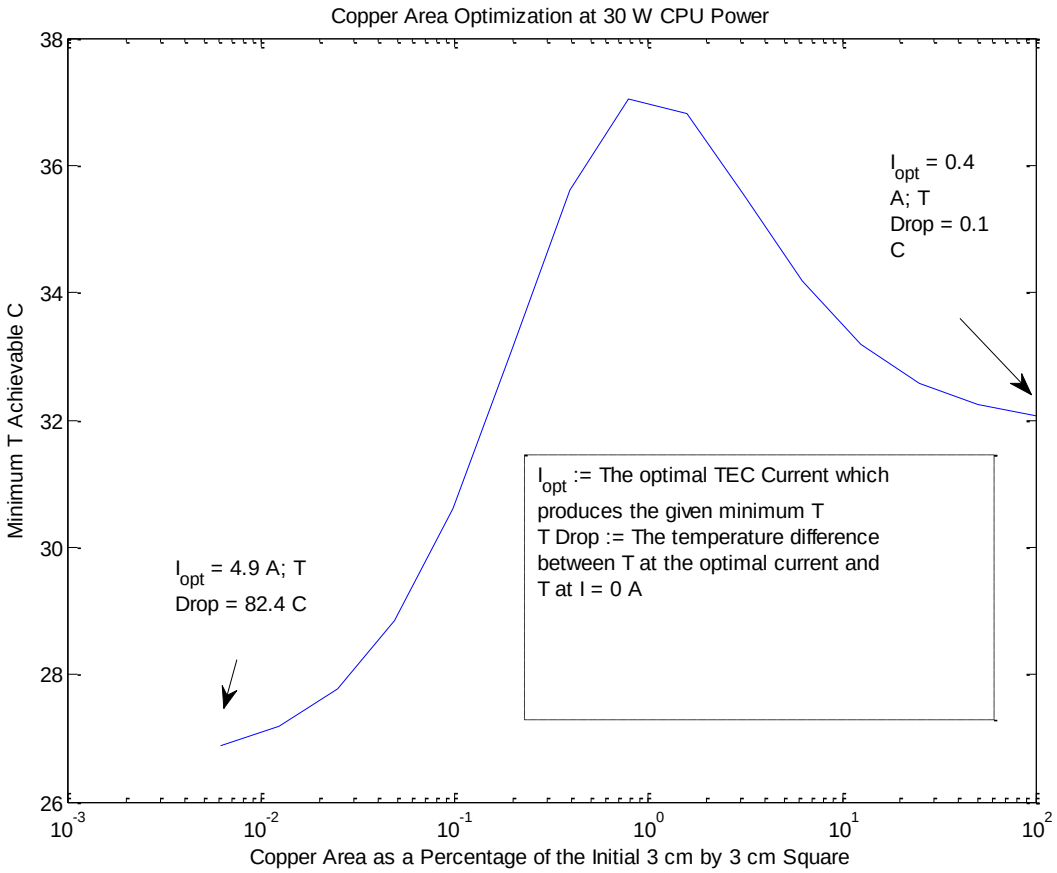

Figure 17 Copper Area Optimization at 30 W CPU Power

In Figure 16 and Figure 17, the x-axis shows copper area as a percentage of the 3 cm x 3 cm starting point. The x-axis is on a log-scale to highlight the asymptotic behavior at high percentages and low percentages. If copper area were increased above 100%, the temperature would reach some limit. The same goes if copper area were decreased even further. The temperatures predicted by the model are lower than what one gets on a real system with the TECs; the model underestimates some thermal resistances.

Here, though, only the general trends are relevant. When the CPU power is 40 W, the best performance is achieved when as much copper as possible is used. At 30 W, on the other hand, the best performance comes when as little copper as possible is used. Regardless of CPU power, using an intermediate amount

of copper gives bad results; this is not a general trend though, since applying 60 W of CPU power makes the maximum occur at the lowest percentage (see Figure 18). The TEC currents which produce the optimum temperatures are lower at higher percentages. For the 40 W case, TEC current is 4.9 A on the far left and 0.4 A on the far right. For the 15 W case, TEC current is 4.9 A on the far left and 0.4 A on the far right as well. At the high copper percentages, the TEC has minimal impact. Running the TEC at 0.4 A brings temperature down 0.1 C compared to keeping the TEC off. This emphasizes that adding copper basically eliminates any good effects TECs have. If a very large amount of copper were added, it would be best to keep the TECs off so they don't generate any extra heat due to Joule heating.

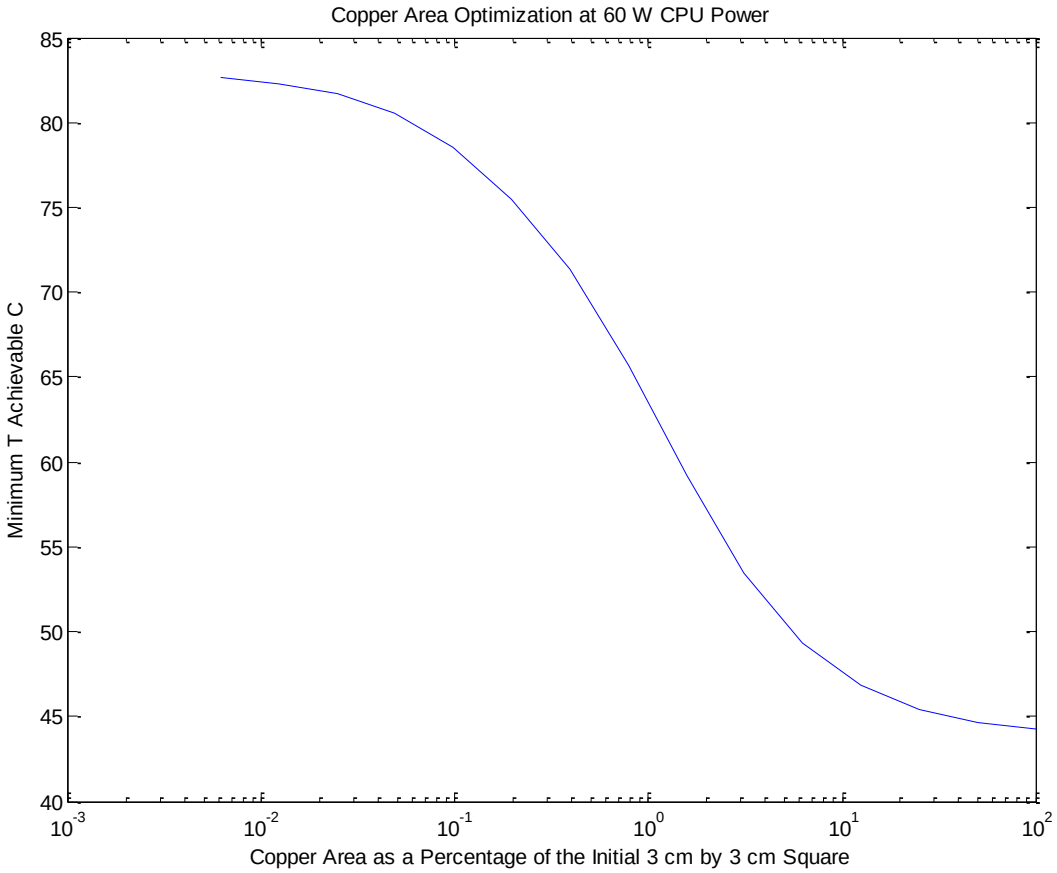

Figure 18 Copper Area Optimization at 60 W CPU Power

To maximize TEC effectiveness, using as little copper as possible is best. However, at some level of CPU power, the TEC thermal conductance K (0.19 W/K for a single TEC) becomes a liability. Temperatures get so high when the TEC is off that turning on the TECs doesn't bring temperature down enough. The TECs still perform some cooling. In the 40 W case in Figure 16, the temperature drop compared to the TECs being off is 94 C. It is just that this cooling is insufficient. Once this CPU power threshold is reached, it's best to have as much copper as possible. The TEC isn't serving its purpose, so it becomes best to effectively not have it at all.

The best design is to have no copper surrounding the TECs. Although the system may not be able to tolerate high CPU loads, it will be effective in cooling up to some point. Future analysis should consider whether these patterns hold when the TEC parameters are different (e.g. if thin films are used, the thermal conductance K will be higher).

TECs-on-Die vs. TECs-on-Spreader

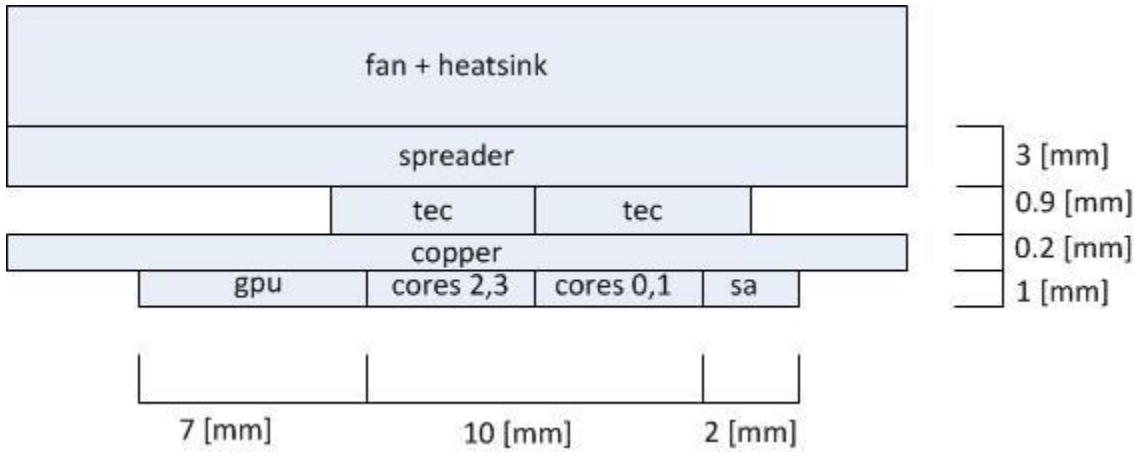

Figure 19 Alternative System Setup

The problem with the experimental setup is that it is ineffective at hotspot cooling. Ideally, engaging 1 TEC would cool one pair of cores more than the other pair. Based upon observation of the real setup, this is not the case. The reason is that the spreader distributes the cooling effect of the TECs equally over the die (see Figure 6 System Diagram). Rather than placing the TECs on top of the spreader, the

alternative setup in Figure 19 places the TECs more or less directly on the die. The copper above the die is very thin and is an attempt at mechanically protecting the die. The copper spreader is above the TECs to provide better thermal contact with the heatsink. The spreader distributes heat across its area and provides a larger surface for the heatsink to contact. These two systems are termed the "TEC-on-spreader" and the "TEC-on-die" systems.

To compare the two designs experimentally, the following system was setup. The system was configured twice, once with the TECs directly on the die, and once with the TECs on top of the spreader. For simplicity, this system used a dual core processor with a die size of roughly 1 cm^2. The quad core processor, on the other hand, has a die size of around 1.6 cm^2.

- Processor. Intel E7400 (Wolfdale) with 3MB cache. Based upon measurements by ruler, the die size is roughly 0.9 cm x 1.0 cm. The die is rectangular. The processor is a dual core processor, and has two thermal sensors, one per core. Interestingly, based on observation, these thermal sensors have a lower limit (roughly 30 C) below which they do not record temperature. The processor has 3 frequency settings: 1.6, 2.13, and 2.8 GHz.
- 2 TECs. The model is the same as used in the other experiments.
- Fan. The same 180 W solution used in the other experiments is installed in this system as well.
- Unlike in Figure 19, the setup did not include a protective layer of copper above the die for the TEC-on-die case.
- Miscellaneous. 2 GB RAM, 250 GB hard disk, the same OS as in the other experiments, Gigabyte GA-G31M-ES2L motherboard, OCZ 500 W power supply.
- Benchmarks. gcc, povray, and omnetpp at 1.6 and 2.8 GHz.

An important physical property of the setup is that 2 TECs take up 144 mm^2 of area, while the silicon die takes up only 100 mm^2. The TECs extend beyond the dimensions of the die. This is visually represented for one dimension in Figure 20 below. In the other dimension as well, the TECs are roughly 12 mm long while the die is only 10 mm long. The TECs are centered on the die so that each TEC lies over one of the cores. The TEC-on-spreader setup swaps the order of the spreader and the TECs.

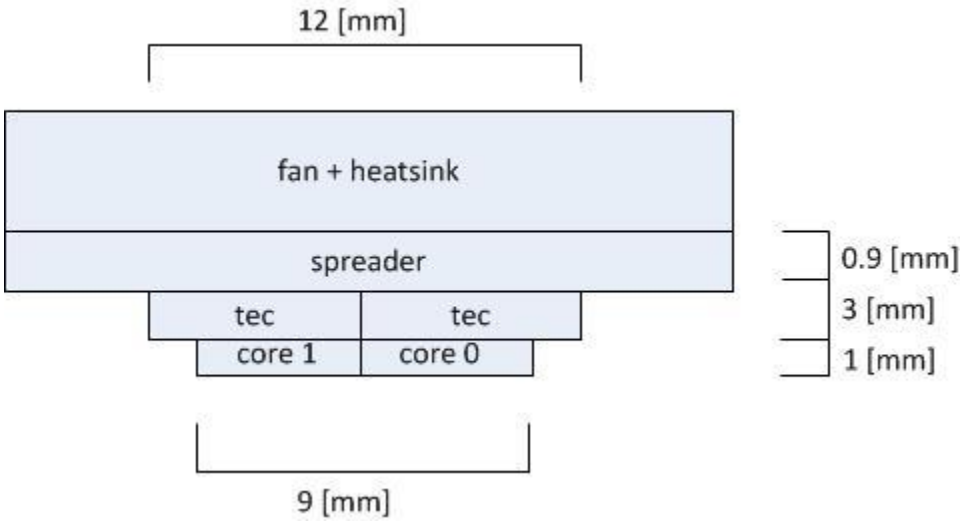

Figure 20 Dual Core System

The system thermal resistance K/W is depicted in Figure 21 below for both setups.

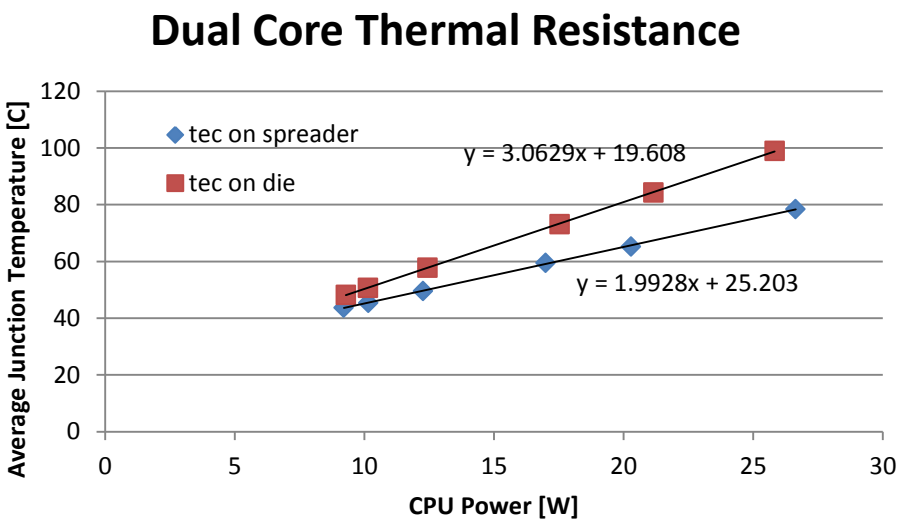

Figure 21 Dual Core Thermal Resistance: TEC-On-Die vs. TEC-On-Spreader. The system with the TECs on the silicon die has significantly higher thermal resistance K/W, as evidenced by the slopes of the lines.

