Platform-Dependent, Leakage-Aware Control of the Driving Current of Embedded Thermoelectric Coolers

Mohammad Javad Dousti and Massoud Pedram
University of Southern California
Los Angeles, CA, USA
{dousti, pedram}@usc.edu

Abstract

One of the biggest stumbling blocks for the successful continuation of the Moore's law is the substrate temperature of VLSI circuits. Thermoelectric cooling is one of the promising cooling methods to combat high die temperatures. This method provides key benefits such as compactness, high reliability, and exceptionally high heat-pumping capability. On the other hand, even with the recent advances in the fabrication techniques, thermoelectric coolers (TECs) are suffering from a poor coefficient of performance (COP), which denotes the ratio of heat removed per second to the power needed to drive the TEC, is rather low. In this paper, different techniques to improve the performance of a TEC, when it is embedded inside a processor package, are investigated. In particular, first the COP of TECs is reformulated to consider the leakage power, which is exponentially dependent on the die temperature. Next it is demonstrated that the TEC driving current that yields the maximum decrease in the die temperature is quite different from the one that runs the TEC in its highest COP state. Based on these observations, a platform-dependent, leakage-aware cooling policy in which the TEC driving current is set based on the target specs (high-performance vs. low-power) and actual conditions of the chip (emergency vs. preventive thermal management) is proposed. Experimental results show that, with this policy, one can reduce the temperature of chip hotspots while achieving a high COP.

Keywords

Thermoelectric coolers, TEC, thermal management, platform-dependent policy, low-power.

1. Introduction

One of the key challenges in the successful continuation of the Moore's law is that of overcoming thermal issues on the chip. Temperature has always been a big barrier to increasing performance. Elevated die temperatures result in higher leakage power dissipation, faster device aging, and can even cause permanent damage to a chip. The purpose of a thermal management system is to stop the temperature increase beyond a certain threshold, even if the required action is to power off the chip.

Different thermal management solutions have been proposed during the past decade. These solutions tend to negatively impact the performance. One solution which does not degrade the performance is to use more advanced cooling material and techniques. Cooling solutions are generally classified as *passive* and *active cooling* [1]. Passive cooling techniques have no moving parts and do not need any power source to operate, whereas active cooling methods exploit moving parts and/or require an external power source. Reference [1], which provides an extensive evaluation of many passive and active methods, points out that a common disadvantage of various techniques is their low heat-pumping capability. In particular, none of the traditional techniques has the ability to pump heat fluxes higher than 1,000 W/cm^2. Note also that active methods suffer from reliability issues and some of them (like the direct jet impingement method) cannot be incorporated inside the chip package because of their large size. After enumerating different active and passive cooling solutions, reference [1] introduces a new active cooling method called *thermoelectric cooling* that has recently caught attention especially for the processor chip cooling. *Thermoelectric coolers* (TECs) are active devices that work based on the *Peltier effect*. This effect allows the device to absorb the heat from one side and release it on the other side when electrical current passes through it. The amount of cooling is linearly proportional to the amount of passing current. Notable features of TECs are as follows: 1) Compact size: TECs can be built as thin as tens of micrometers and their area can be smaller than 1 mm^2. These devices have the right size to exclusively cover typical hot spots on a chip. 2) Fast response time: Thin-film TECs have very fast response times in the order of a few milliseconds. 3) High reliability: These devices have no moving parts, and hence, can last longer than other active cooling solutions. Commercial TECs are expected to work for more than 11 years [2]. 4) High controllability: TECs can be controlled at the granularity of fractions of a degree of Celsius and can cool down a chip below the ambient temperature. 5) High heat pumping rate: It has been shown that thin-film TECs can pump high heat fluxes as large as ~1,300 W/cm^2 [3].

The unique features of TECs make them a perfect candidate for cooling. Unfortunately, Joule heating occurs as an adverse phenomenon during the cooling process by TECs, which causes the device to dissipate heat when current flows through it. As a result, a major drawback of TECs is their rather poor *coefficient of performance* (COP), which is defined as the ratio of heat removed in a unit of time to the total power used to drive the TEC. Many studies have been focused on adapting TECs for microprocessor cooling. Reference [4] suggests to counter the low-COP problem of TECs by limiting their use to the chip hotspots. Accordingly, very few TECs are selectively deployed on the chip surface. Although this recommendation has been widely accepted, it has two shortcomings:

1) It limits the usage of TECs to high-performance applications since low-power applications remain sensitive to low COP values of even a small number of deployed TECs.

2) Recent state-of-the-art multi-core chips have dozens of hot spots, which demand aggressive deployment of TECs. Again the low COP value of TECs poses a serious problem.

In this paper, we take on the challenge of improving the COP of TECs incorporated in a processor package. In particular, first we redefine the COP in order to capture the effect of chip leakage power, which is exponentially dependent on temperature. Using this new definition, we show that the COP of a cooling system (comprised of the chip, the TEC element, and a heat sink) versus the TEC driving current changes so as to exhibit a peak value for a driving current level based on the thermal chip condition. This is in clear contrast to the traditional COP vs. current curve (i.e., when excluding the leakage power consumption from consideration), which shows a constant peak value irrespective of chip condition. In particular, we show that TECs can increase the COP of a cooling system by 7% while decreasing the temperature by 6℃. Using these observations, we present a platform-dependent, leakage-aware policy to apply an appropriate current level to the TECs based on the target platform/application (high-performance vs. low-power) and the actual conditions of the chip (emergency vs. preventive thermal management.)

