Novel Junction Level Cooling in Pulsed GaN Devices

Tapan G. Desai[1], Daniel Piedra[2], Richard Bonner[1], Tomas Palacios[2]
[1]Electronics Products Group
Advanced Cooling Technologies, Inc.
1046 New Holland Ave.
Lancaster, PA, USA 17601
[2]Dept. of Electrical Engineering and Computer Science
Massachusetts Institute of Technology,
Cambridge, MA, USA 02139

Email: tapan.desai@1-act.com

ABSTRACT

Gallium nitride (GaN) based RF power transistor technology offers the unique combination of higher power, higher efficiency and wider bandwidth. However, the extremely high power densities create new challenges for heat dissipation. In pulsed GaN devices, each duty cycle consists of an active period of heat generation followed by an inactive period. The device temperature oscillates causing thermal stresses leading to device fatigue and life reduction. In this paper, we present a novel junction level cooling technique based on a compact thermal storage design that involves phase change material (PCM) filled micrometer-sized grooves etched in the semiconductor substrate. PCM located close to the junction absorbs waste heat during the active period and dissipates the heat to a heat sink during the inactive period. Computational simulations proved the feasibility of the concept and showed reduction in junction level temperatures. High electron mobility transistors (HEMTs) were fabricated on a GaN-on-silicon wafer with micrometer-sized grooves. The selection of appropriate PCM (critical for concept's success) to completely fill the grooves was done by performing wetting tests. DC and pulsed characterization of the new PCM-enabled devices showed up to 10% improvement in the electrical device performance (due to enhanced thermal management) compared to baseline GaN transistors.

KEY WORDS: Junction level cooling, GaN devices, phase change material, pulsed devices, thermal storage.

NOMENCLATURE

T_m	melting temperature, °C
C_p	specific heat, J/m^{3}K
I_{ds}	Drain current at a given drain-to-source voltage
$I_{ds,max}$	Maximum drain current
V_g	gate voltage, V

Greek symbols

ρ	mass density (kg/m^3)
κ	thermal conductivity (W/mK)

Introduction

RF and microwave applications are becoming increasingly important with the huge demand for wireless telecommunications which is commensurate with the need for high power, high frequency transistors. The RF transistors are in wide demand to support the mobile phone network and radar applications. Until 2005, silicon LDMOS (laterally diffused metal oxide semiconductor) covered about 90% of the high power RF applications in the 2GHz and higher frequency range; the 10% remaining market share was addressed by gallium arsenide (GaAs) high electron mobility transistor (HEMT) technology. This equilibrium is changing considerably by the introduction of gallium nitride (GaN) HEMT technology. The next generation GaN-based RF power transistor technology offers the unique combination of higher power, higher efficiency and wider bandwidth than competing GaAs and silicon based technologies.

AlGaN/GaN HEMTs have become the preferred option for solid state amplifiers in the 1-40 GHz frequency range. With an output power density of more than 40 W/mm at 4 GHz [1], these devices offer 10 times higher power density, higher efficiency levels, higher maximum operating temperatures, and better impedance matching than silicon-based electronics. GaN transistors enable devices to run at higher voltages (bias voltages of 48 or 65V) and hence higher output power. Another important application of GaN transistors is high voltage switches in the next generation of power electronics. It has been predicted that a wider use of power electronics could reduce the consumption of electricity in the U.S. by 15-20%.

However, the extremely high power densities available in GaN devices create new challenges for heat dissipation. Therefore, regardless of the unprecedented power densities demonstrated in these devices, commercial GaN devices typically operate at much lower power densities (4-6 W/mm, expressed in power per unit length of gate) to stay within a maximum allowable junction temperature. The device junction temperature needs to be maintained below 150-175ºC to minimize degradation in the transport properties of the semiconductor and, more importantly, to assure good reliability. Even when operating at the relatively low power densities of 4 W/mm, these devices

are grown on expensive Silicon Carbide (SiC) substrates, which have a thermal conductivity 3 times higher than silicon.

The thermal management technologies for heat dissipation involve using high thermal conductivity substrates such as SiC [2] and diamond [3, 4]. In addition, they need to be attached to oversized heat sinks, which significantly reduce the system scalability. The heat sinks are cooled by convective or forced air cooling. Novel heat dissipation approaches that can efficiently extract the heat from the junction level will allow the nitride transistors to perform to their true potential.

