



# Time scale matching of dynamically operated devices using composite thermal capacitors



Craig E. Green, Andrei G. Fedorov\*, Yogendra K. Joshi

George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, GA 30332-0405, United States

## ARTICLE INFO

### Article history:

Received 22 November 2013  
Received in revised form  
14 May 2014  
Accepted 19 May 2014  
Available online 14 June 2014

### Keywords:

Multicore  
Hotspot  
Core migration  
Thermal  
Pulsed RF  
Computational sprinting

## ABSTRACT

A new thermal management solution is proposed to maximize the performance of electronics devices with dynamically managed power profiles. To mitigate the non-uniformities in chip temperature profiles resulting from the dynamic power maps, solid–liquid phase change materials (PCMs) with an embedded heat spreader network are strategically positioned near localized hotspots, resulting in a large increase in the local thermal capacitance in these problematic areas. The resulting device, called composite thermal capacitor (CTC), can theoretically produce an up-to-twenty-fold increase in the time that a thermally constrained high heat flux device can operate before a power gating or core migration event is required. A prototype CTC that monolithically integrates micro heaters, PCMs and a spreader matrix into a Si test chip was fabricated and experimentally tested to validate the efficacy of the concept and to gain an insight into phase change heat transfer in a spatially-confined environment on the microscale. As the most significant result, an increase in allowable device operating times by over  $7\times$  has been experimentally demonstrated, while operating a device at heat fluxes approaching  $400\text{ W/cm}^2$ .

© 2014 Elsevier Ltd. All rights reserved.

## 1. Introduction

Dynamic operation and control is an essential tool for the thermal management of a number of next generation electronic devices that suffer from localized hotspots with large heat fluxes that cannot be dissipated by the baseline cooling system designed for dissipation of the time-averaged power load. Examples of such devices with pulsed or time varying power loads include (i) RF transmitters, which generate heat in short, periodic, pulses during the data transmission process; (ii) power electronics, which use short modulated voltage pulses to perform tasks such as conversion of power from direct to alternating current; and (iii) many-core microprocessors, which use techniques such as thread migration to actively move high power consumption computations from hotter to cooler areas of the die to lower peak temperatures and temperature gradients.

Due to limited baseline cooling resources, the time that many high heat flux devices can operate before load mitigation approaches must be employed is limited. In some cases, such as data transmitters, the dynamic operation of the device is coupled to its functionality, while in others (e.g. microprocessors) the dynamic architecture is driven by thermal limitations. Both types of systems can benefit from a thermal solution that is specifically geared

towards addressing the dynamic nature of the devices' heat generation.

For applications where the duration of the operational pulse is integral to the device's functionality, it is imperative that the component is able to operate for the entire pulse without exceeding temperature limits. Radar systems, for example, transmit and receive data in high power pulses that can exceed several  $\text{kW/cm}^2$  [1] and in some cases must operate in thermally challenging environments such as outer space or unmanned aerial vehicles [2]. While the transmitters' power consumption during transmission can be large, their duty cycles – the fraction of time when the device is actively consuming power – are typically less than about 25% [3]. In power electronics, extremely large voltages (sometimes kV) are switched at varying pulse lengths using the pulse width modulation technique in order to convert the electrical signal to the desired output [4]. The heat flux associated with these switching events can also be several  $\text{kW/cm}^2$  on time scales ranging from ms to hundreds of ms [5,6].

The exponential growth in the number of on chip transistors so reliably predicted by Moore's law has proven to be a powerful driver for increases in computing performance over the past 40 years, although limitations associated with wire delay [7], power consumption, and heat generation [8] have recently become significant challenges to traditional transistor scaling. The desire to maintain the historic rate of advancement in the microelectronics industry, while avoiding the roadblocks associated with power consumption and wire delay have led to the consideration of

\* Corresponding author. Tel.: +1 404 385 1356.

E-mail addresses: [AGF@gatech.edu](mailto:AGF@gatech.edu), [cgreen8@gatech.edu](mailto:cgreen8@gatech.edu) (A. G. Fedorov).

several disruptive design strategies for next generation devices, including many-core processors and 3D vertical integration [9,10].

In a many-core system, the thermal profile across the chip can be made more uniform by actively migrating computations from hotter to cooler areas of the chip, reducing the problem of localized hotspots that have become a major challenge in modern architectures [11]. While this Dynamic Core Migration (DCM) scheme can mitigate hotspots for most cores, serial cores with their potentially higher power densities, larger size, and smaller number may still experience hotspots [12]. To compensate for the higher power densities the serial cores will either experience more throttling events during an intra-migration time slice [13], higher migration frequencies, or a dedicated local hotspot cooling solution would be required to handle the additional thermal overhead [14]. In DCM schemes, there is parasitic computational cost associated with each throttling event that can become significant over time when the cycling is too rapid [15]. Furthermore, rapid thermal cycling can lead to reduced lifetime reliability for the chip [16]. To minimize the performance losses associated with these gating and throttling events, an optimized system should be designed that can operate for longer periods without requiring an idle for cool-down, and have as short of an idle time as possible.