The rest of this paper is organized as follows. Section 2 explains the principles of thermoelectric cooling. Section 3 provides a review of the previous work. Section 4 introduces a new formulation for COP to account for the leakage power dissipation. Section 5 presents the platform-dependent, leakage-aware policy for setting the current of TECs. Section 6 presents the experimental results performed by a TEC simulator (called *Teculator*) which is developed based on the suggested new COP formulation. Section 7 concludes the paper.

2. Background

In this section, first key principles of thermoelectric cooling are reviewed. Equations are well known in the field of thermodynamics. Interested readers may refer to reference [5] for detailed discussions. Next, the assembly of TEC modules inside a microprocessor cooling package is explained. This assembly is used throughout the paper.

2.1. Principles of TEC Operation

Thermoelectric coolers are compact devices which are made of pairs of N-type and P-type semiconductor pellets. When current flows through a P-type pellet (from the positive terminal to the negative terminal), heat flows in the same direction, i.e., heat is absorbed from the positive side, which is called *cold side*, and released to the negative side, which is called *hot side*. The heat flow direction in an N-type pellet is the reverse of that of the P-type pellet. Usually several N-P pairs are connected electrically in series and thermally in parallel to increase the amount of heat rejection. Figure 1 shows a 3×3 array of TECs (a total of 9 N-P pairs).
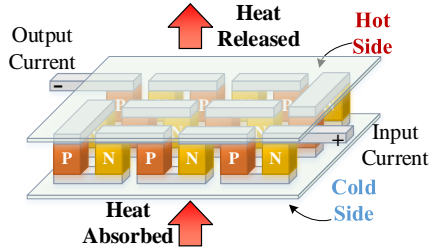

Figure 1: A 3x3 array of TECs.

The heat absorbed per unit time from the cold side is denoted by q_c and calculated as

$$q_c = N\left(\alpha T_c I - K_{TEC}\Delta T - \frac{1}{2}R_{TEC}I^2\right),\qquad(1)$$

where N is the number of TECs connected electrically in series, α is the Seebeck coefficient, T_c is the temperature of the cold side (in Kelvin), K_{TEC} is the thermal conductance of the TEC, ΔT is the temperature difference between the hot side and the cold side $(T_h - T_c)$, R_{TEC} is the electrical resistance of the TEC, and I is the current which flows through the TEC. The first term in this equation captures the *Peltier effect* which is the cooling phenomenon, the second term signifies the heat conductivity of the material from hot side to the cold side, and the third term is the Joule heating effect. Note that the second and the third terms have adverse effects in the cooling applications and hence have a negative sign. Moreover, the $\frac{1}{2}$ coefficient for the Joule heating is added because it is approximated that half of the Joule heating is released in the cold side and the other half is released in the hot side. Also note that the Joule heating quadratically depends on the current whereas the Peltier effect linearly depends on it. Similarly, the heat released per unit time to the hot side is denoted by q_h and can be written as

$$q_h = N\left(\alpha T_h I - K_{TEC}\Delta T + \frac{1}{2}R_{TEC}I^2\right),\qquad(2)$$

where T_h denotes the temperature in the hot side. In equations (1) and (2), the *Thomson effect* is not considered because of its negligible effect. Figure 2 shows how the current flows through a TEC N-P pair. The pink dashed arrow shows the direction of the current flow.
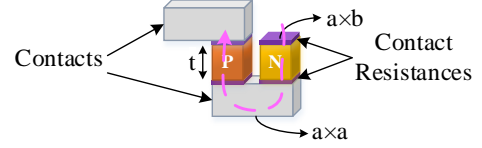

Figure 2: A TEC N-P pair. The aspect ratio of elements is not accurate and sizes are exaggerated. The dashed arrow shows the direction of current flow.

The contact resistance between the pellets and the metal contact increases the TEC resistance. If the contact resistivity is assumed as ρ_{contact} (with the unit of $\Omega.m^2$), the resistance caused by these contacts can be calculated as shown in Equation (3). Note that the factor of 4 in this Equation is added to account for the four contacts that each pair of pellets has with metals.

$$R_{contacts} = 4\rho_{\text{contact}}\frac{1}{a\times b}\qquad(3)$$

Using Equation (3), the total resistance of a TEC pellet (R_{TEC}) can be calculated as

$$R_{TEC} = R_{contacts} + 2\rho_{TEC}\frac{t}{ab},\qquad(4)$$

where ρ_{TEC} is the average electrical resistance of N and P-pellets (i.e., $\rho_{TEC} = (\rho_N + \rho_P)/2$) and t is the thickness of the TEC. The coefficient 2 is added to the second term in order to account for both pellets. The cooling performance of a TEC is linearly proportional to α and inverse proportional to K_{TEC} and R_{TEC}. Hence a natural way of defining *figure of merit* (Z) for a TEC device is

$$Z = \frac{\alpha^2}{R_{TEC}K_{TEC}} = \frac{\alpha^2}{\rho_{TEC}k_{TEC}}.\qquad(5)$$

The simplification in Equation (5) is done using the relations $R_{TEC} = \rho_{TEC}\frac{t}{ab}$ and $K_{TEC} = k_{TEC}\frac{ab}{t}$. Note that in the second relation, capital K_{TEC} is thermal conductance whereas small k_{TEC} is the thermal conductivity. Figure of merit is defined in such a way to be independent of TEC geometry and its input current. In order to make it a dimensionless quantity, ZT_{avg} is usually used. T_{avg} is the average temperature between the hot and the cold side temperatures of a TEC. ZT_{avg} value for the state-of-the-art TECs are as high as 2.1 in 300K [3].