Considering a simple model of the system, the thermal resistances of the two setups should be equivalent. Swapping the order of the TECs and the spreader should not make a difference. However, it turns out that the TEC-on-die system has a higher net thermal resistance, which is indicated by the slope of the line. It is roughly 1 K/W higher. One potential reason for this is that the entire lower surface area of the TECs does not contact the die. As mentioned above, the TECs have a larger area than the die. As a rough justification of this, imagine two extremes. First, suppose the die only contacted the TECs at a tiny point. In this case the net thermal resistance would be even worse. On the other hand, imagine, in Figure 20, a layer of copper in between the die and the TECs. This would improve performance by distributing heat to the entire surface of the TECs.

The thermal resistance of the dual-core TEC-on-spreader system is 2.0 K/W, while it is 1.7 K/W for the quad-core system (see page 35, Figure 11). Since the fan accounts for the majority of the thermal resistance, and the same fan is used in both the dual-core and quad-core setups, the thermal resistances should be closer to each other. The most likely explanation is that the quad-core setup was assembled in

a better fashion, with lower thermal contact resistances between surfaces. This anomaly is irrelevant to the discussion here because results for the dual-core system only are being compared.

In terms of hotspot cooling, however, the TEC-on-die system performs better. Povray at 2.8 GHz was run for 120 s, but only the TEC above core 1 was engaged. The average temperature over time was collected for each core. The TEC-on-die system runs hotter because of the higher thermal resistance. However, it's evident that core 1 is cooled disproportionately more than core 0. In the TEC-on-spreader system, the cores cool more uniformly.

	Avg. T junction core 0 (C)	Avg. T junction core 1 (C)
TEC-on-die	90	77
TEC-on-spreader	66	61

Table 4 Dual Core System Hotspot Cooling

A central problem with the TEC-on-die alternative design is that it is mechanically risky. First, the processor ships with the integrated heat spreader on top of the die (see Figure 8 on page 25). Removing the integrated heat spreader is difficult. Additionally, exposing the die risks damaging it; the heat spreader acts as mechanical protection. Once this mechanical protection is removed, the die is susceptible to cracks, for example. The TEC-on-die system is also difficult to design such that it has the same thermal resistance as the TEC-on-spreader system. Although hotspot cooling is an important aspect of thermoelectric cooling, for the purposes of this study, using the TEC-on-spreader system makes more sense.

4. Dynamic Thermal Management

Introduction

This chapter first characterizes the TEC system behavior, and then uses this information to implement a controller for the TECs. The system characterization is primarily a comparison of a standard system with no TECs to the TEC system. The comparison is done over a range of TEC currents. The system characterization also examines how dynamic voltage and frequency scaling (DVFS) couples with the TEC system. The controller optimizes performance by varying TEC current and processor frequency, subject to a temperature and power consumption constraint. The inputs to the controller are temperature, CPU power, and TEC power. The controller section examines how varying the temperature constraint and power consumption constraint affects the performance of the controller.

System Characterization

Procedure

The purpose of this set of experiments is to examine how thermoelectric coolers can complement dynamic voltage and frequency scaling. Typically, DVFS is used to trade off performance, power consumption, and temperature. For example, the adaptive thermal management system in the Intel i5-3450S uses DVFS to bring down temperatures if they ever exceed a critical limit [13, pp. 53–57]. Scaling down the clock frequency will lower dynamic power, and permits lowering voltage as well. Voltage affects static power consumption too. The cost of DVFS is performance, since the clock frequency is scaled down. TECs provide an alternative mechanism for decreasing temperatures, but come at the cost of consuming extra power consumption at the system level. As an example, consider running an application at 2.8 GHz that overheats the CPU. Rather than scaling down frequency, engage the TEC. Some performance benefit is gained at the cost of the power that the TEC consumes. This section considers this power-performance tradeoff.

Two benchmarks from the SPEC2006 suite were run on the system: povray and gcc. povray does ray tracing, and gcc is a C compiler. Only two benchmarks are used because the system characterization section is concerned mostly with average benchmark temperature and power consumption over time. When averaging over time, which benchmark is used doesn't really matter, because the transient behavior of the benchmark is smoothed away. Basically, each benchmark is run at a variety of {DVFS setting, TEC current} pairs. Also, there needs to be some reference against which the TEC system is compared. This reference is the system without the TECs, but with everything else identical. This system is termed the "standard system." Ideally, the TEC system should obtain temperatures equivalent to or lower than the standard system.

In more detail, for each benchmark, one trial consists of the following:

- On the TEC-on-spreader system
    - Run the benchmark at all combinations of DVFS setting and TEC current. The 13 available DVFS settings are between 1.6 and 2.8 GHz inclusive, in 0.1 GHz steps. Having 13 steps is advantageous because it allows the experimenter to vary frequency in very fine steps and infer general trends. The TEC currents used for these experiments were 0, 1, 2, 3, 4, and 5 A. TEC current refers to the current that each TEC receives. Each TEC has its own power supply. Generally, when TEC current is referenced it refers to the level of current that each TEC receives; e.g. 4 A refers to each TEC receiving 4 A of current. Regarding TEC power consumption, it is reported for both TECs combined unless otherwise specified.
    - For each combination of DVFS setting and TEC current.
        - Run two instances of the application. Running two applications makes the CPU generate sufficient levels of power to make the experiments interesting. Run

one instance on core 0 and the second instance on core 3. Core 0 and core 3 are on the edges of the block of cores. It is realistic to schedule the applications on these cores since the cores are as far apart from one another as possible. This will minimize core temperature.

        - Run the applications for 120 s. In between runs, wait for 120 s to let the CPU cool down and return to room temperature. When processing data, discard the first 30 s of runs. Analyze the data once the system has reached some steady state.
        - Sample data every second. The data includes multimeter readings, temperature sensor readings, etc. More details are in the section Experimental Setup.
- On the standard system,
    - For each DVFS setting, run the benchmark. Record CPU power consumption, etc. The specifications on number of applications, runtime, and sampling rate from above apply here as well.

Once the data is collected, the following analysis can be done.

- *Total power consumption*. The CPU power consumption and TEC power consumption can be totaled.
- *Energy (E) and energy-delay product (EDP) metrics*. When considering energy alone, the best metric would be runtime times average total power consumption. However, since applications are run for only 120 s, runtime isn't available. Runtime is $\frac{\#\ instructions}{f*IPC}$; IPC is instructions per cycle. An approximation of the metric that has similar properties is $\frac{f}{total\ power}$. Here, we are looking at the reciprocal of the power-delay product (energy), so higher numbers are better, since they indicate more performance per power. Alternatively, delay can be included in the

metric by considering the product of energy and delay. An approximation of this is $\frac{f^2}{total\ power}$, since EDP is runtime*energy. Again, higher numbers are better for this $\frac{f^2}{total\ power}$ metric. Both metrics allows comparisons between different scenarios involving a single benchmark. (For a given benchmark, IPC is roughly fixed, and increasing f will reduce the runtime).

- *Maximum, minimum, and average temperature*. This can be done on a per-core basis. Alternatively, temperature can be averaged over all 4 cores. The percent temperature difference between adjacent cores can be computed as well. The cores with applications running on them should be at higher temperatures, although how accurately the temperature sensors are calibrated will affect whether this really holds true.
- *Standard system comparison*. The temperatures achieved by the TEC system can be compared against the standard system temperatures.

Temperature vs. TEC Current

To begin with, the behavior for the benchmark gcc will be considered. Then the behavior for povray will be compared. The most basic plot of system behavior shows the maximum temperature reached by any of the cores during the duration of the run against TEC current. From a thermal management/reliability perspective, maximum junction temperature (rather than average junction temperature) is the relevant quantity. From the perspective of constructing a model for the TEC system, average junction temperature is a better quantity to consider.

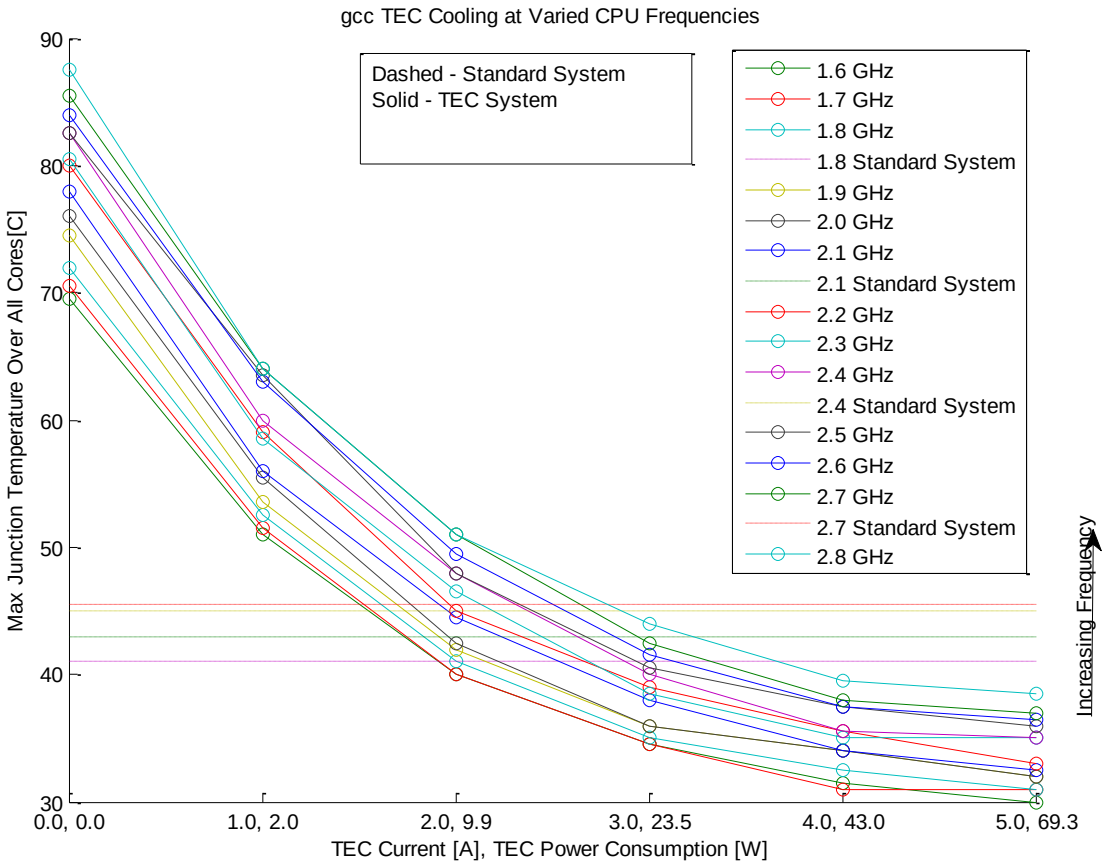

Figure 22 gcc TEC Cooling (Maximum Temperature). To help in distinguishing between curves, note that lower frequencies always run cooler.

In Figure 21, the x-axis shows the TEC current input into each TEC. It also shows the total power consumption of the TECs. TEC power consumption grows quadratically ($P = S\Delta TI + I^2R_e$). For a given current, all the frequencies had roughly the same TEC power consumption. That is, in this range, TEC power consumption did not vary much with CPU power consumption, even though ΔT is changing, because the Joule heating dominates. The curves show TEC system performance at different frequencies. The dashed lines show standard system performance at different frequencies.

Ultimately, a central question is whether the TEC system can cool below the standard system. In the case of gcc it can, regardless of frequency. Higher frequencies will have higher temperatures, which makes picking out which curve represents which frequency easier.

If the plot uses average temperature (over all cores) rather than maximum core temperature, as in Figure 23, the curves look more regular. Maximum temperature is a hard-to-predict quantity. For example, with a given TEC current, sometimes several different frequencies obtain the same maximum core temperature. Average temperature, on the other hand, will smooth away time variations in CPU power for a benchmark. In terms of cooling to standard system levels, both maximum and average temperature give similar results. In both cases, between 2 and 3 A of TEC current (10 to 20 W of TEC power) is required to cool to standard system levels. To give a specific case, for gcc 1.8 GHz, roughly 2 A (10 W of TEC power) must be used to cool to the 1.8 GHz standard system level.

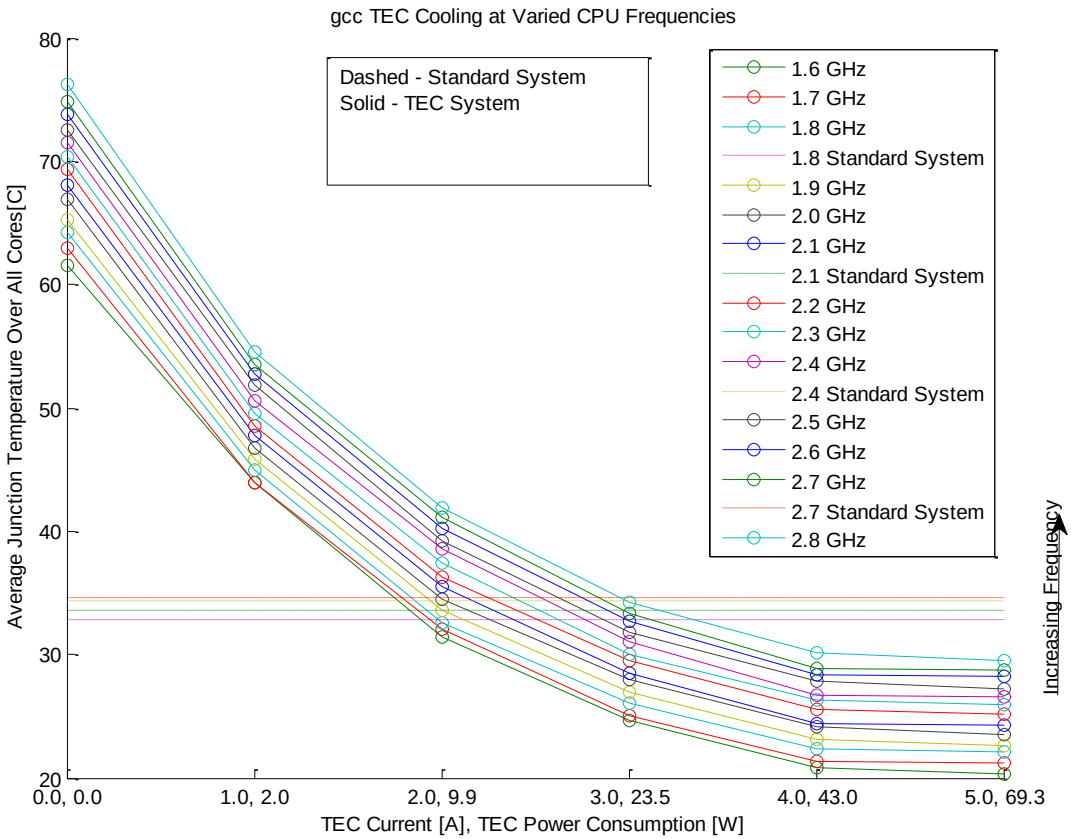

Figure 23 gcc TEC Cooling (Average Temperature). To help in distinguishing between curves, note that lower frequencies always run cooler.