Novel Junction Level Cooling Concept

Transmit/Receive (T/R) devices, such as those used in radar applications, typically operate in a pulsed mode, i.e., an "ON" or active period followed by an "OFF" or inactive period. On a GaN chip, the heat is generated at discrete locations due to the electrical resistance faced by current as it traverses from source to drain circuitry, with the heat generation slightly favoring the drain side [5]. During the device operation, the GaN material at the chip level rapidly rises in temperature, as its low mass is not able to absorb the generated heat. However, towards the heat sink end of the stack-up, only the average power dissipated by the chip is observed and not the transient pulses of the T/R devices. The microsecond pulses are dampened by the mass of the thick materials in the stack-up. Hence, a lower resistance heat sink cannot reduce the peak junction temperature or the transient temperature swings (in the range of 50°C) where the majority of the thermal problems reside.

One of the state-of-the-art solutions to alleviate these thermal problems is employment of thermal vias in the substrate. Thermal vias were first utilized in the design of packaging and printed circuit boards (PCBs). Lee *et. al.* studied arrangements of thermal vias in the packaging of multichip modules (MCMs) and found that as the size of thermal via islands increased, more heat removal was achieved but less space was available for routing [6]. Li studied the relationships between design parameters and the thermal resistance of thermal via clusters in PCBs and packaging [7]. However, thermal vias will not be able to reduce the temperature swings (in the range of 50°C) seen in the high power pulsed devices.

In this paper, we present a novel junction level cooling technique for GaN devices operating in the pulsed mode. The technique is based on a compact thermal storage design that involves phase change material (PCM) filled micrometer-sized grooves etched in the semiconductor substrate. The PCM absorbs the heat during the active period and dissipates the heat to the heat sink during the inactive period.

This cooling concept is designed to reduce the peak junction temperature and the transient temperature changes by increasing the effective heat capacitance of the material layers (via the high heat of fusion of PCM) nearest to the chip junction. The melting point of the PCM creates a "stop" for the peak junction temperature provided there is sufficient PCM to absorb all of the heat generated during the active period and relatively long inactive time to release the heat (to the heat sink) and refreeze all of the PCM. An optimal amount of PCM will prevent the chip's transient temperature from rising and falling rapidly, inducing isothermal conditions over the entire duty cycle. Maintaining a constant junction temperature is ideal in terms of increased reliability of the device.

The concept is more suitable in high frequency applications where the transients are on the order of microseconds and only a small amount of PCM (less than μ-gm) is required; which makes the integration of PCM into the device level feasible.

The paper is organized as follows. First, in the "Computational Modeling" section, preliminary simulation results are presented with the selected PCM (erythritol) to validate the concept. The focus of this paper is fabrication and testing of the PCM based junction level cooling concept. Hence, details are provided on the PCM selection procedure by performing wetting and microscopy tests in the "Wetting of PCM in Grooves" section. Then, in the "Prototype Design and Fabrication" section details are provided on the fabrication procedure, followed by the "Electrical Testing" and "Summary & Conclusions" sections.

Computational Modeling

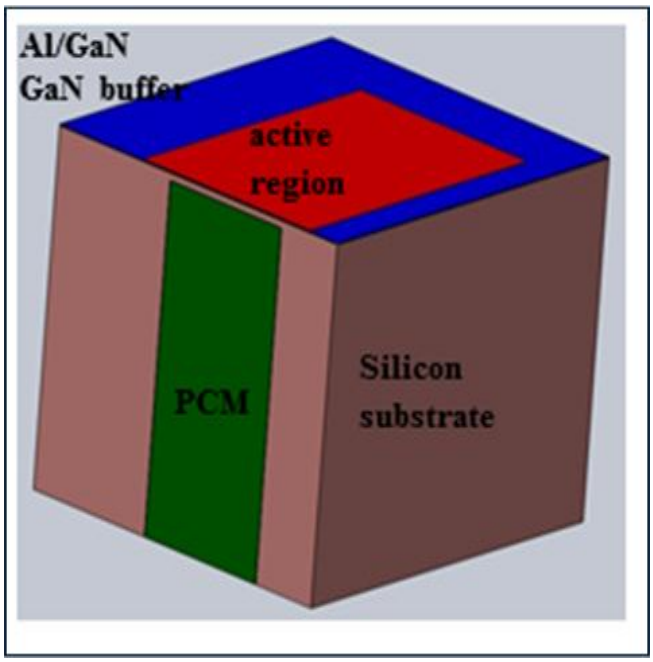

Figure 1: A three-dimensional model consisting of AlGaN, GaN buffer on silicon substrate with embedded PCM. The active (heat source) region is shown in red.