Power consumption of N TECs is the difference between q_h and q_c and may be written as follows:

$$P = q_h - q_c = N(\alpha\Delta TI + R_{TEC}I^2).\qquad(6)$$

A useful metric is the *coefficient of performance* which we call it COP^{basic}. This metric is traditionally defined as the ratio of the rejected heat (q_c) per unit time and the input power to the TEC (P):

$$COP^{basic} = \frac{q_c}{P} = \frac{\alpha T_c I - K_{TEC}\Delta T - \frac{1}{2}R_{TEC}I^2}{\alpha\Delta TI + R_{TEC}I^2}\qquad(7)$$

2.2. TEC Assembly

Figure 3 shows a typical cooling package assembly of a microprocessor in which TEC modules are incorporated. As can be seen, TECs are immersed inside the thermal interface material (TIM) for better heat conductivity between the chip and TECs as well as between TECs and the heat spreader. The heat spreader is also connected to the heat sink through another layer of TIM.
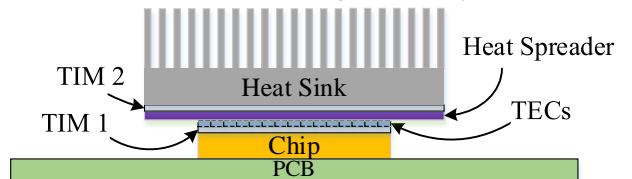

Figure 3: A chip assembly with its cooling solution

Using the duality between thermal and electrical phenomena, an electrical circuit equivalent to a thermal system can be built. This duality is summarized in Table 1. An electrical system is handy as it can be easily analyzed using well-known circuit laws (such as KVL and KCL) and it can also be simulated using circuit simulators such as SPICE.

Thermal Quantity	Unit	Dual Electrical Quantity	Unit
Temperature (T)	K	Voltage (V)	V
Power (P)	W	Current (I)	A
Thermal resistance (R_{th})	K/W	Electrical resistance (R)	Ω
Specific heat (C_{th})	J/K	Electrical capacitance (C)	F

Table 1: Thermal quantities and their electrical duals

3. Previous Work

Many studies have been conducted in the area of thermoelectric cooling. Most of them focus on improving the material that the device is made of and the manufacturing process. In other words, their aim is to improve the figure of merit of TECs. Reference [1] presents a comprehensive survey on TEC principles and the manufacturing advances in the recent years.

Reference [6] tries to increase COP^{basic} by restricting the ΔT ($= T_h - T_c$) to smaller values. However, this is not a practical solution in the microprocessor cooling application, because the TECs are sandwiched between the heat spreader and the TIM and as a result T_h cannot be directly controlled. One can still use a better heat sink and fan assembly in order to insure that T_h does not go beyond a certain value. This solution is not cost efficient and sometimes due to the system form factor, it is not possible to install a larger heat sink or fan.

Reference [7] uses TECs in order to cool down microprocessors in a datacenter and reduces the total cooling cost while maintaining the same reliability. This paper mainly focuses on the steady-state analysis of TECs and uses a constant COP^{basic} for modeling TECs which is too coarse grain.

Reference [8] shows the significance of the transient behavior of TECs. It presents two simple controllers: *threshold based controller,* which turns on or off the TEC when the temperature goes above or below a certain temperature, and *maximum cooling based controller,* which uses two different temperature thresholds and hysteresis to decrease the number of ON and OFF transitions of the TEC. In both controllers, TECs are supplied with a constant current to effect a state change.

References [9] and [10] formulate the selective deployment of TECs on top of a chip in order to achieve the maximum cooling (lowest temperature). The idea is that excessive deployment of TECs adversely affects the temperature of the device because of lateral heating which negatively affects other TECs. Moreover, deploying unnecessary TECs increases the power consumption of the cooling solution. The focus of these two papers is on the steady-state analysis of TECs.

Reference [11] tries to improve the performance of TECs by optimizing the dimensions of N and P-pellets.