As noted above, TEC power consumption does not vary much as a function of CPU power. At 1 A TEC current, power consumption is roughly the same (+/- 0.5 W) at both 1.6 GHz and 2.8 GHz. Thus, Figure 24 above shows TEC power consumption averaged over all frequencies.

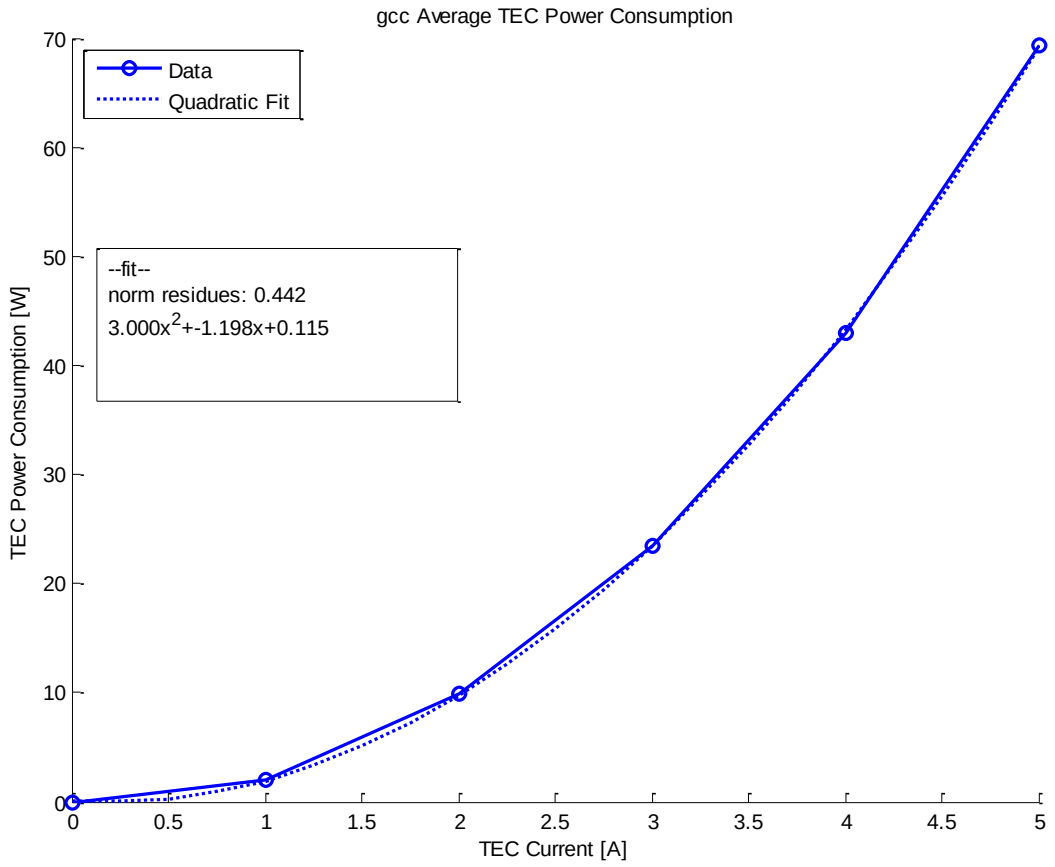

Figure 24 gcc TEC Power Consumption

Since many of the plots here are done for different DVFS settings, it is worthwhile noting that CPU power as a function of DVFS setting is described by the following equation $P = \alpha f V^2 + \beta V$. The Greek letters represent constants, f represent frequency, and V represents processor voltage. The first term represents dynamic power due to switching and the latter term represents leakage; dynamic short-circuit power is neglected because in modern processors it has a negligible effect [34, Ch. 5]. Alternatively, the short-circuit power can be considered a fixed percentage of the switching power, and α can be modified appropriately. α depends on factors like the capacitance on the microprocessor routing, and β depends on factors like leakage current. If voltage stays relatively fixed as frequency increases, the growth in power will be linear as a function of DVFS setting. However, the trend could

potentially be different because voltage needs to increase to enable increases in frequency. The CPU power consumption as a function of DVFS setting is below in Figure 25, for reference. The trend looks mostly linear.

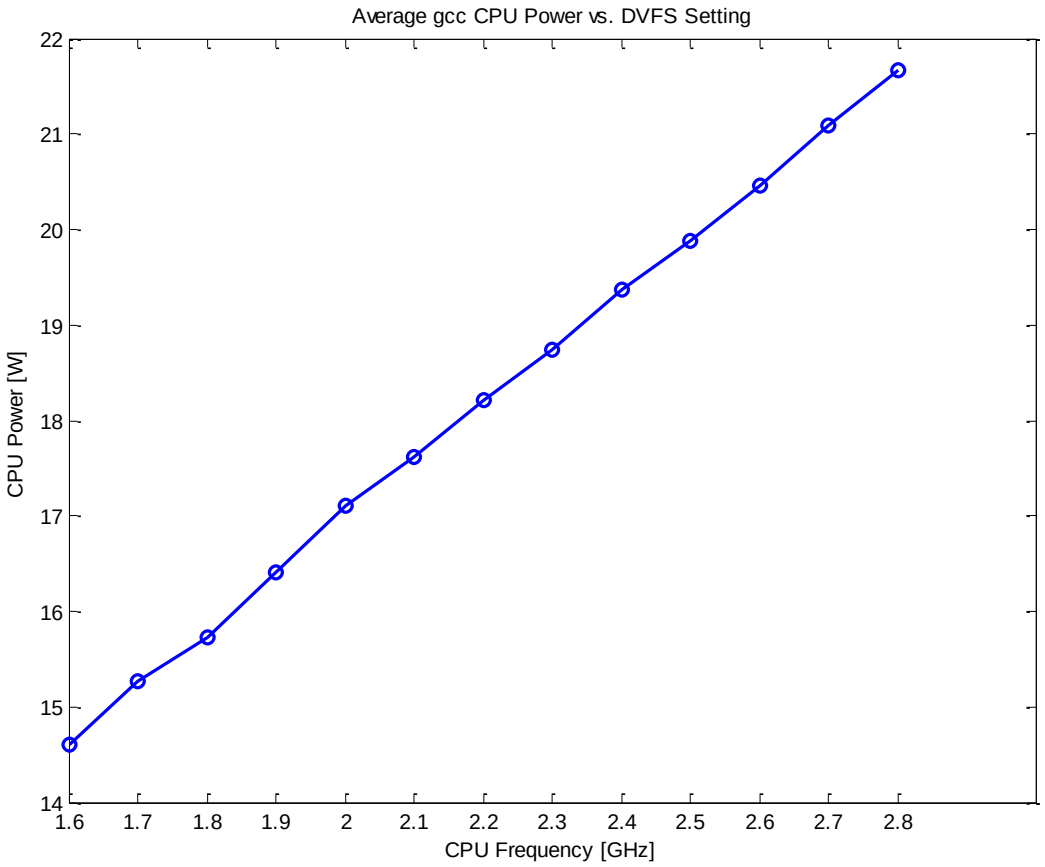

Figure 25 gcc CPU Power Consumption vs. DVFS Setting

The behavior of the experimental data against the analytical model was considered previously in the Steady State TEC Equations section on page 29. Here, it suffices to say that the data does match expectations. To demonstrate, pick a particular frequency for the TEC system. The rate of temperature change decreases as current increases. As current increases, there are diminishing returns.

The effectiveness of the TEC diminishes at higher frequencies. In Figure 22 on page 57, consider operating at 5 A TEC current. One notices that the difference between the 2.7 GHz standard and TEC

plots is smaller than the difference between the 1.8 GHz standard and TEC plots. Figure 26 displays this trend. It shows how far below the standard system the TEC system can cool, but a function of CPU power rather than DVFS setting. As CPU power increases, the TEC system's ability to cool below the standard system tapers. The plot shows much random variation since it uses maximum temperature, but the overall trend is visible.

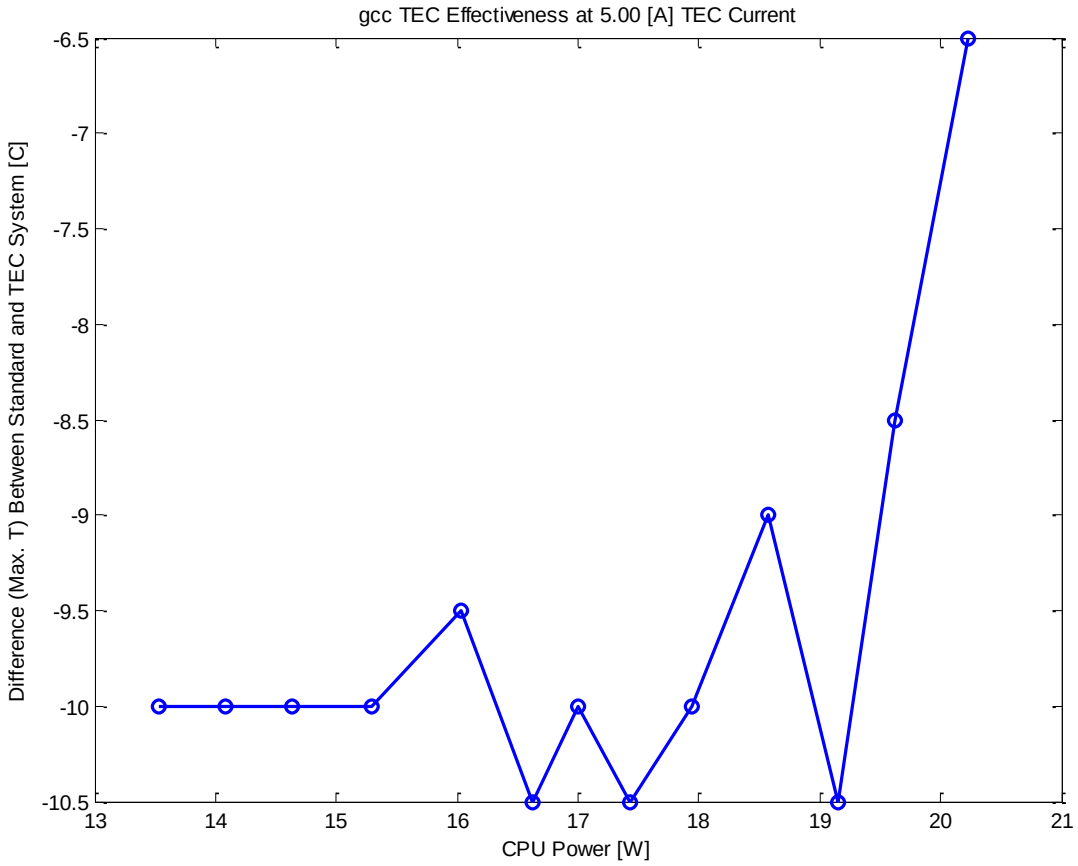

Figure 26 gcc TEC Effectiveness

The theoretical explanation for this is that the derivative of TEC cold side temperature with respect to CPU power is $\frac{dT_c}{dP} = \frac{1-RSI+KR}{SI-RS^2I^2+K}$ (see Steady State TEC Equations on page 29). Assigning typical parameters of R = 0.4 K/W, S = 0.021 V/K, and K = 0.83 W/K (experimental value), we get $\frac{dT_c}{dP} \approx (1 - .01I + .33)/(.021I - 0.0002I^2 + .83)$. At a current of 4 A, this evaluates to 1.4 K/W, while at 0 A this

evaluates to 1.6 K/W. On the other hand, with the standard system, temperature grows at a rate of about 0.44 K/W. As power increases, temperature grows at a faster rate in the TEC system than in the standard system. The difference in rates is at best 1 K/W. This accounts for the decrease in effectiveness. The slope of the line in Figure 26 (at some intermediate point) is closer to 0.5 K/W. A discussion similar to this one was given in the System Thermal Resistance section on page 34.

Consider operating at 2.8 GHz on the TEC system, and the amount of TEC power required to cool to various standard system levels. Figure 27 below shows how much TEC power is required to cool to the 1.6 GHz standard system, 1.7 GHz standard system, etc. Again, this plot shows random variation because it is computed based upon maximum temperature rather than average temperature.

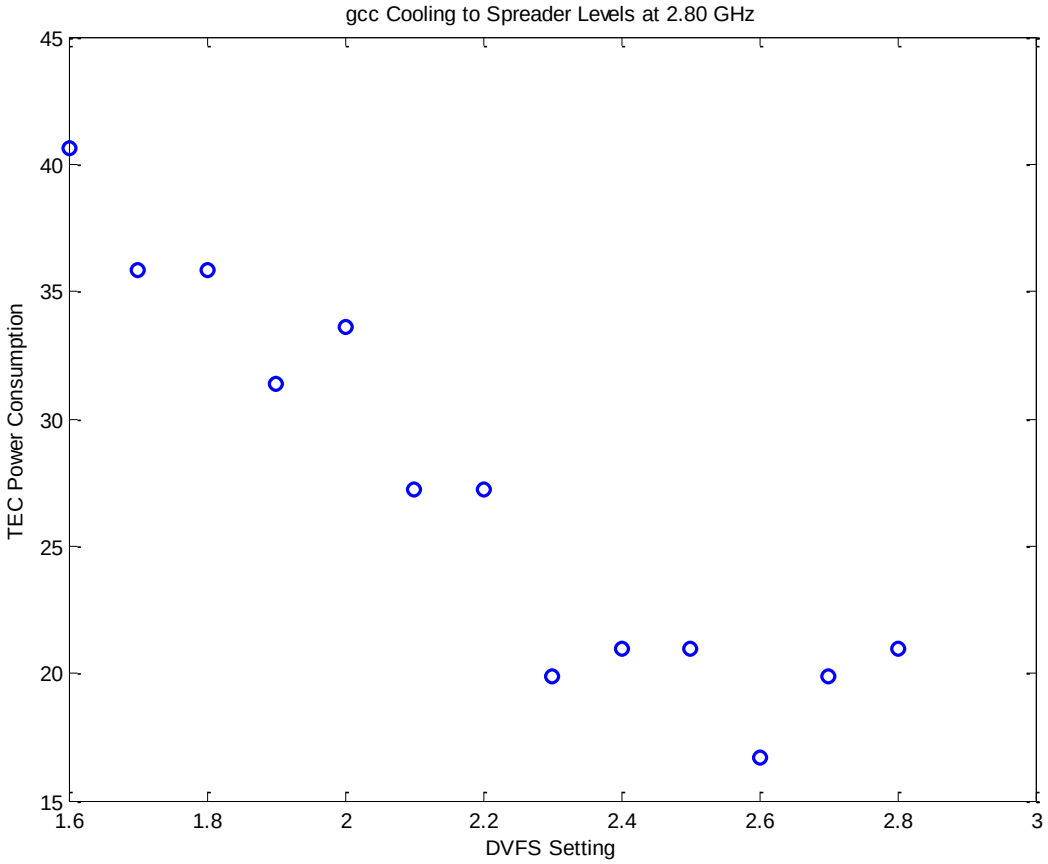

Figure 27 gcc Cooling to Spreader Levels at 2.8 GHz

2 gcc applications by themselves consume around 14 W at 1.6 GHz and 20 W at 2.8 GHz. It's a bad sign that the TECs usually consume more power than the application. Consider the frequency-per-total-power, or 1/E, metric, which should be as high as possible. The 2.6 GHz plot at 3 A roughly intersects the 1.8 GHz standard system line in Figure 22 on page 57, so these two 41 C points are good for comparison. Some roughly equivalent operating points are listed in Table 5 below. Although 2.6 GHz gives more performance, it costs disproportionately more power. The same is true with 1.9 GHz, which gives just incrementally better performance. The metric at 2.6 GHz and 2 A is 0.08, but the cooling is only to 50 C. All of this assumes that performance will scale linearly with frequency, although this might not always be the case. An extreme case is an application that spends most of its time waiting for memory or IO operations to complete rather than doing computation.