The simulations were carried out as preliminary calculations to validate the junction level cooling concept using the software tool, *CFDesign*. In the experiments, a mask with the several transistors was placed on a copper chuck attached to a thermal heater that allowed the temperature control of the

chuck. No heat sink was employed and the mask was not subjected to any forced convection. This no-heat sink condition does not simulate the real device condition, but rather the conditions that the fabricated devices were exposed during testing in this study. Under these circumstances, the boundary conditions imposed on the top and bottom surfaces of the computational model were natural convection and a convective heat transfer coefficient of 8 W/m^2K. The thermophysical properties assigned to the materials used in simulation are summarized in Table 1. The value of latent heat of fusion for erythritol was chosen from literature as 340 KJ/kg.

Table 1. Thermophysical properties of materials

Material	ρ	C_p	κ
Silicon	2330	664	148
Erythritol	1480(s)	1383(s)	0.733 (s)
	1300 (l)	1100 (l)	0.326 (l)
GaN buffer	1,293	490	6.7
Al/GaN	1,495	490	85

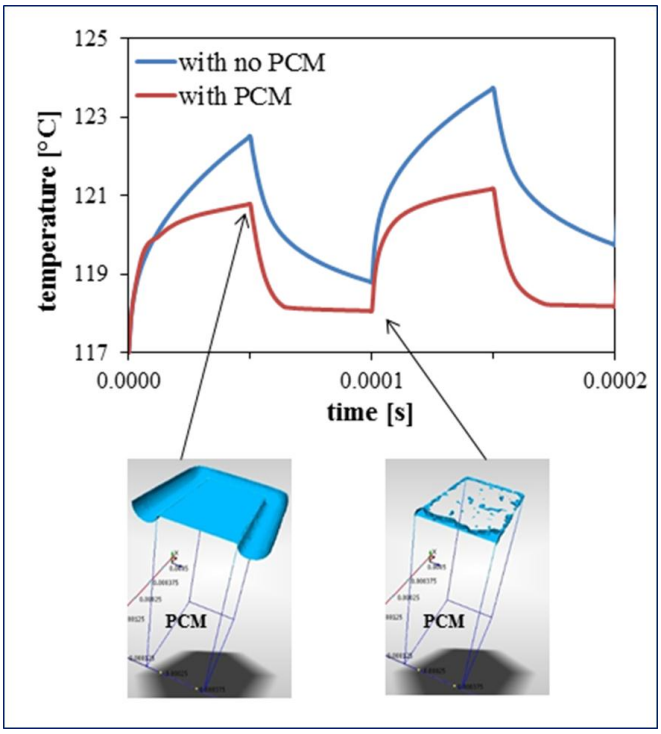

Figure 2: Maximum temperature rise experienced by a transistor in 100 μs pulsed mode with 50% duty cycle. The device with the PCM (red line) experiences lower temperatures than the device with no PCM (blue line). The bottom panel shows the 118 °C isotherm, which represents the PCM solid-liquid interface.

Figure 2 shows the transient behavior of the maximum temperature of a simulated pulsed multifinger transistor operating at 100 μs pulses with 50% duty cycle for devices with PCM and without PCM. During the first 50 μs (on-state of the device), a constant heat power of $4x10^6$ W/m^2 was imposed. During the following 50 μs period (off-state of the device), no heat power was imposed. The initial temperature of the entire model was set at 115°C. As Figure 2 shows, the maximum temperature of the transistor with the PCM (in red) increases to 119°C, 1°C above the melting temperature of erythritol, 118°C, and then the rate of temperature rise reduces. This indicates that the melting of the PCM absorbs the high heat load imposed on the system. In contrast, the maximum temperature of the transistor with no PCM (in blue) shows a continuous increase with time. The bottom panel in Figure 2, shows a 118 °C isotherm, which inside the PCM represents the liquid-solid interface. The two instances indicate the melting-solidification process of the PCM during the cycles. The rise in peak temperature in both simulations, with and without PCM, from one cycle to another is attributed to the chosen value of convective heat transfer coefficient of 8 W/m^2K. If the chosen value is not large enough to dissipate the heat flux imposed on the device during each cycle, then the device temperature will keep climbing with each cycle.

These preliminary simulations indicate that the maximum temperature peaks are lower for device with PCM, showing that the melting of the PCM during the on-state caps the transistor's maximum temperature. This preliminary model can be used as an optimization tool for the amount of PCM, groove geometry and location after including accurate interfacial resistances and boundary conditions.