4. Redefining the COP

The major drawback of TECs is their low COP^{basic}. Any value lower than one means the device adds more heat to the system than the cooling it provides. Even COP^{basic} values slightly higher than one are problematic since the system would require larger heat sinks/ and/or stronger fans to dissipate the excessive heat that is generated by TECs. Differentiating COP^{basic} in Equation (7) with respect to I gives the current value that maximizes COP^{basic} [1][4]. This current is called $I_{COP(basic),opt}$ and is:

$$I_{COP(basic),opt} = \frac{\alpha \Delta T}{R_{TEC}\sqrt{ZT_{avg} + 1} - 1} \qquad (8)$$

where T_{avg} is defined as the average of T_h and T_c. Plugging $I_{COP(basic),opt}$ into Equation (7) gives the maximum value COP^{basic}_{max}, which may be written as follows:

$$COP^{basic}_{max} = \frac{T_c\left(\sqrt{1 + ZT_{avg}} - \frac{T_h}{T_c}\right)}{\Delta T\left(1 + \sqrt{1 + ZT_{avg}}\right)} \qquad (9)$$

As can be seen, COP^{basic}_{max} is proportional to T_c and ZT_{avg}. Moreover, it is inversely proportional to ΔT. These relations suggest three ways for increasing COP^{basic}_{max}:

1. Using materials with high figure of merit. This can be done by selecting better materials and improving fabrication techniques. These methods are outside the scope of this paper.
2. Limiting ΔT to low values. As has been discussed previously, this solution is not possible in many applications/platforms.
3. Increasing T_c. An important observation is that TECs perform efficiently when T_c reaches its maximum tolerable value.

Figure 4 shows the dependency of COP^{basic}_{max} on the aforementioned parameters. As can be seen, in order to achieve COP^{basic}_{max} values higher than 4, ΔT should be limited to ~15°C-25°C.
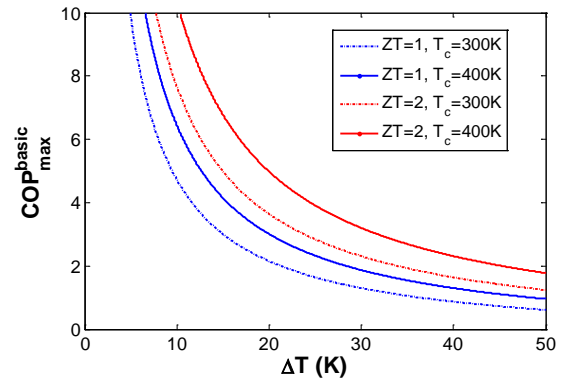

Figure 4: Dependency of COP^{basic}_{max} on ΔT, ZT, and T_c.

Increasing T_c is a possible solution for some applications (other than processor cooling). However, it comes at the cost of increasing the leakage power, which is exponentially dependent on the die temperature [12]. Unfortunately, COP^{basic} does not capture the effect of the leakage power.

Based on the thermal-electrical duality explained in Section 2, a TEC inside a processor package can be modeled using an electrical circuit. This model is shown in Figure 3. The Peltier effect is modeled by two current sources: one is at the bottom which has a negative value, and hence, *absorbs heat* and one is at the top which has a positive value and *releases heat*. The Joule heating effect is also modeled as a current source which charges a capacitor. This capacitor signifies the specific heat of the TEC material. When a TEC is turned on (or its driving current is changed), the Peltier effect appears quickly but the Joule heating effect appears gradually. The reason is that the Joule heating needs to overcome the specific heat of the TEC material (charge the capacitor) whereas the Peltier effect only pumps heat from one side to the other side [8][13]. The two RC networks model the rest of thermal package at the top and the bottom of the TEC. The novelty in this model is the addition of $P_{leakage}$ as a function of T_{die}. Note that the temperature which affects the leakage power (T_{die}) is not equal to T_c but it is a function of it.

Using the model given above, the *system COP* or COP^{sys}, which captures the die temperature-dependent leakage power of the system, is written as follows:

$$COP^{sys} = \frac{N\left(\alpha T_c I - K_{TEC}\Delta T - \frac{1}{2}R_{TEC}I^2\right) - P_{leakage}(T_{die})}{N(\alpha \Delta T I + I^2 R_{TEC}) + P_{leakage}(T_{die})} \qquad (10)$$

The leakage power decreases the amount of cooling (nominator) and increases the power consumption of the TEC (denominator). COP^{sys} equals zero when the cooling and the heating amounts are

identical. Note that in this formulation, we do not consider $P_{dynamic}$ for the system as its value is not controlled by TEC (neither directly by the TEC current nor indirectly by the temperature). Maximizing COP^{sys} is equivalent to achieving the maximum cooling while expending the least amount of power; this is called the *maximum CoP cooling* (MCPC) strategy. Defining COP^{sys} helps find the MCPC current for driving TEC. This current is a function of leakage power and it changes based on the chip condition, whereas COP^{basic} is independent of the chip condition. Differentiating Equation (10) with respect to I does not give a closed-form expression like the one presented in Equation (8). As a result, we perform different experiments with several current values to find the one that maximizes COP^{sys}. Although this method seems to be time consuming, in fact it is not an expensive proposition, because this is done offline during the design phase.
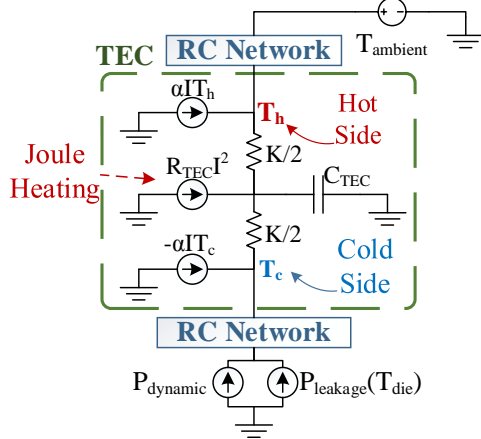