Frequency (GHz)	TEC Current (A)	1/E Metric (GHz/W)	1/EDP Metric (GHz^2/W)	Max. Core T (C)	Note
1.8	N/A	0.1	0.174	41	Standard system
1.9	2	0.07	0.139	42	
2.6	3	0.06	0.158	41.5	

Table 5 gcc 1/E and 1/EDP Metrics at Equivalent Operating Points. Note that larger values for the metrics are better.

Even with the 1/EDP metric (which should be as high as possible) the 2.6 GHz and 3 A operating point is worse than operating at 1.8 GHz and 0 A. If delay were given even more priority, say with a $1/ED^2$ (energy delay-squared) metric, then the 2.6 GHz operating point would look attractive. For example, {2.6 GHz, 3 A} would have an $f^3/(total\ power)$ of 0.41, while {1.8 GHz, 0 A} would have a metric of 0.31. It is difficult to assign meaning to the specific numbers the metrics take on. However, the broader point is that the TECs provide performance gains but are not energy efficient. If one really values performance over energy consumption, then the TECs are attractive.

For both the E and EDP metrics, the metric has a maximum as TEC current increases for a fixed frequency, as shown in Table 6 for 2.8 GHz gcc. This suggests that if the TECs were slightly more power

efficient – that is, if the optimal coefficient of performance were at a higher current – they would be a design win.

TEC Current (A)	1/E (GHz/W)	1/EDP (GHz^2/W)
0	0.110	0.307
1	0.116	0.325
2	0.091	0.255
3	0.064	0.178
4	0.044	0.124
5	0.031	0.088

Table 6 2.8 GHz gcc 1/E and 1/EDP Metrics. Note that larger values for the metrics are better.

Extending the notion of equivalent operating points, the following contour plot in Figure 28 shows which {DVFS,TEC current} pairs are equivalent in terms of operating temperature. In Figure 22, for example, the standard system 1.8 GHz temperature is 41 C. Based on the plot below, 41 C can be met at 2.8 GHz and 4 A, or at 1.6 GHz and 2 A. Both CPU power and TEC power are listed on the plot so performance-per-power metric values can easily be computed by hand if necessary.

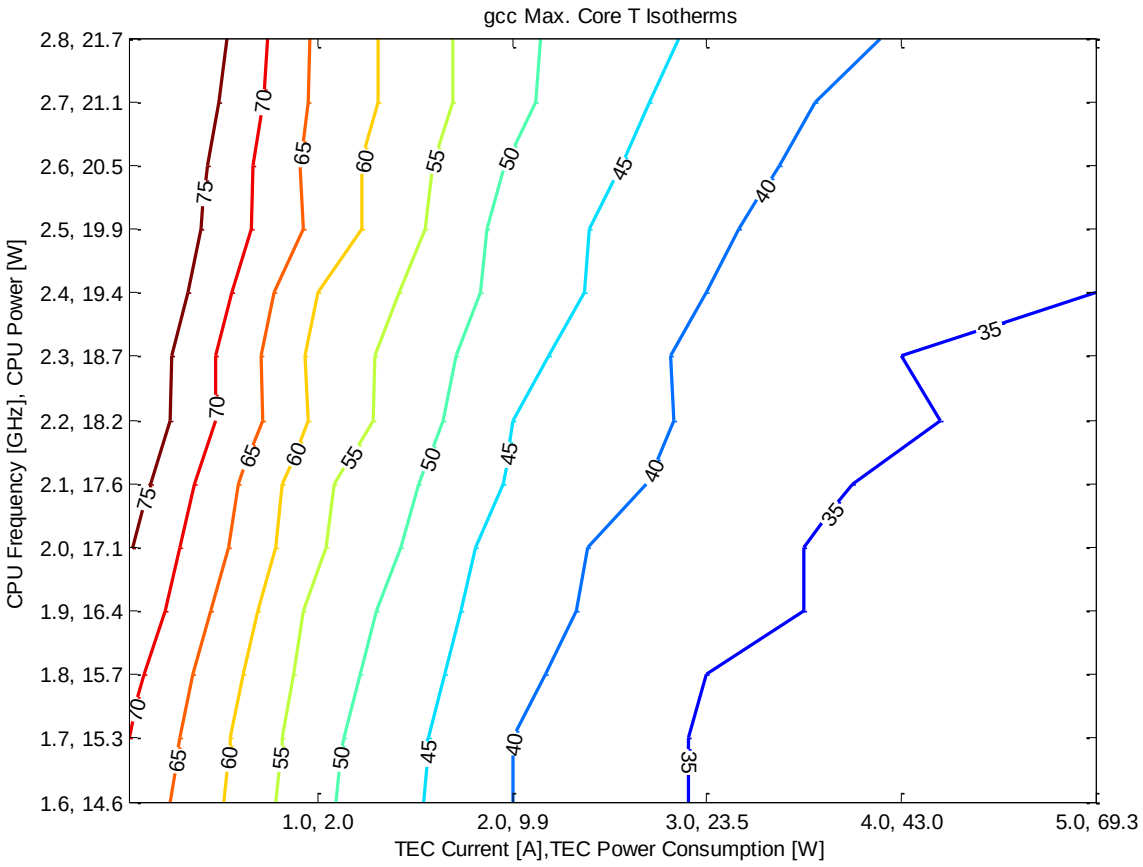

Figure 28 gcc Isotherms

For a fixed frequency, the curves get increasingly spaced as current increases. This matches up with the temperature vs. current curves, which show that increasing current has diminishing returns. Put another way, to obtain a fixed temperature change at each step, more and more increase in current is required. For fixed current, on the other hand, the curves are equally spaced. This matches up with the TEC effectiveness curve, which shows the behavior of temperature as a function of power. The TEC effectiveness curve should be roughly linear based upon calculated values, so the equal spacing in the contour plot makes sense.

The new insight that the contour plot provides is that, for a fixed temperature, as CPU power increases, more and more current is required. The effect is more pronounced at lower temperatures. At low

temperatures, the isotherms turn quite sharply to the right. This shows that to maintain a fixed temperature, disproportionately more power from the TEC is required as CPU power increases.

In terms of isotherm plot inaccuracies, the plots were generated by linearly interpolating the temperature vs. current curves, even though the temperature vs. current curves are nonlinear. So, the contour plot is only an approximation. The lack of smoothness in the curves comes from the interpolation, and from random experimental variation.

Benchmark Comparison

A benchmark that consumes more power than gcc is povray. The overall trends in terms of DVFS and TEC tradeoff are similar. The main difference between gcc and povray is that, since CPU power increases for povray, the TEC becomes less effective. The TEC system is less and less able to reach the temperature levels of the standard system. At 2.8 GHz gcc was able to cool 6.5 C below standard levels, while povray is only able to cool 4.5 C below, as seen in Figure 29. Extrapolating the trend, by 35 W of CPU power consumption, the TEC system will not be able to cool to the standard system levels.

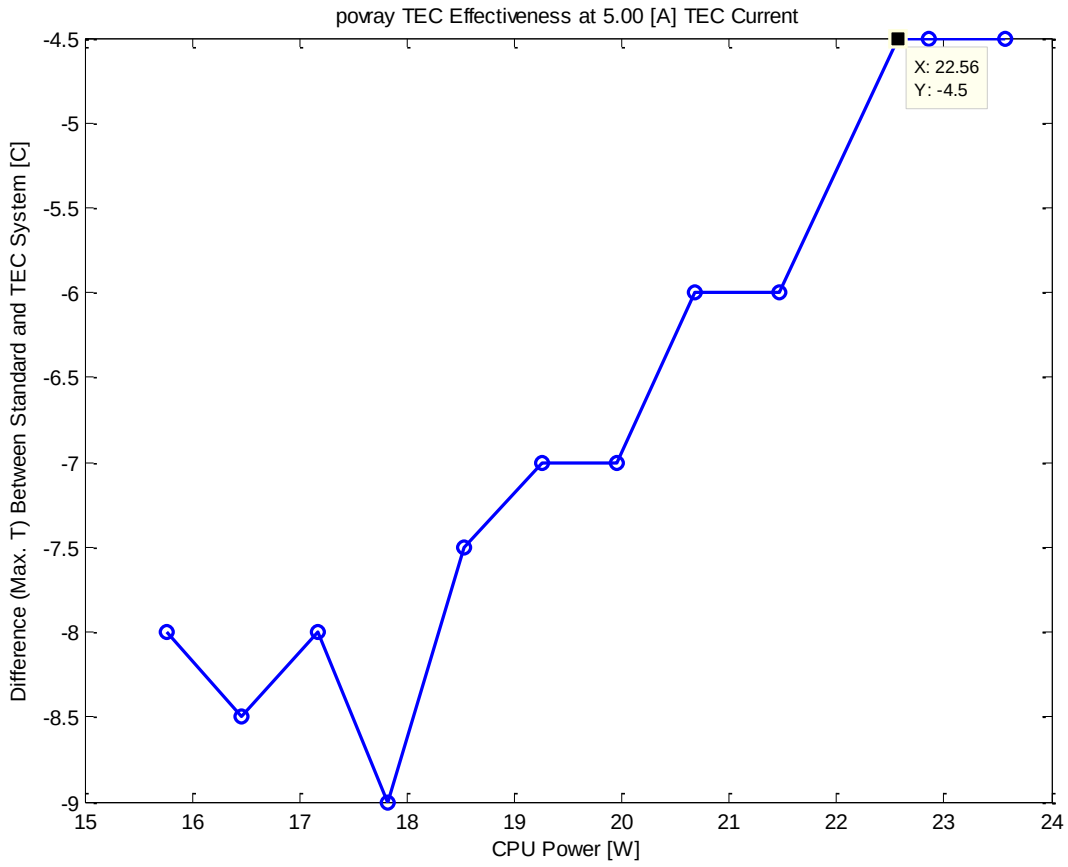

Figure 29 povray TEC Effectiveness. This figure shows how far below the standard system the TEC system can cool, as a function of CPU power. This plot is analogous to Figure 26, except that this plot is for povray rather than gcc.

A final point is that the overall system behavior does not vary significantly with benchmark. As long as two benchmarks are consuming the same amount of power, then behavior is similar, when averaged over time.

Thermal Management

Given the understanding of DVFS and TECs developed, this section implements a simple feedback controller for the thermoelectric coolers. The controller will attempt to maximize performance while adhering to temperature and power constraints. More concretely, the controller optimizes benchmark runtime, since runtime is the best measure of performance. The temperature constraint is a restriction

on the maximum die temperature, which is measured using the digital thermal sensors embedded in the microprocessor. The power constraint accounts for the power consumption of both the CPU and the thermoelectric coolers. To achieve these criteria, the controller will vary the CPU operating frequency and the TEC current.

A central theme will be the comparison of a DVFS-only controller with a DVFS+TEC controller. The DVFS-only controller attempts to meet the constraints by only varying CPU frequency. On the other hand, the DVFS+TEC controller has both the TECs and DVFS (dynamic voltage and frequency scaling) at its disposal. Consider a scenario in which the junction temperature exceeds the maximum allowed temperature. Rather than decreasing processor frequency, the controller can engage the TECs to bring temperature down while maintaining performance. Finally, this section examines how varying the values of the constraints (e.g. choosing a temperature constraint of 70 C rather than 80 C) affects controller performance.

Controller Design

The easiest way to describe the controller is in a flowchart. First, here are some definitions.

- I – TEC current, in A. Increasing TEC current will bring down temperature, but at the cost of additional power consumption. Decreasing current will increase temperature, but decrease power.
- f – CPU frequency, in GHz. Increasing frequency will increase temperature and power. Decreasing frequency will decrease temperature and power.
- P – The current total power consumption, in W. The total power is the sum of the CPU and TEC power consumption.
- T – The highest temperature out of all the cores, in C.
- T_{max} – The temperature constraint, in C.

- P_{max} – The total power constraint, in W.
- I_{max} – The maximum current that the TECs can use, in A. This can be a physical device limit that the manufacturer specifies. It can also be the optimal current, discussed in the section Steady State TEC Equations, after which the TECs stop cooling to lower temperatures, even though they consume more power. Note that in doing the system characterization, regardless of CPU power, this optimal current hovered around 4 A. Therefore, in the controller design, I_{max} will be fixed at 4 A.
- I_{min} – A minimum threshold below which the TEC current cannot drop, in A. This will usually be 0 A. A case in which I_{min} might be nonzero is when the system requires the TEC to be on continually to prevent overheating.
- f_{min}, f_{max} – The maximum and minimum processor frequencies, in GHz.
- T_{hyst} – The temperature hysteresis built into the controller, in C. For example, if the operating temperature is 1 C below the constraint, it may be better to maintain the present state rather than increasing the frequency. Stepping the frequency up will most likely cause the constraint to be exceeded. This 1 C threshold is T_{hyst}.
- P_{hyst} – The power hysteresis built into the controller, in W. This is similar to T_{hyst}. If the current power is only slightly under the constraint, it may be better not to increase frequency or current.
- The controller rate is the rate at which the controller makes decisions. For example, the controller might re-evaluate and decide new current and frequency settings every 10 seconds.

In addition, there are two operations.

- ++. This increments the variable by a predetermined step-size. For example, I++ will increment the TEC current by a step. f++ will increment CPU frequency by a step. On the processor used in

these experiments, the minimum step size is 0.1 GHz. f and I can be considered as discrete variables.

- --. This decrements the variable by the predetermined step-size.

The flowchart for the controller algorithm is below in Figure 30. The flowchart considers the case of the hysteresis values being 0, to avoid adding unneeded complexity. See the Controller Psuedocode section in the appendices for a complete description of the algorithm. Also, the flowchart assumes that the increment and decrement operations tell the user whether they succeeded or not. The operations will fail if an increment would push the variable beyond its maximum constraint (i.e. f_{max}, I_{max}), or a decrement would drop the variable below its minimum constraint (i.e. f_{min}, I_{min}). A fundamental assumption of the controller algorithm overall is that TEC current is capped at the optimal current, so that increases in current always produce drops in temperature.

This algorithm executes at every timestep the controller is invoked. There are four main cases to consider: (1) neither the temperature nor the power threshold is exceeded (2) the temperature constraint is exceeded but the power constraint is not (3) the power constraint is exceeded but the temperature constraint is not (4) both constraints are exceeded.