Wetting of PCM in Grooves

The selection of the PCM is critical for the success of this cooling technology. The device performance depends on good wetting of PCM in the groove. A good wetting ensures the void free filling of micron-sized grooves in the silicon substrate and a low interfacial thermal resistance for the heat to transport from the substrate to PCM inside the groove. Furthermore, a high thermal conductivity of the PCM will aid in transporting the heat quickly to the heat sink. For high performance and improved reliability, it is desirable to maintain the peak junction temperature in GaN devices below 160°C. A preferred operating peak junction temperature would be 120°C. Considering these temperatures, wetting angle tests were performed on two classes of PCM: metallic and organic.

Metallic PCM

Metallic PCM were first tested due to their high thermal conductivity in comparison to the traditional organic PCMs. Two indium (In) based PCMs were tested, pure In [T_m = 157°C] and tin-indium ($Sn_{48}In_{62}$) alloy [T_m = 118°C]; both do not form a solution with silicon in the melt stage. However, the operational challenge was the prevention of In oxidation. A self passivating oxide layer of 80-100 Å and T_m of 1910°C is formed on In's surface at room temperature. Furthermore, the oxide layer thickness increases near In's melting temperature. To prevent oxidation, the wetting studies (contact angle measurement) were performed under inert environment.

Commercially available In shots ordered from Sigma-Aldrich and Indium Corporation were found to have a poor quality and arrived with a thick coating of indium oxide. Better quality was observed when ACT made In and $Sn_{48}In_{62}$ beads in-house

under inert atmosphere from 99.999% high quality ribbons ordered from Indium Corporation. The molten bead was slowly cooled and then cut into two halves for wetting studies.

The solid hemispherical bead was placed on the silicon wafer on the heater plate in the inert compartment box. In Figure 3, we present the photographs taken to estimate the wetting angles; the inset shows the bead before melting for comparison. When the In bead was melted in air (Figure 3(a)), a thick oxide layer was formed that encompassed the melt In similar to a water balloon. This encapsulation effect occurs because the indium oxide layer is under tension and pure In under compression during heating. For flow, In has to break the strong oxide layer. However, under nitrogen flow, the thick oxide layer is not formed and the In bead melts into a droplet form as seen in Figure 3(b). A high contact angle of 115° was measured. Similarly, $Sn_{48}In_{62}$ shows a high contact angle of 130° on silicon in Figure 3(c).

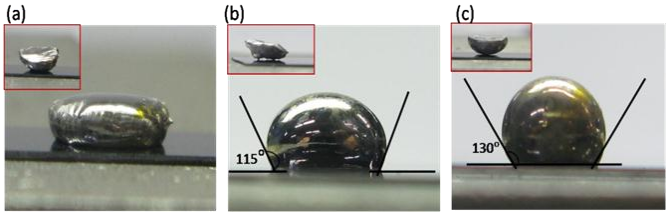

Figure 3: Wetting angle studies of In and $Sn_{48}In_{62}$ on silicon wafer: (a) In melt in air (b) In melt under nitrogen (c) $Sn_{48}In_{62}$ under nitrogen. Inset shows the solid bead before melting.

These results indicate that the In and $Sn_{48}In_{62}$ would not wet the silicon groove and impose a difficulty in filling the grooves. Also, continuous caution for prevention of In oxidation limits the applications for the devices. Gold plating on silicon is known to increase the wetting of liquid metals [8]. In has a wetting angle of 20° on gold coated silicon [9]. However, in addition to oxidation prevention, there are several other concerns in using gold plated silicon substrate:

(1) Uniform gold plating inside the micron-sized grooves is challenging
(2) Molten In forms an alloy with gold, $AuIn_2$, that melts above 540°C
(3) Consumption of In to form $AuIn_2$ reduces the amount of available In for thermal storage

Organic PCM (Erythritol)

Erythritol ((2*R*,3*S*)-butane-1,2,3,4-tetraol) is a sugar alcohol (or polyol) which has been approved for use as a food additive in the United States and throughout much of the world. Due to erythritol's high latent heat of fusion (340 kJ/kg) and T_m of 117.7°C, it is also considered as a PCM candidate. The low thermal conductivity is its only drawback, which it makes up for with an order of magnitude higher latent heat of fusion than In (28.47 kJ/kg). Erythritol, as an organic liquid has a low surface tension and, hence, wets a silicon wafer better than In or $Sn_{48}In_{62}$. This is confirmed by a low wetting angle of 40° shown in Figure 4.