Figure 5: An electrical model for a TEC embedded inside a processor package

5. Platform-Dependent, Leakage-Aware Cooling Policy for TECs

As it will be demonstrated in the next section, the driving current of TECs for the MCPC strategy is quite different from that of the *maximum temperature reduction* (MTR) strategy. Based on this observation, one can establish a platform-dependent, leakage-aware cooling policy according to the target platform/application (high-performance vs. low-power). The first target platform (high-performance) employs the MTR policy whereas the second one (low-power) adopts the MCPC strategy. The optimum current which is suitable for the MTR case is called I_{MTR} whereas the optimum current for the MCPC case is called I_{MCPC}. As explained previously, the Peltier effect appears before the Joule heating. This behavior is usually used for transient cooling. Hence, for each platform type, a set of currents should be found; one that works best in the steady state and another one which is suitable for the transient cooling.

Based on the aforementioned explanations, Algorithm 1 describes a platform-dependent, leakage-aware cooling algorithm, which determines the TEC driving current for both steady state and transient regimes of operation. In this algorithm, *platform type* is set based on the requirements of the target hardware or application; *chip condition* refers to current die temperature, which can be read from temperature sensors deployed on the chip surface. The set of TEC currents for different conditions and the *thermal network time constant (TRC)* are determined based on a thorough analysis of the TEC thermal behavior. Notice that one can extend this algorithm to provide an (online) *adaptive cooling policy,* which uses a *peak COP tracking* method (via a look-up table or an online optimizer) in order to set the driving current of the TECs at a finer time granularity based on dynamically-updated die temperatures. Details are omitted.

Transient cooling is superior only for the duration of T_{RC}, hence, a timer is set up in order to stop using transient current if the

---

Algorithm 1: Platform-dependent, leakage-aware cooling policy for setting the current of TECs

Given: platform type, chip condition, $\{I_{MTR}^{Emergency}, I_{MCPC}^{Emergency}, I_{MTR}^{Steady}, I_{MCPC}^{Steady}\}$, and T_{RC}.
Determine: I_{TEC}.
1   **IF** *platform type* = high-performance **THEN**
2   **IF** *chip condition* = emergency **THEN**
3   **IF** *Timer<T_{RC}* **THEN**
4         $I_{TEC}=I_{MTR}^{Emergency}$
5   **ELSE**
6         $I_{TEC}=I_{MTR}^{Steady}$
7   **END IF**
8   **ELSE**
9       $I_{TEC}=I_{MTR}^{Steady}$
10   Reset the *Timer*.
11   **END IF**
12 **ELSE**   // platform type = low-power
13   **IF** *chip condition* = emergency **THEN**
14   **IF** *Timer<T_{RC}* **THEN**
15         $I_{TEC}=I_{MCPC}^{Emergency}$
16   **ELSE**
17         $I_{TEC}=I_{MCPC}^{Steady}$
18   **END IF**
19   **ELSE**
20       $I_{TEC}=I_{MPC}^{Steady}$
21   Reset the *Timer*.
22   **END IF**
23 **END IF**

---

emergency condition lasts more than T_{RC}. Without this timer, if the emergency situation takes longer than T_{RC}, the Joule heating effect will dominate the Peltier effect and the policy will not perform well.

6. Experiments and Discussions
6.1. Simulation Setup

To evaluate the new definition of COP and find the optimum TEC current values for the proposed cooling algorithm, we developed a tool called *Teculator* (i.e., a TEC Simulator) to simulate the behavior of TECs and evaluate their effect in a processor package assembly. This tool is implemented as an extension to HotSpot 5 [14]. Each TEC is modeled in three layers:

1) The bottom layer, which is called the *heat absorption layer*, accounts for the Peltier cooling effect. It also characterizes the thermal resistance and capacitance of the bottom contacts.
2) The middle layer, which is called the *heat generation layer*, captures the Joule heating effect. It also signifies the heat conduction of TEC between the cold and hot layers. The thermal capacitance of this layer allows simulating the transient behavior of a TEC.
3) The top layer, which is called the *heat rejection layer*, models heat rejection. Similar to the cold layer, it accounts for the thermal resistance and capacitance of the top contacts.

The TEC parameters are mostly taken from reference [3]. Missing parameters are taken from other references that use a similar experimental setup. Table 2 provides used references. The only missing information for calculating R_{TEC} is the area of N and P-pellets. Using the 92% packing factor (which is reported in reference [15]), it can be estimated that 46% of the total area of a TEC is occupied by a P-pellet and another 46% is occupied by an N-pellet. Based on this 'assumption' and Equation (4), R_{TEC} is calculated to be $4.98\times10^{-3}\,\Omega$.

The processor package assembly used for simulations has a similar configuration to Figure 3. Table 3 shows dimensions, thermal resistivity, and specific heat of each layer (except the TEC layer, which was discussed earlier). The surface of the chip is tiled

with 16×16 TECs (for a total of 256 TECs). All of these TECs are connected serially and driven by the exact same current value.