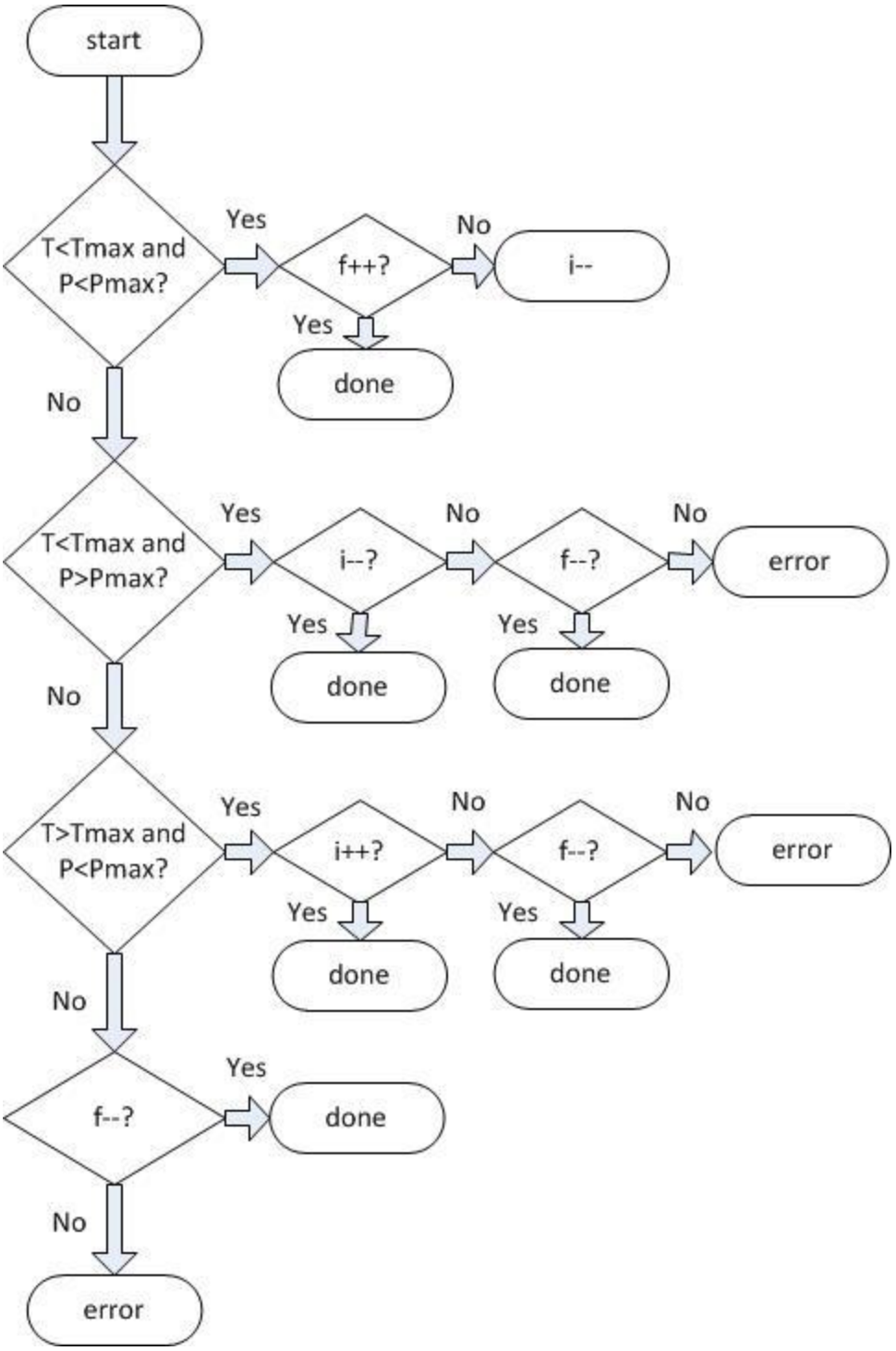

Figure 30 Controller Algorithm Flowchart. The diamonds represent decision points, and the ovals represent starting or ending points. The -- and ++ operations, in addition to decrement or incrementing, return to the user whether they executed successfully or not.

In words, if neither constraint is violated, the controller first attempts to increase frequency to improve performance. If it cannot do so because the maximum frequency has been reached, it decreases current to save power. If only the power constraint is violated, the controller first attempts to decrease TEC current, since some thermal slack is available. If TEC current is at the minimum, the controller decreases frequency. On the other hand, if only the temperature constraint is violated, the controller first increases TEC current if possible, in order to cool the chip down. TEC current can increase because there is some portion of the power budget available. If that is not possible, frequency is decreased. Finally, if both constraints are violated the controller attempts to decrease frequency. It doesn't attempt to decrease current, because doing so would exacerbate the temperature constraint violation.

One takeaway is that the controller always tries to increase frequency if possible. Also, the controller always attempts to increase TEC current before decreasing frequency, in order to maintain optimal performance. In the T<Tmax and P<Pmax case, the current decrement is done to save power.

The parameters available for tuning the controller are: the constraint values; the hysteresis values; the predetermined step sizes at which current and frequency change; the controller rate. Also, if the TEC current is fixed at a particular value, then the only variable available for manipulation is the processor frequency, so a DVFS-only controller is obtained.

Experimental Results

Two experiments were conducted. Unlike with the system characterization experiments, here the benchmarks were run to completion. The runtime of the benchmark is the ultimate measure of performance.

First an unlimited P_{max} was used at T_{max} of 70, 75, 80, and 85 C to observe how, in the best case, the DVFS+TEC controller improves on the DVFS-only controller. Three gcc applications were run. Note that the DVFS-only controller fixed TEC current at 0.5 A rather than 0 A because of a minor power supply

issue detailed below. This is essentially irrelevant because at 0.5 A the TECs neither consume much power (around 0.1 W) nor have any significant cooling effect. 0.5 A is mostly equivalent to 0 A in this case.

In the second experiment, two gcc applications were run. P_{max} was varied at various fixed T_{max} constraints. The T_{max} constraints chosen were 45, 50, 55, and 60 C. This low range was chosen because the standard system operates in these temperatures when running two gcc applications run at once. The idea is to create realistic scenarios in which the TEC system is actually cooling below standard system levels.

Table 7 summarizes the experimental setup details. Some of the motivations behind the experimental setup are discussed in the bullet points below.

	Experiment 1	Experiment 2
T_{hyst} (C)	5	5
P_{hyst} (W)	3	3
I_{min} (A)	0.5	0.5
I_{max} (A)	0.5 (DVFS-only) or 4 (DVFS+TEC)	4
f_{step} (GHz)	0.1	0.1
I_{step} (A)	0.5	0.5
Controller Rate (Hz)	0.1	0.1

Table 7 Setup for Controller Experiments

- The SPEC2006 benchmark gcc was run. gcc was chosen as the benchmark to run because it has quite variable power consumption over time and is a good challenge for the controller. Either two or three instances of the benchmark were run at the same time. The affinities were set to cores 0, 1 or 0, 1, 3. Running two instances is convenient because the system characterization used two instances of gcc. Running three instances provides some diversity. Future work should consider a wider variety of benchmarks, since each benchmark has different behavior over time. For example, some benchmarks like hmmer have relatively flat power consumption over time, while others are more variable.

- The controller triggered every 10 s. Data on temperature, power, etc. was externally logged every 1 s.
- The minimum TEC current was set at 0.5 A. If the power supplies for the TECs are set to 0 A, and an application is running on the CPU, the power supplies get put into an error state. The reason is that when the TECs are not engaged, they develop a voltage due to the Seebeck effect. The power supplies see this voltage, and disable themselves when the voltage gets large enough. The DVFS+TEC controller must keep the TEC current at a minimum of 0.5 A to prevent the power supplies from being disabled, since it engages them from time to time. To make a fair comparison between the DVFS-only and DVFS+TEC controllers, the DVFS-only controller maintains a fixed TEC current of 0.5 A as well.
- The experiments described here were performed on the quad-core Intel i5-3450S processor, using two TECs. Each TEC received the same amount of current, since the experimental setup is not conducive to hotspot cooling. Any operating point achieved by running each TEC at different currents x and y can be achieved by running each TEC at some identical current z. Using non-identical currents would simply add complexity to the controller. As a reminder, whenever TEC power is mentioned it refers to the power consumed by both of the TECs combined.
- Each experiment was performed only once.

The first set of data concerns a controller with no power constraint. This data will contrast a DVFS-only controller and a DVFS+TEC controller. The DVFS-only controller has a fixed current of 0.5 A. The DVFS+TEC controller has I_{max} = 4.0 A. Plots of the two controllers for 3 applications of gcc running simultaneously are in Figure 31 and Figure 32.

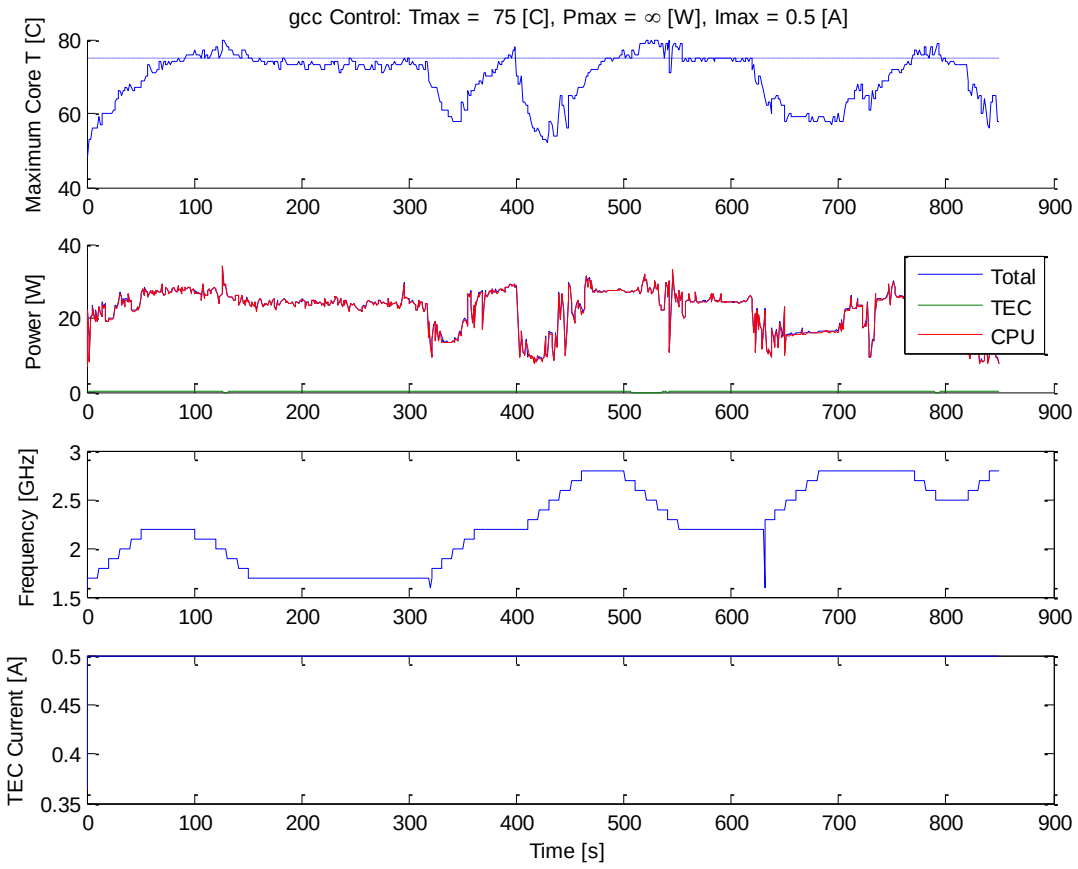

Figure 31 DVFS-Only Controller with No Power Budget. With this controller, TEC current is fixed at 0.5 A, as seen in the plot of TEC Current. The TEC power consumption is negligible, and the total power consumption matches the CPU power consumption. In the frequency plot, the glitch between 600 and 700 seconds is unaccounted for, and may be due to the processor's own adaptive thermal management. The dashed line in the temperature plot is the temperature constraint.

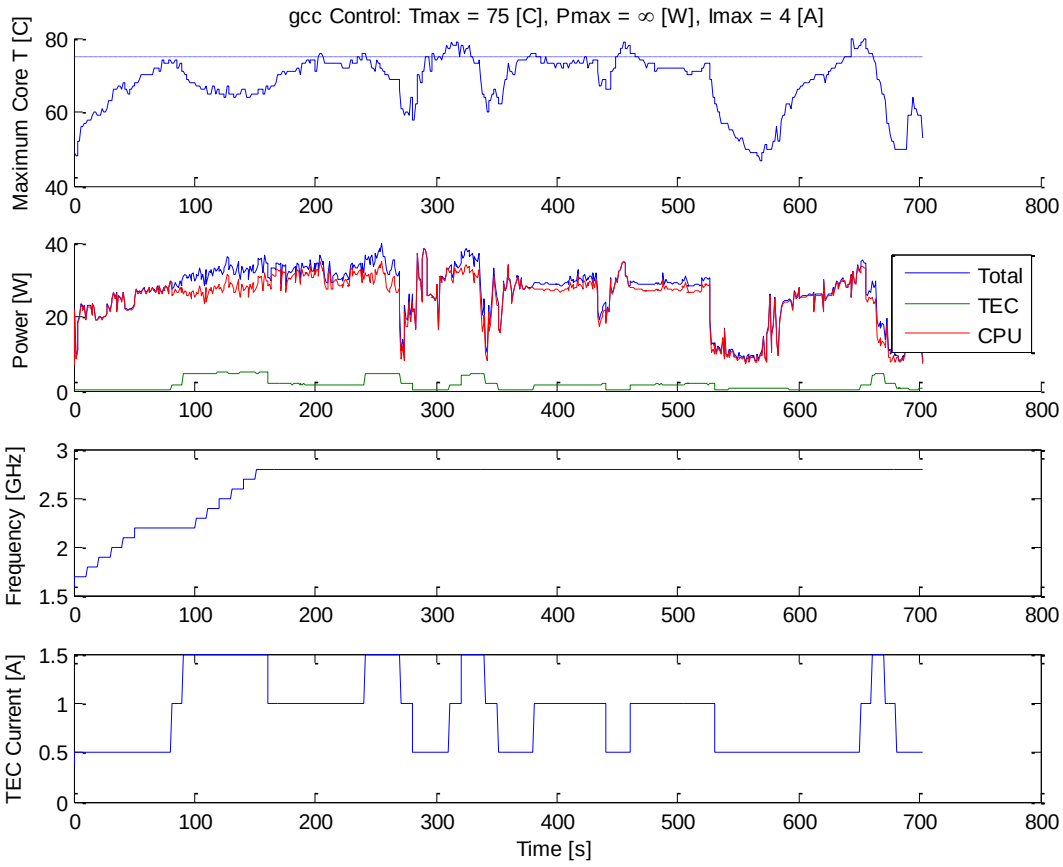

Figure 32 DVFS+TEC Controller with No Power Budget. In the plot of power, to distinguish TEC power note that it is the lowest of all three types plotted. The transient behavior of the temperature is visible. For example, the TEC ramps up to 1.5 A at 100 s but the temperature only reaches a steady state at 150 s. The 100 s point is also notable because it shows the controller increasing the TEC current in order to permit an increase in frequency.

Both controllers are relatively effective at maintaining temperature below the constraint. They both also maintain the temperature in a stable fashion. However, the DVFS-only controller must decrease frequency at times. On the other hand, the DVFS+TEC controller reaches 2.8 GHz and stays there. For this 75 C temperature constraint, the average operating frequency improves by 19%, and runtime decreases by 17%. Note that increases in frequency may not always create decreases in runtime, especially if the benchmark is bottlenecked on IO or memory operations.

The energy consumption (average total power times time) of the DVFS+TEC controller is 18840 J while for the DVFS-only controller it is 19272 J. The fact that the DVFS+TEC controller consumes less energy

seems to be a contradiction, since the System Characterization section (page 53) emphasized that the TEC system is not energy efficient. However, it should be noted that here a comparison is not being done between (a) the DVFS-only controller on the standard system, and (b) the DVFS+TEC controller on the TEC-on-spreader system. Instead, the DVFS-only controller is running on the TEC-on-spreader system as well. Thus, these experiments are not meant to contrast the energy consumption of the two controller types. They simply show that the controller is designed well enough to give increases in frequency and decreases in runtime. The ideal experiment (which should be conducted in the future) would compare the DVFS-only controller on the standard system with the DVFS+TEC controller on the TEC-on-spreader system, while using much lower temperature constraints closer to 45 C. This type of experiment would accurately characterize the energy tradeoffs. The experiments described later in this section run at lower temperature constraints, and are used to approximate this ideal experiment.