## 2. Proposed cooling method

A commonly used approach to handle hot spots is to bring an embedded liquid cooler to the hotspot, to locally enhance heat transfer. Such dedicated hotspot coolers can significantly increase the complexity of the overall thermal solution, often requiring additional coolants, piping, or off chip regeneration [17,18]. Furthermore, hotspot coolers are often steady state solutions, as operating a liquid cooler in a dynamic fashion requires the use of valves or other active flow control measures to synchronize the coolant delivery with the operation of the electronic device. Recognizing these challenges of attempting to locally increase a cooler's heat transfer coefficient, the approach investigated in this work instead seeks to locally increase the thermal capacitance in thermally troublesome areas of the chip to maximize the time that a core or device can operate before reaching its thermal threshold.

As shown schematically in Fig. 1, for dynamically operated electronics, increasing the thermal capacitance of a device can significantly decrease the frequency of core hopping, gating, or throttling events. This in turn reduces the parasitic computational overhead associated with the DCM implementation. Thus, matching a device's thermal capacitance to its intrinsic dynamics of power dissipation can “homogenize” the thermal time scales of devices with very different power dissipation profiles. Furthermore, for high power devices with defined pulse lengths such as

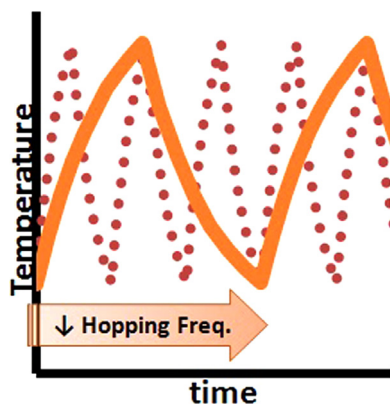


Fig. 1. Impact of increased thermal capacitance on core hopping frequency.

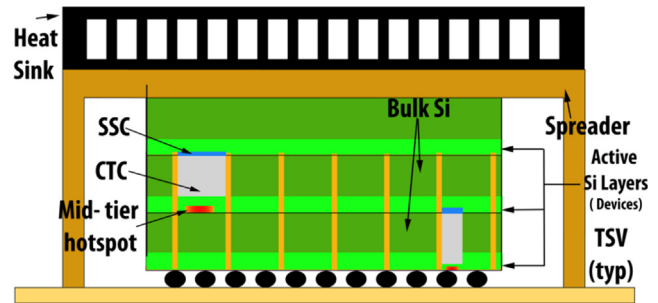


Fig. 2. Schematic of CTC integration in a 3D chip stack.

RF electronics, a sufficiently large increase in local thermal capacitance can ensure that the device can operate for its entire pulse without exceeding its temperature limits.

In order to locally alter the thermal capacitance of the devices, a portion of the substrate (Si, SiC, etc.) on the inactive back side of the chips can be etched away and a material with a higher thermal capacitance, for example solid–liquid phase change materials (PCMs), can be placed in the cavity created by removal of substrate material. An embodiment of this approach is shown schematically in Fig. 2, where PCMs have been inserted into the inactive back side of several dies in a 3D chip stack. The PCMs, named because of their ability to reversibly melt/solidify during heating/cooling processes, can absorb a large amount of thermal energy at a relatively constant temperature. One challenge of utilizing PCMs is that their typically low thermal conductivities ( $\kappa$ ) limit the amount of material that can be melted prior to the device reaching its threshold temperature due to significant temperature non-uniformity. This can be mitigated by using a “composite thermal capacitor” (CTC), consisting of PCM incorporated into a high thermal conductivity matrix to enhance heat spreading and therefore improve PCM utilization. The CTC can be manufactured using standard batch microfabrication techniques, making the proposed solution amenable to the level of high volume integration that is needed for devices in the consumer electronics market.

## 3. Performance characterization

### 3.1. Analysis of the impact of spreading on device operating time modulation

A key feature of the CTC design philosophy is the enhancement of lateral spreading and energy storage into the PCM by improving the effective thermophysical properties, specifically the effective thermal conductivity ( $\kappa_{eff}$ ) of the composite matrix. It is valuable to first examine how  $\kappa_{eff}$ , along with the other relevant PCM properties – density ( $\rho$ ), specific heat ( $c_p$ ), and latent heat of solid to liquid phase change ( $h_{sl}$ ) – affect the physics of the problem and, in turn, the achievable device operating times. Concentrating on the contribution of lateral spreading to the achievable enhancements in device operating times will allow a more informed decision on whether design of the overall CTC should focus just on the area directly above the hotspot, or on using a larger cross-sectional area accessible through thermal spreading.

A simple model that can be used to study the impact of lateral spreading on device operating times is an annular region of PCM surrounding a cylindrical block of Si of radius  $R_{Si}$ , and height  $z$ , as shown in Fig. 3. At the bottom of the Si region is a heat flux boundary condition that represents a localized hotspot. Because the PCM is confined to the annular region at the periphery of the hotspot, this arrangement highlights what can be gained from lateral spreading specifically.

Download English Version:

<https://daneshyari.com/en/article/547100>

Download Persian Version:

<https://daneshyari.com/article/547100>

[Daneshyari.com](https://daneshyari.com)