Parameter	Value
Seebeck coefficient (α)	3.01×10^4 V/K
TEC electrical resistivity (ρ)	1.08×10^{-5} Ω.m
TEC thermal conductivity (k)	1.2 W/(m.K)
TEC specific heat capacity (C) [13]	1.20×10^6 J/(m³K)
TEC dimension [3][16]	0.5mm×0.5mm×8μm
TEC contact dimension (each side)	0.5mm×0.5mm×46μm
TEC-metal contact thermal resistivity	8×10^{-6} (m².K)/W
TEC-metal contact electrical resistivity	10^{-10} Ω.m²

Table 2: TEC parameters used in the simulations

Layer	Thermal Resistivity (m.K)/W	Specific Heat J/(m³.K)	Dimensions
Chip	1.0×10^{-2}	1.75×10^6	8mm×8mm×150μm
TIM 1	2.5×10^{-1}	4.00×10^6	1.6mm×1.6mm×20μm
Heat Spreader	2.5×10^{-3}	3.55×10^6	30mm×30mm×1mm
TIM 2	2.5×10^{-1}	4.00×10^6	30mm×30mm×1mm
Heat Sink	2.5×10^{-3}	3.55×10^6	60mm×60mm×6.9mm

Table 3: Thermal resistivity, heat specific and dimensions of each layer of the chip package

McPat [17] is used in order to estimate $P_{leakage}(T_{die})$. A Xeon processor (whose model comes with the tool) is simulated using the 32nm CMOS process technology. The simulation is done for nine temperature values distributed evenly in the range of 310K to 390K. Next, a 4th order approximation of these values is derived by using Excel. Figure 6 shows the curve fitting result. Note that the power value is normalized to the area in order to find the power density.
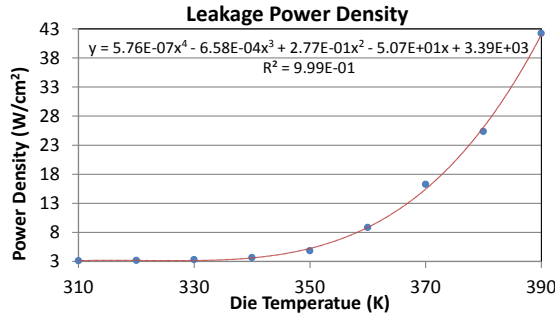

Leakage Power Density

$y = 5.76E\text{-}07x^4 - 6.58E\text{-}04x^3 + 2.77E\text{-}01x^2 - 5.07E\text{+}01x + 3.39E\text{+}03$
$R^2 = 9.99E\text{-}01$

Figure 6: Curve fitting for the leakage power density of a Xeon processor in 32nm process technology

It is assumed that the chip has a uniform heat flux of 70 W/cm². A hot spot is created at the center of the chip with a variable additional heat flux taking values of 500 and 1,000 W/cm². The area of this hot spot is 0.5mm×0.5mm.

6.2. Simulation Results

Figure 7(a-d) shows the result of several steady-state experiments with different TEC current values ranging from 0A to 11A. For every current value, two local (hotspot) heat fluxes, i.e., 500 and 1,000 W/cm², are considered. Figure 7(a) shows the temperature of hot spot. It can be seen that for both heat flux values, I_{TEC}=5A gives the maximum temperature decrease compared to the I_{TEC}=0A case. This decrease is equal to 14.7℃ and 14.2℃ for the high heat flux and the low heat flux cases, respectively. An interesting point is that the amount of temperature drop for the high heat flux case is somewhat larger than that of the low heat flux case. This confirms the claim that TECs work better in higher temperatures. As a result of this experiment, I_{MTR}^{Steady} is set to 5A.

Figure 7 (b) shows COP^{sys} for different current values. It can be seen that I_{TEC}=1A maximizes COP^{sys} for both heat fluxes. This experiment reveals four important points:

1) The current value that maximizes COP^{sys} is not equal to the current that maximizes the temperature decrease. This

emphasizes the distinction between two different objectives, i.e., MTR and MCPC.

2) It is interesting that COP^{sys} has a value higher than unity when I_{TEC}=0A, i.e., the TECs eventually cools the chip even when they are off. This is due to the high heat conductivity of the TECs. In other words, considering Equation (10), ΔT takes a negative value, which leads to a high positive value for COP^{sys}. Note that COP^{basic} (which is not shown in the figure) does not behave in the same way as COP^{sys}. Indeed, COP^{basic} is undefined when the current is equal to zero since the denominator is equal to zero. Hence, this second point could not be stated if COP^{basic} were used instead of COP^{sys}. Moreover, note that COP^{basic} is independent of the leakage power, which results in a fixed optimum current level for driving TECs irrespective of the chip temperature.

3) The COP^{sys} value for I_{TEC}=1A is larger than that of I_{TEC}=0A. This means that turning on TECs not only cools down the processor by more than 6℃ but also the cooling acts more efficiently by 7% and 5% for the high and the low heat fluxes cases, respectively. Again, note that the TECs have higher COP^{sys} values when they are working at higher die temperatures (i.e., higher hotspot heat fluxes).

4) I_{MCPC}^{Steady} can be set to 1A.