A drawback of both controllers is their response time. For example, it takes over 100 s for the DVFS+TEC controller to reach 2.8 GHz. The leaves room for improvement using a proportional or proportional-integral controller.

Table 8 examines the DVFS-only and DVFS+TEC controllers at different temperature constraints. The table does not show energy consumption because, as described above, it is a misleading statistic in this case.

		Tmax 70 (C)		Tmax 75 (C)		Tmax 80 (C)		Tmax 85 (C)	
	Controller type	DVFS -only	DVFS+ TEC	DVFS -only	DVFS+T EC	DVFS -only	DVFS+ TEC	DVFS -only	DVFS+ TEC
Benchmark Data	Maximum Real Temperature Reached (C)	75	79	80	80	88	84	91	89
	Average f (GHz)	1.8	2.7	2.2	2.7	2.4	2.7	2.6	2.7
	Average I (A)	0.5	1.0	0.5	0.9	0.5	0.7	0.5	0.6
	Average TEC Power (W)	0.2	2.4	0.1	1.7	0.1	1.0	0.1	0.6
	Average CPU Power (W)	21.0	23.5	22.5	25.1	23.7	25.2	24.6	27.6
	Average Total Power (W)	21.1	25.9	22.7	26.8	23.7	26.2	24.7	28.2
	Runtime(s)	806	764	849	703	820	723	818	679
	Percentage of Time in Violation of Constraints	25	10	15	10	12	8	8	7
Calculations	**Percentage Increase in CPU Frequency**	--	46.4%	--	19.3%	--	11.2%	--	5.1%
	Percentage Decrease in Runtime	--	5.2%	--	17.2%	--	11.8%	--	17.0%
	(TEC Power)/(Total Power) as a Percentage	0.7%	9.1%	0.5%	6.2%	0.4%	4.0%	0.4%	2.0%
	Maximum Overshoot (C) (Maximum Real T - Tmax)	5	9	5	5	8	4	6	4

Table 8 DVFS+TEC vs. DVFS-only Comparison. In these experiments, Pmax = ∞ (really 200 W). The DVFS-only controller uses Imax = 0.5 A. The DVFS+TEC controller uses Imax = 4 A. Total power is the sum of the CPU power and TEC power. The values in the table are rounded, TEC and CPU power may not add up to the Total exactly. Note that even if CPU frequency corresponded to runtime exactly, the percentage increase in frequency would not equal the percentage decrease in runtime since they are both slightly different calculations; use the formula runtime = (# instructions/(f*IPC)) to see this.

Based upon the data, in the best case (a 70 C constraint) the DVFS+TEC controller provides a 46% improvement in frequency. The maximum possible improvement is 75%, since f_{max} = 2.8 GHz and f_{min} = 1.6 GHz. The best improvement in runtime was 17%, which somehow occurred with the 75 C constraint rather than the 70 C constraint. Particularly when examining average operating frequency, as the temperature constraint decreases, the DVFS+TEC controller performs better and better than the DVFS-only controller. The improvement in runtime shows this trend to some extent, but is more randomly distributed. This is because two gcc runs may not be identical. The controller may behave slightly differently during each run. Additionally, since all the runs were performed successively, caching will have an impact. Previous runs will warm up the cache for subsequent runs. These problems can be fixed by flushing the cache in between runs, and by conducting multiple trials.

The DVFS-only controller must cut frequency to meet the temperature constraint, while the DVFS+TEC controller maintains a constant frequency of operation. However, the DVFS+TEC controller increases TEC power consumption, and does so more and more as the temperature constraint decreases, as seen in the "ptec/ptotal" row of the table. In terms of controller behavior, in all cases the temperature is in violation of the constraint for some percentage of the time. The temperature exceeds the constraint by a maximum of 9 C. While this appears excessive, the plots in Figure 31 and Figure 32 show that the temperature constraints are really only significantly violated for short periods of time.

The second experiment examines how varying the power budget affects CPU frequency and TEC power consumption. The assumption is that CPU frequency corresponds with performance and runtime, which may not always be accurate. This experiment runs two gcc applications at once rather than three, as in the previous experiment.

Figure 33 and Figure 34 show the controller behavior for Tmax = 45 C and power budgets of 40 W and 25 W. In the 25 W scenario, because of the strict power budget, the controller cannot increase TEC

current as much as it wants to meet the temperature constraint. The controller is forced to instead cut CPU frequency.

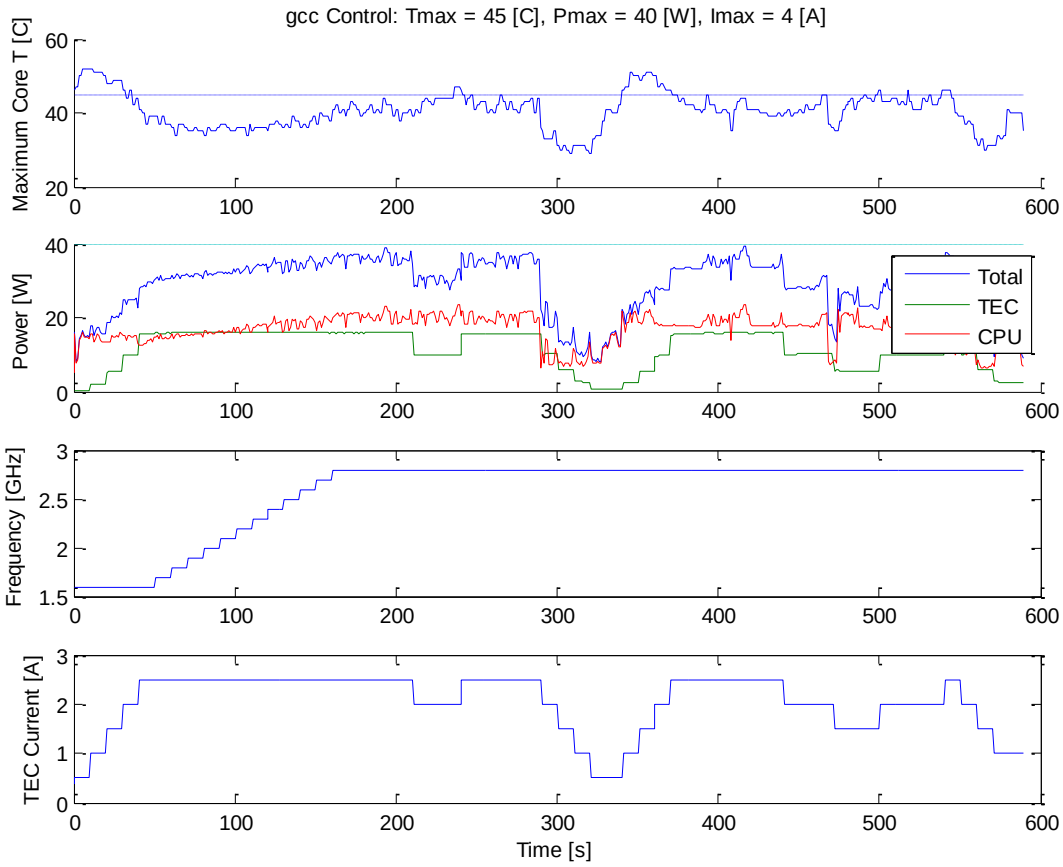

Figure 33 gcc Control with Tmax = 45 C and Pmax = 40 W. At some points, TEC power consumption almost exceeds the power consumption of the CPU. Note that TEC power consumption includes the power consumption of both TECs.

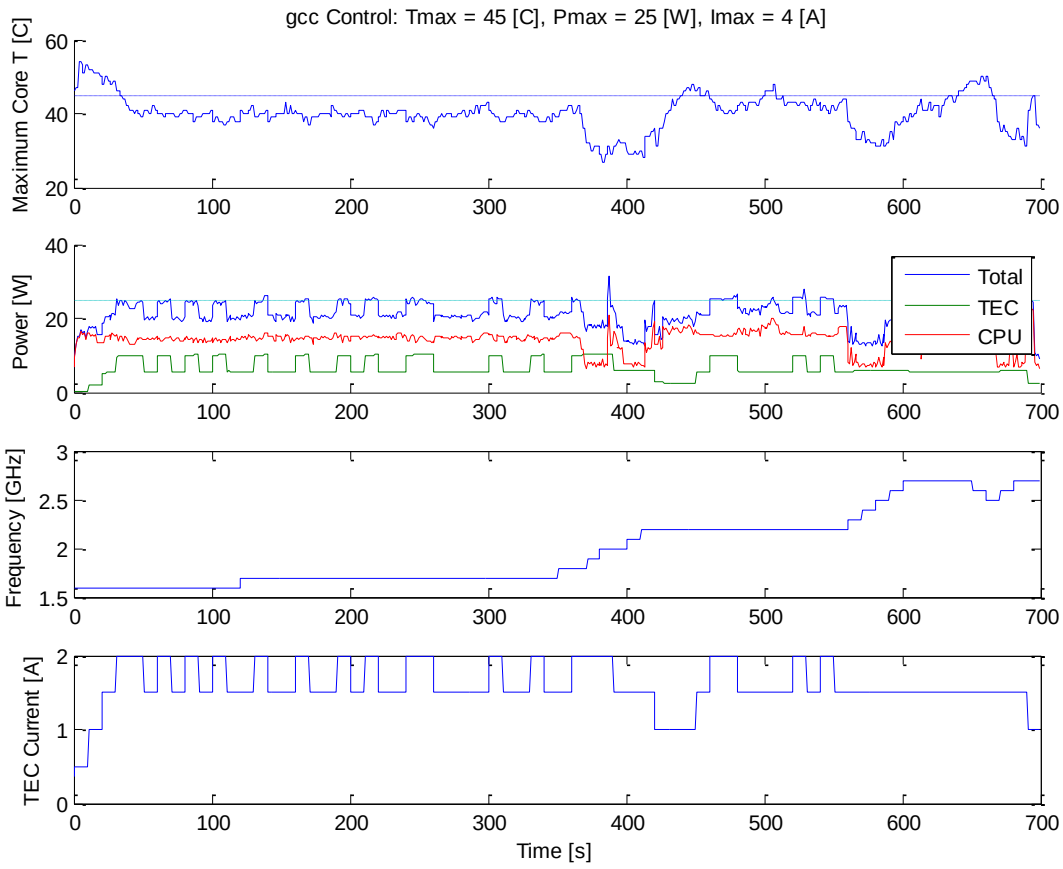

Figure 34 gcc Control at Tmax = 45 C and Pmax = 25 W. Compared to the 40 W power budget, the operating frequency is not 2.8 GHz most of the time.

Figure 35 shows how controller performance varies as a function of power budget, for fixed temperature constraints. Lower temperature constraints require higher power budgets since they need more TEC current to be supplied. This is why each of the curves starts at a different point. Additionally, for lower temperature constraints, the growth rate of frequency as a function of power budget is lower. Compare the Tmax = 45 C and Tmax = 60 C curves. At the lowest power budget available, the 45 C curve starts out growing slower compared to the 60 C curve.

Consider a fixed temperature constraint. At lower frequencies, increasing the power budget slightly gives larger gains in performance, compared to increasing the power budget by the same amount at a

higher frequency. This trend is exacerbated at lower temperature constraints, particularly at 45 C. This trend is expected. The temperature isotherms plot in Figure 28 on page 66 in the System Characterization section shows that, for a fixed temperature, achieving higher performance requires more and more TEC power. (In fact, this plot is very similar to the isotherms plot, except that the total power budget is plotted on the x-axis rather than TEC current).This shows that the trends in Figure 35 are not specific to the controller implementation, but are more general.

The underlying explanation for the trends is illustrated in Figure 36, which reproduces the temperature vs. TEC current curves investigated in the System Characterization section. As frequency changes, each of the curves is equally spaced. For gcc, as frequency increased, CPU power more or less grew linearly (see Figure 25 on page 61 in the System Characterization section). Thermal resistance K/W is constant for a fixed TEC current (see Figure 12 on page 39 in the Modeling chapter), and stays within some limited range over all TEC currents. Therefore, the spacing between the curves will be the change in CPU power times the thermal resistance. For each transition to a new frequency while maintaining a fixed temperature, incrementally more and more current is required. Remember that TEC power grows quadratically with current.

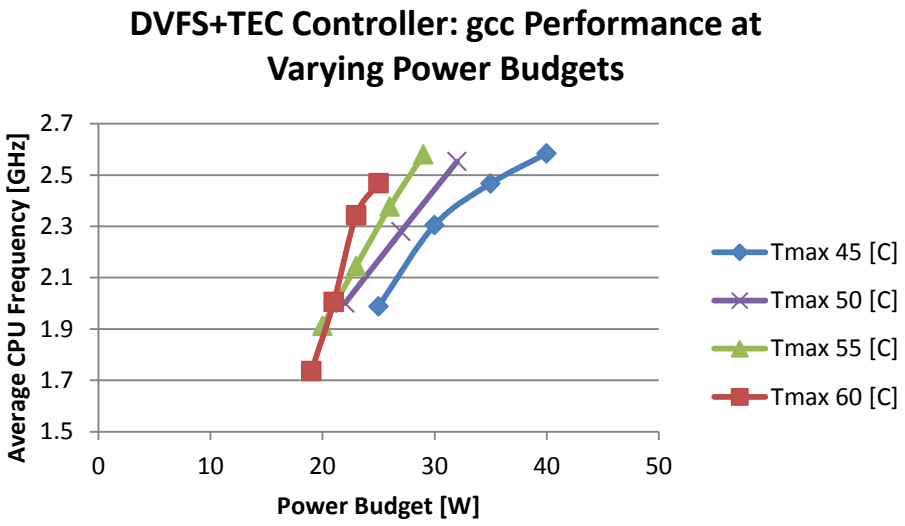

Figure 35 DVFS+TEC Controller: gcc Performance at Varying Power Budgets. Lower temperature constraints require a larger power budget. Second, as the power budget increases, to achieve a fixed increase in frequency requires more and more power. This trend is more visible for lower temperature constraints.

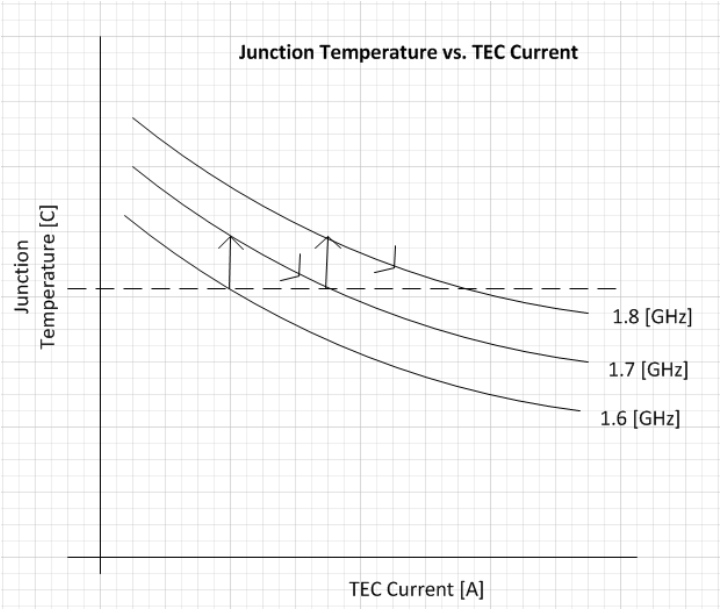

Figure 36 Junction Temperature vs. TEC Current. This plot shows frequency transitions while maintaining a fixed temperature constraint.