Figure 7(c) shows the total leakage power in the chip. Note that since T_{die} is a function of T_c (temperature of the cold side of TEC), leakage power is minimized when T_c is minimized.

Figure 7(d) depicts the absorbed heat per unit time by all TECs deployed on the surface of the processor for different current values. As can be seen, this value monotonically increases with current. Most of this heat is the heat generated by the Joule heating effect as well as the heat generated because of the increase in the leakage power. Note that only part of this heat is pumped by the Peltier effect and the other part is exchanged through the heat conduction because of the negative ΔT that exists across some TECs. Also since the processor cooling package cannot dissipate this much heat (which are absorbed from one side and released to the other side of TECs), the temperature of the hot spot rises after I_{TEC}=5A.

To study the transient cooling behavior of TECs, an experimental setup similar to the steady-state case for the low heat flux (500 W/cm²) is used. At time instance 0.1s, the heat flux is increased to 1000 W/cm² and this elevated heat flux lasts until time instance 1.1s. (This increase can be the result of increase in the dynamic or the leakage power of the chip.) Finally, the hot spot heat flux value is reset back to 500 W/cm². Figure 7(e,f) shows the results of this experiment. Before the high heat pulse and after that, I_{TEC} is set to I_{MTR}^{Steady} or I_{MCPC}^{Steady} based on the type of the target system and objective function (MTR or MCPC). During the high heat flux, the current is increased to values higher than their steady state (as high as 11A). The initial die temperature is set to 68.69℃ and 76.9℃ for MTR and MCPC scenarios respectively. Notice that these values are equal to the steady-state temperature of the hot spot in each scenario.

Figure 7 (e) shows the temperature change during the heat flux pulse in the MTR case. I_{TEC}=5A is retained as a reference, which means that the TEC driving current is not changed during the heat flux pulse. For clarity, we only show two main cases; other cases produce inferior results. We find that when the transient current is set to 6A, the resultant temperature is below the baseline's temperature during the pulse period, although the temperature difference decreases as time passes. On the other hand, when the current is set to 8A, the temperature drops quickly but after ~0.3s, it exceeds that of the baseline (I_{TEC}=5A). This experiment suggests that we set $I_{MTR}^{Emergency} = 6A$.

Figure 7(f) presents the COP^{sys} change during the heat pulse. Current values higher than 2A (e.g., 3A as shown in the figure)
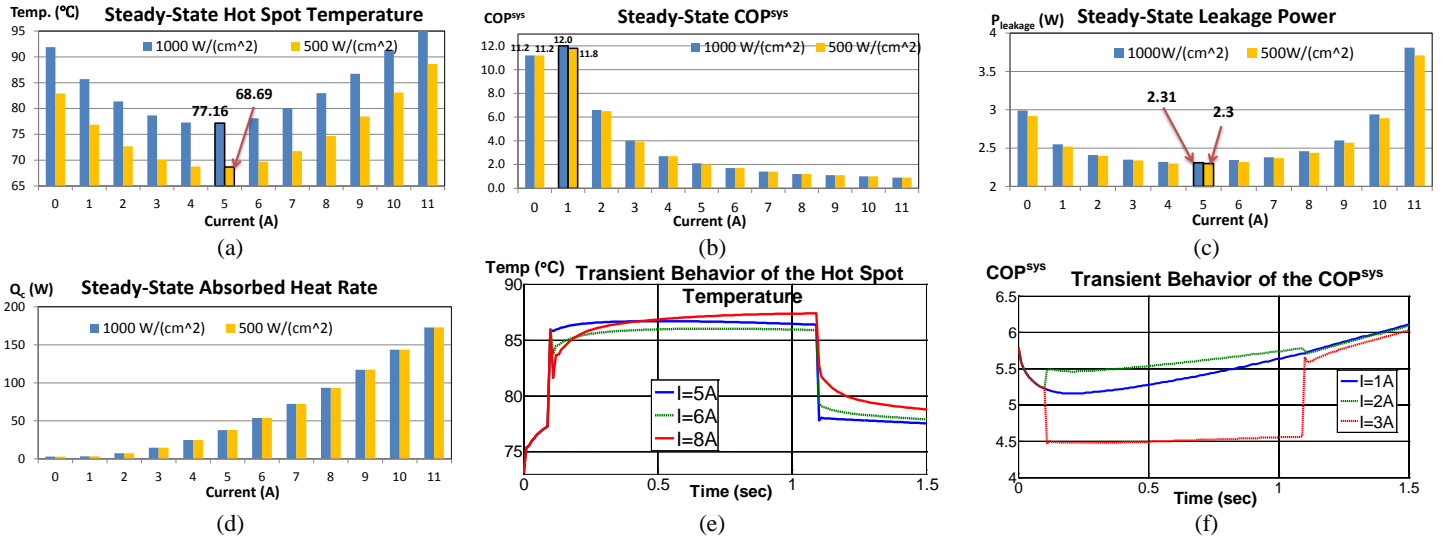
Figure 7 (a-d) Results of steady state experiments with TECs. **(a)** Hot spot temperature, **(b)** COP^{sys} values, **(c)** leakage power, and **(d)** absorbed heat per unit time by all TECs for different current values ranging from 0A to 11A. **(e-f)** Results of transient cooling experiments with TECs. **(e)** Hot spot temperature change and **(f)** COP^{sys} change when applying a one-second heat pulse to the center of the chip to make a hot spot.

drastically degrade COP^{sys}. Conversely, with I_{TEC}=2A, COP^{sys} is improved during the heat pulse. This improvement fades out at the end of the pulse. This experiment suggests that we set $I_{MCPC}^{Emergency} = 6A$. Also based on these two experiments, the thermal network time constant (T_{RC}) should be set to a value slightly higher than 1s.