Figure 37 shows the proportion of total power taken up by TEC power consumption. As the power budget increases, the percentage of TEC power approaches some limit. This is because after a certain

power budget, no more increases in performance can be made, so the controller won't increase TEC current any further. Also, as Tmax decreases, the proportion of TEC power grows rapidly.

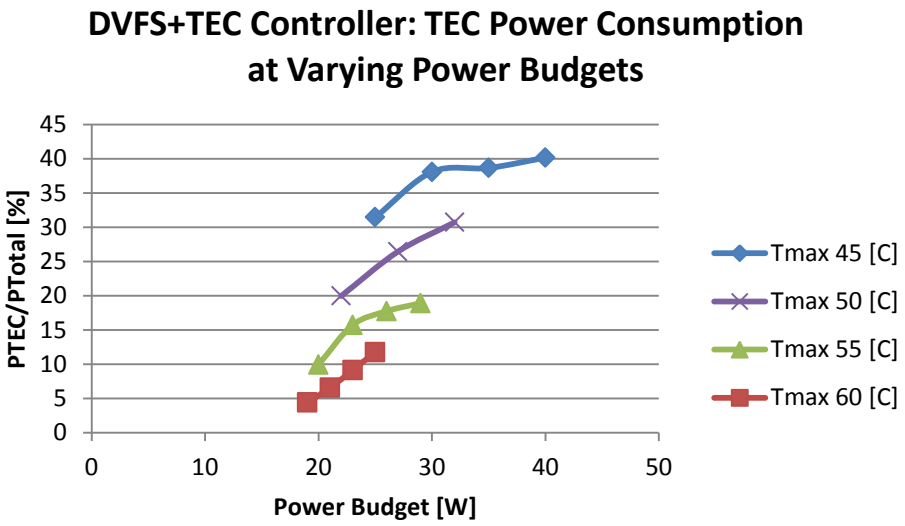

Figure 37 DVFS+TEC Controller: TEC Power Consumption at Varying Power Budgets and Temperature Constraints. As the power budget increases, the proportion of TEC power approaches some limit.

Showing the data in Figure 37 another way, the TEC coefficient-of-performance q_c/P_{TEC} is plotted as a function of power budget in Figure 38. In the TEC setup used in this thesis, the TEC is forced to pump all of the CPU power, so q_c is the CPU power. Therefore, the COP is P_{CPU}/P_{TEC}. q_c is not usually fixed when measuring the COP for a TEC, so this plot is not the standard plot of COP as a function of TEC current (see [21] for an example of the typical plot). Nevertheless, in this plot here, the TEC is more efficient at lower power budgets and higher temperature constraints. As the power budget increases, performance increases as the CPU runs at a higher frequency and the TEC current also increases to maintain the temperature constraint. So, while both CPU power and TEC power are increasing, it is clear that TEC power is increasing more quickly since the COP decreases as a function of power budget. For a fixed power budget and increasing temperature constraint, CPU frequency increases while TEC current decreases. Thus CPU power increases while TEC power decreases, so the COP rises.

For a given temperature constraint, the COP has a lower bound because after the power budget increases beyond a certain point, the CPU hits the maximum 2.8 GHz while meeting the temperature constraint, and there is no need for further increases in TEC power since there is no more performance to be had. Finally, if there were a temperature constraint and power budget at which the controller would run without the TECs on, the COP would be infinite.

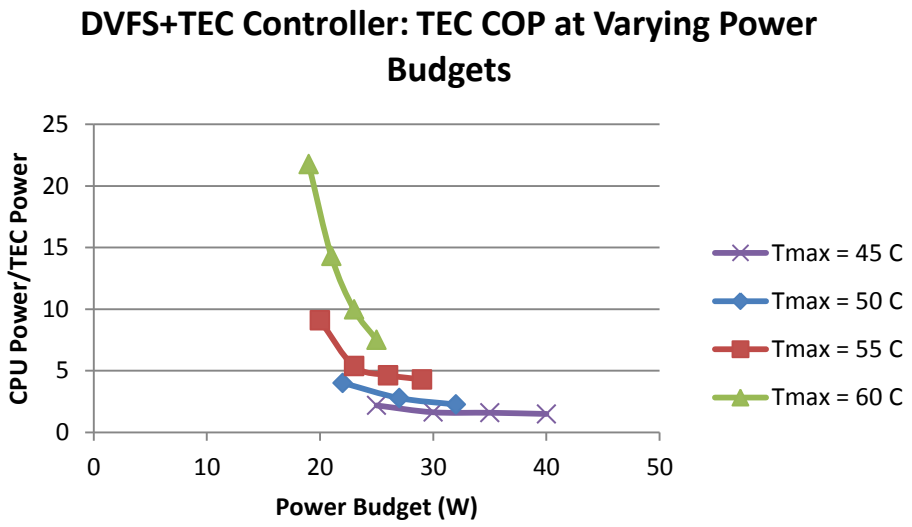

Figure 38 DVFS+TEC Controller: TEC COP at Varying Power Budgets and Temperature Constraints. Strictly speaking, the TEC COP is plotted. However, the COP of a TEC device is usually measured when the TEC hot-side temperature is fixed, rather than when fixing q_c. In these experiments, the TEC is forced to pump all of the CPU power, so q_c is fixed.

Table 9 shows the data used in Figure 35, Figure 37, and Figure 38.

	Power Budget (W)	Avg. f (GHz)	Run time (s)	Avg. TEC P (W)	Avg. CPU P (W)	Avg. Total P (W)	**Percent Increase in f**	**Percent Decrease in Runtime**	(TEC Power)/ (Total Power) %	Est. % Change in f, vs. DVFS-only on Std. System
Tmax = 45 (C)	25	2.0	699	6.6	14.4	21.0	--	--	31.5%	-13.0%
	30	2.3	640	9.7	15.8	25.4	15.9%	8.4%	38.1%	0.0%
	35	2.5	657	10.2	16.2	26.4	24.1%	6.0%	38.7%	8.7%
	40	2.6	589	11.3	16.9	28.2	30.0%	15.7%	40.2%	13.0%
Tmax = 50 (C)	22	2.0	767	3.6	14.5	18.2	--	--	20.0%	--
	27	2.3	616	5.8	16.3	22.1	13.9%	19.7%	26.4%	--
	32	2.6	546	7.9	17.8	25.6	27.6%	28.8%	30.7%	--
Tmax = 55 (C)	20	1.9	666	1.7	15.3	17.0	--	--	9.9%	--
	23	2.1	653	2.9	15.7	18.7	12.2%	2.0%	15.7%	--
	26	2.4	585	3.7	17.2	20.9	24.3%	12.2%	17.7%	--
	29	2.6	564	4.1	17.7	21.9	35.0%	15.3%	18.9%	--
Tmax = 60 (C)	19	1.7	661	0.7	15.4	16.2	--	--	4.4%	--
	21	2.0	656	1.1	16.0	17.1	15.5%	0.8%	6.5%	--
	23	2.3	588	1.7	17.4	19.1	35.0%	11.0%	9.1%	--
	25	2.5	595	2.4	18.0	20.4	42.2%	10.0%	11.8%	--

Table 9 DVFS+TEC Controller Behavior at Varying Power Budgets. For this data, 2 gcc applications were run simultaneously to completion. The final column approximates a comparison to the DVFS-only controller running on the standard system. Please see the discussion in the text on why this is only an estimate.

In the System Characterization section (page 53), two gcc applications running for 120 s on the standard system hit a maximum temperature of 45 C when running at a fixed 2.3 GHz. gcc hits its highest temperatures within the first 120 s of runtime. Thus, a rough comparison between (a) the DVFS+TEC controller running on the TEC-on-spreader system, and (b) the DVFS-only controller running on the standard system, can be made. Any controller run that has an average CPU frequency of 2.3 GHz and temperature constraint of 45 C can substitute for the DVFS-only controller. This substitution depends on the controlled run with an average 2.3 GHz CPU frequency being equivalent to a run with fixed 2.3 GHz frequency. This is obviously inaccurate. Additionally, the initial notion that the fixed 2.3 GHz run accurately represents the DVFS-only controller meeting a 45 C temperature constraint on the standard system is flimsy. It is particularly shaky for a benchmark like gcc, which consumes variable amounts of

power over time. The DVFS-only controller would increase frequency during times of low power consumption, rather than keeping frequency fixed. As a side note, two applications of gcc on the standard system don't go above 50 C in terms of maximum temperature, so no comparison is possible beyond the 45 C point.

Table 9 shows that for a 45 C temperature constraint, if the average CPU frequency is 2.3 GHz, then the average CPU power consumption is 15.8 W, and the runtime is 640 s. (Using another 2.3 GHz point at a different temperature constraint won't do because leakage power increases with temperature). Since the average CPU frequency is 2.3 GHz, this data approximates how the DVFS-only controller would behave on the standard system while meeting a temperature constraint of 45 C. Since the standard system (which has no TECs) is being approximated, the TEC power consumption becomes irrelevant. The standard system would only consume power via the CPU. Thus, the DVFS-only total energy consumption, using just the average CPU power, would be 10100 J. At a 40 W power budget, the DVFS+TEC controller on the TEC-on-spreader system consumes 28.2 W of average total (TEC and CPU) power over a runtime of 589 s. The energy consumption is 16600 J. Overall, comparing the DVFS-only and DVFS+TEC controllers, with the DVFS+TEC controller runtime decreases by 8.0%, while energy consumption increases by 64.3%.

5. Conclusion

Summary

This thesis investigated a real TEC system that can cool a microprocessor, and then developed a controller to thermally manage the system with the TECs. When running two applications of gcc simultaneously with a 40 W power budget and a 45 C temperature constraint, the thesis estimated that the DVFS+TEC controller provides an 8% decrease in runtime compared to a DVFS-only controller, at the cost of a 64% increase in energy consumption. Additionally, the transient behavior of the system as the DVFS+TEC controller ran was investigated. For the scenario just described, the maximum overshoot above the temperature constraint is 7 C. The temperature constraint is violated 13.8% of the time. In terms of response time, temperature reaches a steady state within 200 seconds.

At the beginning, the *Modeling* chapter presented the basic TEC system setup with a real quad-core processor. Two TECs are placed on top of the integrated heat spreader. The fan and heat sink go on top of the pair of TECs. The alternative designs of (1) placing copper in parallel with the TECs and, (2) putting the TECs directly on die, were deemed less effective in minimizing junction temperature. The inefficacy of (2) is not a general result.

The steady state model developed for the TEC system gives an estimate of junction temperatures, as a function of CPU power and TEC current. Comparing the steady state model to the experimental data, the model performs well if the correct model parameters are chosen. The model parameters may have to be determined empirically, instead of by purely analytical methods. Nevertheless, the model predicts the optimal TEC current which minimizes junction temperature. Also, in terms of predicting junction temperature over the entire range of TEC currents, the mean error for the model compared to the actual data is less than 3 C.

In the *Dynamic Thermal Management* chapter, the system characterization compared a standard system with the TEC system. To characterize the TEC system, two SPEC2006 benchmarks (povray and gcc) were run at various processor frequency settings and TEC currents. The TEC system has a thermal resistance of 1.6 K/W when the TECs are off, while the standard system has a thermal resistance of 0.4 K/W. While this seems a daunting obstacle for the TEC system to overcome, if the TECs are engaged at sufficient current levels, the TEC system can cool to temperatures below those of the standard system. This is true only up to maybe 30 W of CPU power consumption. This is the primary limitation of the TEC system. Realistically, at 30 W CPU power, junction temperatures are not high enough to justify needing the extra cooling the TECs provide.

In any case, the cost of this cooling is TEC power consumption, but the major benefit, as the controller demonstrates, is performance. This thesis concluded that, for this particular setup, while the TECs can cool below standard system levels and provide additional performance, they do not do so efficiently. For example, running two gcc applications, the standard system can run at 1.8 GHz and 41 C junction temperature. The TEC system can do the same at 2.6 GHz and 3 A of TEC current distributed to each of the TECs. At this 2.6 GHz and 3 A setting, the TECs consume more than 20 W of power by themselves, which is greater than what the CPU consumes. However, running at 2.6 GHz also improves performance significantly.

The system characterization also determined equivalent {DVFS setting, TEC current} pairs that all maintain the same maximum junction temperature. Lower operating temperatures require more TEC power consumption. Also, as DVFS setting and performance increase, the rate at which TEC power consumption grows increases. These observations were confirmed with the controller when the power budget and temperature constraints were varied.

Finally, the controller section contrasted a DVFS-only controller with a DVFS+TEC controller. With an unlimited power budget, as the temperature constraint decreases, the DVFS+TEC controller maintains performance while the DVFS-only controller is forced to cut frequency to meet the temperature threshold.

Directions for Further Work

The downsides of the TEC system are that it can only cool up to a certain CPU power limit, and that its power consumption is relatively high. Therefore, one line for new work is to improve on system designs, particularly by considering thin-film thermoelectrics. A second path would be to develop a transient model of the TEC system and fine-tune the controller using it. As is, all of the modeling work done in this thesis is based on a static TEC model. For example, the controller can be improved to work optimally over all possible patterns of CPU power consumption, or to have a better response to step changes in CPU power consumption.

6. Appendices

Data Reliability

This study considers two SPEC2006 benchmarks: gcc and povray. Each benchmark is run over CPU core frequencies ranging from 1.6 to 2.8 GHz (in 0.1 GHz increments) and TEC currents range from 0 to 5 A (in 1 A increments). (The TEC current refers to the current that each TEC receives. There are 2 TECs in the system.) There are 13 frequencies, and 6 current levels. A single trial for a benchmark is defined as running that benchmark 78 times, each time at a different (core frequency, TEC current) pair. A run is defined as a benchmark run at a particular core frequency and TEC current. In tabular form, a single trial looks like the following. Each run is a single row in the table.

Frequency (GHz)	TEC Current (A)	Data
1.6	0	...
...	0	...
2.8	0	...
1.6	1	...
...	1	...
2.8	1	...
...	...	...

Two trials are run for each benchmark, to account for random variation; any analysis uses the average of the two trials. Nevertheless, there should be minimal variation between trials because running benchmarks is a mostly deterministic process. The benchmark, frequency, and TEC current should uniquely determine the CPU temperature. The experiments control for factors like CPU fan speed and thread affinities.[5] CPU fan speed is fixed at 2000 rpm, since the fan has an attached potentiometer to control rpm. Two instances of the benchmark, say gcc, are run simultaneously. Instance 1 is fixed to run

[5] A potentiometer controls the fan speed. For affinities, see Linux's sched_setaffinity and sched_getaffinity libc system call wrapper.

on core 0, instance 2 on core 3. While it is difficult to control for variations in ambient temperature, and the behavior of the operating system and caches, these are minor effects.

The best measurement for comparing equivalent runs in two separate trials is the average temperature, which is averaged both over time and over all four cores. All aspects of the system, including TECs, CPU power, and fan, affect the average temperature, so it is useful measurement for detecting anomalies between trials.