7. Conclusion

This paper investigated various venues to improve the performance of TECs embedded inside a processor package. First a new definition for COP of TECs considering the system's leakage power dissipation, which is exponentially dependent on the temperature, was presented. Next, it was shown that well-tuned TECs in the MCPC mode can improve the COP of an entire cooling system by 7% while reducing the temperature of chip hotspots by 6℃. Moreover, it was shown that the TEC driving current that yields the maximum drop in the chip temperature is quite different from the one that runs the TEC in its highest COP state (5A vs. 1A). Finally, a platform-dependent, leakage-aware cooling policy was proposed in which the TEC driving current was set based on the target platform/application (high-performance vs. low-power) and the actual conditions of the chip (emergency vs. preventive thermal management.)

Acknowledgement

References

[1] A. Bar-Cohen and P. Wang, "On-Chip Thermal Management and Hot-Spot Remediation," in *Nano-Bio- Electronic, Photonic and MEMS Packaging*, Springer, 2010.

[2] "Tellurex - An introduction to thermoelectrics." [Online]. Available: http://www.tellurex.com/technology/design-manual.php. [Accessed: 25-Mar-2013].

[3] 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," *Nature Nanotechnology*, vol. 4, no. 4, pp. 235–238, 2009.

[4] J. Sharp, J. Bierschenk, and H. B. Lyon, "Overview of Solid-State Thermoelectric Refrigerators and Possible Applications to On-Chip Thermal Management," *Proceedings of the IEEE*, vol. 94, no. 8, pp. 1602 –1612, Aug. 2006.

[5] D. M. Rowe, Ed., *Thermoelectrics Handbook: Macro to Nano*, 1st ed. CRC Press, 2005.

[6] J. Bierschenk and D. Johnson, "Extending the limits of air cooling with thermoelectrically enhanced heat sinks," in *ITherm*, 2004, pp. 679–684 Vol.1.

[7] 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 *Proceedings of the 38th International Symposium on Computer Architecture*, New York, NY, USA, 2011, pp. 331–340.

[8] B. Alexandrov, O. Sullivan, S. Kumar, and S. Mukhopadhyay, "Prospects of active cooling with integrated super-lattice based thin-film thermoelectric devices for mitigating hotspot challenges in microprocessors," in *Proceedings of Asia and South Pacific Design Automation Conference*, 2012, pp. 633–638.

[9] J. Long, S. O. Memik, and M. Grayson, "Optimization of an on-chip active cooling system based on thin-film thermoelectric coolers," in *Proceedings of the Design, Automation and Test in Europe*, 2010, pp. 117–122.

[10] J. Long and S. O. Memik, "A framework for optimizing thermoelectric active cooling systems," in *Proceedings of the Design Automation Conference*, 2010, pp. 591–596.

[11] P. Y. Hou, R. Baskaran, and K. F. Böhringer, "Optimization of Microscale Thermoelectric Cooling (TEC) Element Dimensions for Hotspot Cooling Applications," *Journal of Electronic Materials*, vol. 38, no. 7, pp. 950–953, Jul. 2009.

[12] W. Huang, K. Rajamani, M. R. Stan, and K. Skadron, "Scaling with Design Constraints: Predicting the Future of Big Chips," *IEEE Micro*, vol. 31, no. 4, pp. 16–29, 2011.

[13] R. Yang, G. Chen, A. Ravi Kumar, G. J. Snyder, and J.-P. Fleurial, "Transient cooling of thermoelectric coolers and its applications for microdevices," *Energy Conversion and Management*, vol. 46, no. 9–10, pp. 1407–1421, Jun. 2005.

[14] K. Skadron, M. R. Stan, W. Huang, S. Velusamy, K. Sankaranarayanan, and D. Tarjan, "Temperature-aware microarchitecture," in *Proceedings of the 30th International Symposium on Computer Architecture*, 2003, pp. 2–13.

[15] K. Wang, R. Baskaran, and K. Bohringer, "Template Based High Packing Density Assembly for Microchip Solid State Cooling Application," in *3rd Conference on Foundations of Nanoscience: Selfassembled Architectures and Devices (FNANO)*, 2006.

[16] M. P. Gupta, M. -h. S. Sayer, S. Mukhopadhyay, and S. Kumar, "On-chip Peltier cooling using current pulse," in *2010 12th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)*, June, pp. 1–7.

[17] S. Li, J. H. Ahn, R. D. Strong, J. B. Brockman, D. M. Tullsen, and N. P. Jouppi, "McPAT: an integrated power, area, and timing modeling framework for multicore and manycore architectures," in *Proceedings of the 42nd International Symposium on Microarchitecture*, New York, NY, USA, 2009, pp. 469–480.