Figure 39 and Figure 40 show the distribution of mean deviation and relative mean deviation (the mean deviation expressed as a percentage of the mean) between trials, by benchmark. For each of the 78 pairs of equivalent runs in the two trials, values are plotted in the histograms. For example, suppose the average temperature of gcc at 1.6 GHz and 1 A during trial 1 is 32 C, and during trial 2 it is 33 C. The mean deviation is 0.5 C, and the relative mean deviation is 100*0.5 C/32.5 C, or 1.5%. The relative mean deviation is more revealing because a fixed absolute deviation is worse at lower temperatures than at higher temperatures. At 4 A TEC current, when temperatures are low and around 20 C, a 1 C deviation means more than at 0 A TEC current, when temperatures may be around 50 C. Since these experiments do not involve many random processes, the histograms for gcc are shown below as an example. The histograms for povray and the standard system data are not shown. The standard system is defined as the system without the TECs.

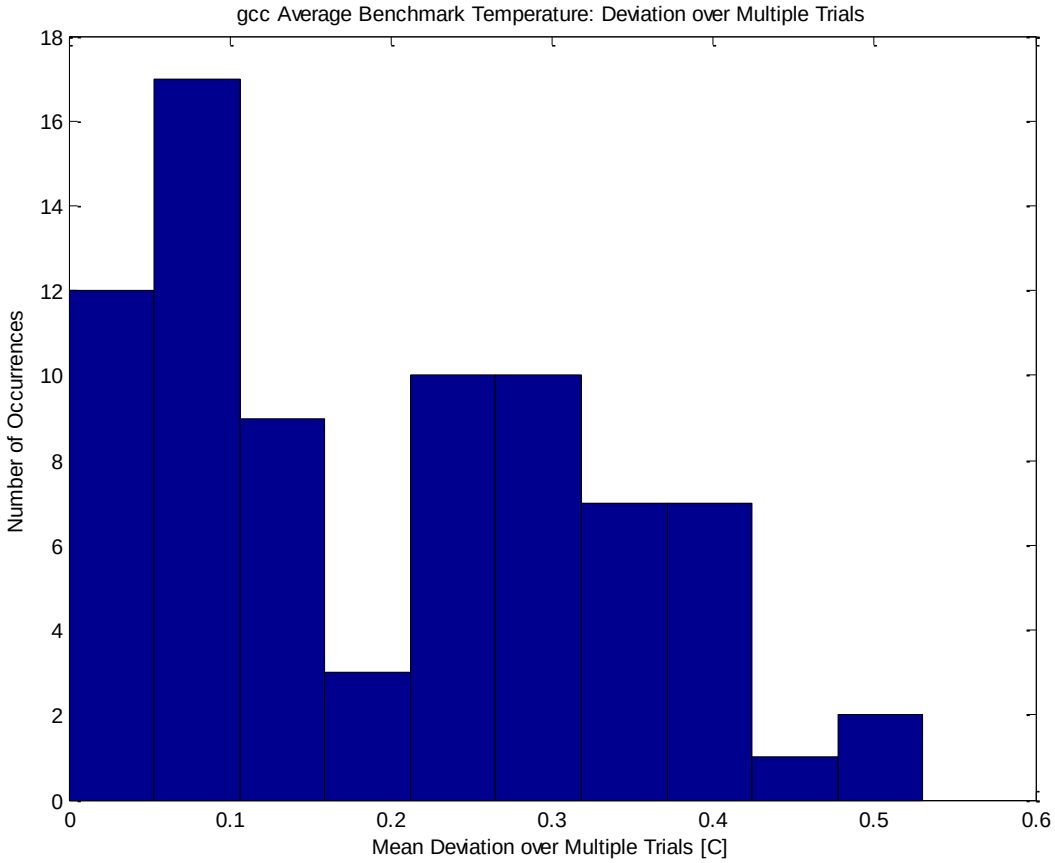

Figure 39 gcc Deviation over Trials

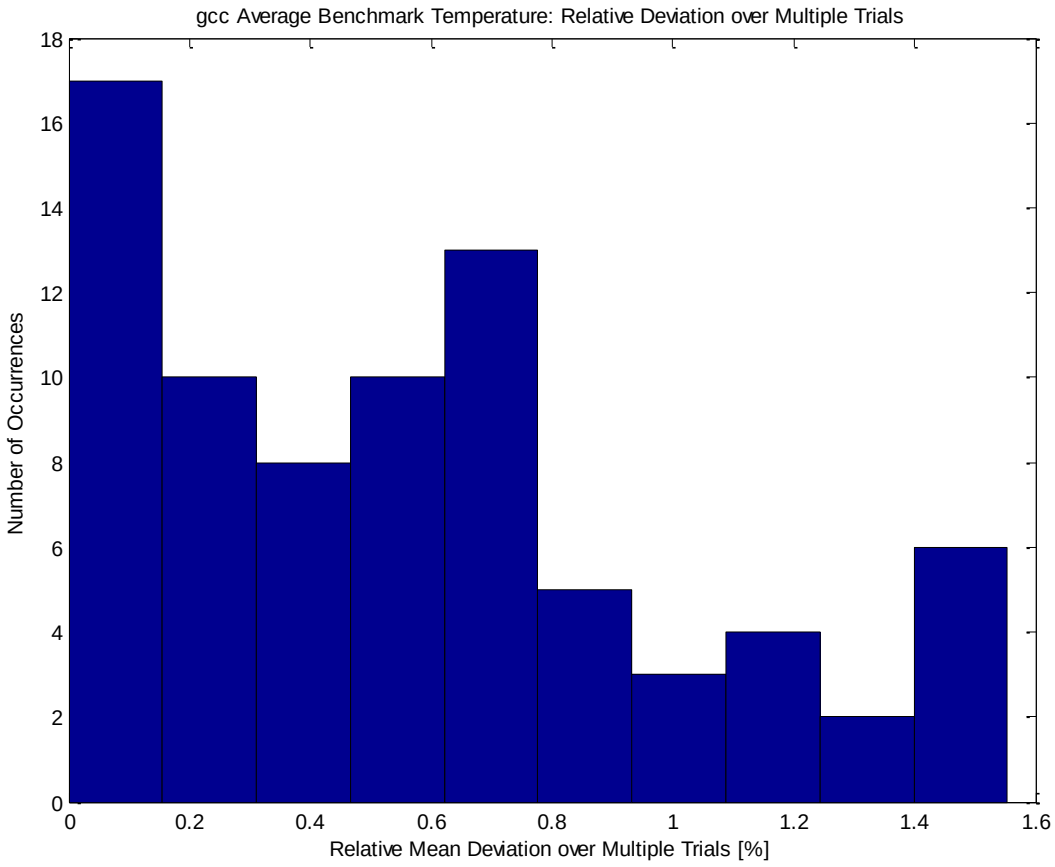

Figure 40 gcc Relative Deviation over Trials

Note that each plot contains deviations for runs at all 13 frequencies and all 6 current steps. The sum of the occurrences in each plot is thus 78. The deviation may be a function of frequency and current. For instance, runs at 2.8 GHz might intrinsically have more deviation across trials than runs at 1.6 GHz. The plots above don't reveal such patterns. They do show that, regardless of frequency, current setting, and benchmark, relative mean deviation stays below 2%. The distribution is skewed towards lower percentages. The majority of deviations are below 1%. This level of deviation is acceptable.

System Pictures

Figure 41 Fan and Motherboard

Figure 42 Power Supply and Multimeter. The bottom photo shows one of the power supplies. The top photo shows the digital multimeter, with the shunt resistor plugged in.

Controller Psuedocode

```
if(t < tmax && p < pmax) {
      /*
      Since neither constraint is exceeded, it is safe to  increase
      frequency if possible. Otherwise, decrease TEC current to save
      power.
      */
      if(t < tmax-t_hyst && p < pmax-p_hyst) {
            if(f < fmax)
                  f++
            else if(i > imin)
                  i--
      }

      /*
      In this case, it isn't safe to increase frequency.  However, if
      the maximum frequency hasn't been reached yet, and there is power
      budget available, try to increase the TEC current to cool the
      processor further. Subsequently, a frequency increase may be
      possible.
      */
      if(t > tmax-t_hyst && p < pmax-p_hyst) {
            if(i < imax && f < fmax)
                  i++
      }

      /*
      Here, decrease TEC current to save power.
      */
      if(t < tmax-t_hyst && p > pmax-p_hyst) {
            if(i > imin)
                  i--
      }

      if(t > tmax-t_hyst && p > pmax-p_hyst) {
            //do nothing
      }
}
```

```
/*
Since the power budget is exceeded but the temperature constraint is
not, try decreasing TEC current first while preserving performance.
*/
if(t < tmax && p > pmax) {
      if(i > imin) {
            i--
      } else if(f > fmin) {
            f--
      } else {
            error(p constraint failed)
      }
}

/*
To decrease temperature, first try increasing current to avoid hurting
performance.
*/
if (t > tmax && p < pmax) {
      if(i < imax && p < pmax-p_hyst) {
            i++
      } else if(f > fmin) {
            f--
      } else {
            error(t constraint failed)
      }
}

/*
If both constraints are violated, try to alleviate both problems by
decreasing frequency. Temperature is considered to have priority over
power, so TEC current is not decreased, since that might increase
temperature.
*/
if (t > tmax && p > pmax) {
      if(f > fmin) {
            f--
      } else {
            error(t and p constraint failed)
      }
}
```

Bibliography

[1] "Radioisotope Thermoelectric Generator," *NASA*, 13-Dec-2012. [Online]. Available: http://solarsystem.nasa.gov/rps/rtg.cfm. [Accessed: 20-Mar-2013].

[2] "Advanced Stirling Radioisotope Generator," *NASA*, 21-Feb-2013. [Online]. Available: http://solarsystem.nasa.gov/rps/asrg.cfm. [Accessed: 20-Mar-2013].

[3] F. J. DiSalvo, "Thermoelectric Cooling and Power Generation," *Science*, vol. 285, no. 5428, pp. 703–706, Jul. 1999.

[4] "FAQ & Technical Information," *TE Technology Inc.*, 26-Jul-2007. .

[5] M. Brain and S. Elliott, "How Refrigerators Work," *HowStuffWorks*, 29-Nov-2006. [Online]. Available: http://home.howstuffworks.com/refrigerator.htm. [Accessed: 27-Feb-2013].

[6] "refrigeration," *Encyclopedia Britannica Online*. .

[7] *1MDL-06-052-XX*. .

[8] M. J. Moran and H. N. Shapiro, *Fundamentals of engineering thermodynamics*. Hoboken, N.J: Wiley, 2006.

[9] M. Kutz, *Temperature control*. Wiley, 1968.

[10] W. E. Boyce, *Elementary differential equations and boundary value problems*, 9th ed. Hoboken, NJ: Wiley, 2009.

[11] C. J. Glassbrenner and G. A. Slack, "Thermal conductivity of silicon and germanium from 3 K to the melting point," *Physical review*, vol. 134, no. 4a, p. A1058, May 1964.

[12] *Materials in Design Engineering*, vol. Materials Selector. Reinhold Publishing Corporation, 1966.

[13] "Desktop 3rd Generation Intel® Core™ Processor Family, Desktop Intel® Pentium® Processor Family, Desktop Intel® Celeron® Processor Family, and LGA1155 Socket Thermal Mechanical Specifications and Design Guidelines (TMSDG)," Intel, Jan. 2013.

[14] A. Wong, "Heatsink Thermal Compounds Review," *SLCentral*, 05-Dec-2003. [Online]. Available: http://www.slcentral.com/thermal-compound/index.php. [Accessed: 24-Mar-2013].

[15] R. C. Dorf and R. H. Bishop, *Modern Control Systems*, 10th ed. Prentice Hall, 2004.

[16] J. F. Shackelford and W. Alexander, *CRC Materials Science and Engineering Handbook, Third Edition*. CRC Press, 2010.

[17] Y. I. Shtern, A. S. Malkova, A. S. Pashinkin, and V. A. Fedorov, "Heat capacity of the n-Bi2Te2.88Se0.12 and p-Bi0.52Sb1.48Te3 solid solutions," *Inorg Mater*, vol. 44, no. 10, pp. 1057–1059, Oct. 2008.

[18] M. Wagner, "Simulation of Thermoelectric Devices," TU Wien, Nov. 2007.

[19] P. E. Gray, *The dynamic behavior of thermoelectric devices*. Technology Press of the Massachusetts Institute of Technology, 1960.

[20] Z. Luo, "A Simple Method To Estimate The Physical Characteristics Of A Thermoelectric Cooler From Vendor Datasheets," *Electronics Cooling*, 01-Aug-2008. [Online]. Available: http://www.electronics-cooling.com/2008/08/a-simple-method-to-estimate-the-physical-characteristics-of-a-thermoelectric-cooler-from-vendor-datasheets/. [Accessed: 07-Apr-2013].

[21] G. J. Snyder, M. Soto, R. Alley, D. Koester, and B. Conner, "Hot spot cooling using embedded thermoelectric coolers," in *2006 IEEE Twenty-Second Annual IEEE Semiconductor Thermal Measurement and Management Symposium*, 2006, pp. 135–143.

[22] I. Chowdhury, R. Prasher, K. Lofgreen, G. Chrysler, S. Narasimhan, R. Mahajan, D. Koester, R. Alley, and R. Venkatasubramanian, "On-chip cooling by superlattice-based thin-film thermoelectrics," *Nat Nano*, vol. 4, no. 4, pp. 235–238, Apr. 2009.

[23] H. Y. Zhang, Y. C. Mui, and M. Tarin, "Analysis of thermoelectric cooler performance for high power electronic packages," *Applied Thermal Engineering*, vol. 30, no. 6–7, pp. 561–568, May 2010.

[24] G. L. Solbrekken, K. Yazawa, and A. Bar-Cohen, "Heat driven cooling of portable electronics using thermoelectric technology," *IEEE Transactions on Advanced Packaging*, vol. 31, no. 2, pp. 429–437, 2008.

[25] R. Alley, M. Soto, L. Kwark, P. Crocco, and D. Koester, "Modeling and Validation of On-Die Cooling of Dual-Core CPU using Embedded Thermoelectric Devices," in *Twenty-fourth Annual IEEE Semiconductor Thermal Measurement and Management Symposium, 2008. Semi-Therm 2008*, 2008, pp. 77–82.

[26] B. J. Huang and C. L. Duang, "System dynamic model and temperature control of a thermoelectric cooler," *International Journal of Refrigeration*, vol. 23, no. 3, pp. 197–207, May 2000.

[27] P. Chaparro, J. González, Q. Cai, and G. Chrysler, "Dynamic thermal management using thin-film thermoelectric cooling," in *Proceedings of the 14th ACM/IEEE international symposium on Low power electronics and design*, New York, NY, USA, 2009, pp. 111–116.

[28] S. Reda and F. Paterna, "Mitigating Dark Silicon Problems Using Superlattice-based Thermoelectric Coolers," *Design, Automation, and Test in Europe 2013*, 2013.

[29] S. Biswas, M. Tiwari, T. Sherwood, L. Theogarajan, and F. T. Chong, "Fighting fire with fire: Modeling the datacenter-scale effects of targeted superlattice thermal management," in *2011 38th Annual International Symposium on Computer Architecture (ISCA)*, 2011, pp. 331–340.

[30] "Desktop 3rd Generation Intel® Core™ Processor Family, Desktop Intel® Pentium® Processor Family, and Desktop Intel® Celeron® Processor Family," Intel, Jan. 2013.
[31] *Intel Ivy Bridge Die Map*. 2012.
[32] "'Making of a Chip' Illustrations 22nm 3D/Trigate Transistors – Version."Intel, Jan-2012.
[33] "ASRock Z77 Extreme4 User Manual." Jul-2012.
[34] N. Weste and D. Harris, *CMOS VLSI Design: A Circuits and Systems Perspective*, 4th ed. Addison-Wesley, 2010.