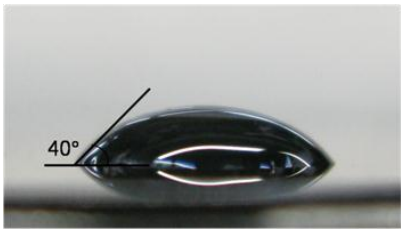

Figure 4: Wetting of erythritol on silicon.

Different sizes and shapes of micron sized grooves were fabricated in the substrate of a GaN-on-Si wafer and were filled with erythritol as shown from optical microscopy images in Figure 5(a) and Scanning electron microscopy (SEM) images in Figure 5(b). Microscopy images confirmed that all grooves are filled by erythritol. Several freeze-thaw cycles were performed to analyze the stability of erythritol inside the groove under cyclic conditions similar to RF devices. Erythritol performed excellently and melted at 120°C during each cycle.

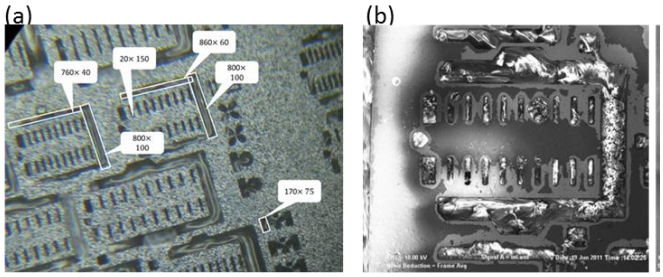

Figure 5: (a) Optical microscopy and (b) Scanning Electron microscopy images of different μm-sized grooves etched on substrate-side of GaN-on-Si filled with erythritol

Thus, erythritol was selected as the PCM candidate. The advantage of erythritol over In based PCM is that while it melts at the temperature similar to $Sn_{48}In_{62}$, there is no need for inert atmosphere during the groove filling procedure. This eases the fabrication and operation issues of transistor devices with PCM. An additional advantage is that erthyritol will not introduce parasitic capacitance due to a weaker electronic component.

It was observed that erythritol present in some of the smaller grooves of dimensions 20×150μm oozed out of the grooves after several solidification-melting, formed a bigger drop and solidified outside the groove. In the device, this would cause a loss of PCM from the groove and reduced the effectiveness of the cooling technology. In addition, this process might be accelerated under gravity during transistor operation. Hence, it is necessary to cap the groove to prevent the loss of PCM. A thin In film was used as the candidate material to cap the grooves.

Prototype Design and Fabrication

In order to test the novel cooling concept, prototype high electron mobility transistors (HEMTs) were fabricated in the clean room facilities at MIT. The starting material was an AlGaN/GaN wafer grown by metal organic vapor phase deposition (MOCVD) by Nitronex Corporation.

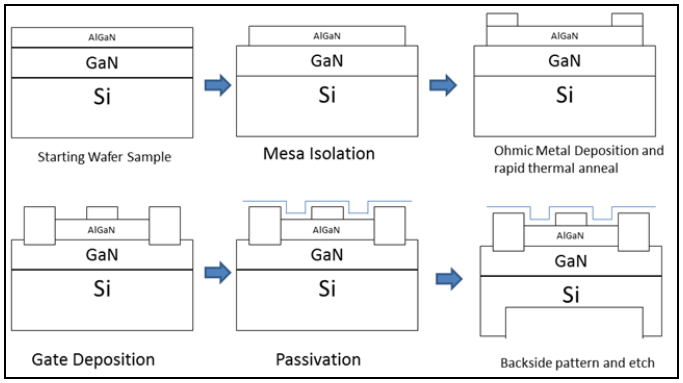

Figure 6: Process flow for the fabrication of PCM-enabled GaN high electron mobility transistors (HEMTs)

The top AlGaN layer had a thickness of 25 nm and the GaN buffer was 1.8 μm thick, while the 550 μm thick Si substrate had a (111) orientation. The process flow of the device fabrication (shown in Figure 6) includes starting with the patterning and isolation of individual transistors. The device isolation was performed through mesa etching with a Cl_2-based plasma etching system. Then, the source and drain ohmic contacts were defined and deposited through electron beam deposition. A Ti/Al/Ni/Au metal stack was used and annealed at 870°C for 30s to form the ohmic contacts. A Ni/Au/Ni gate electrode was then defined by optical lithography and deposited by electron beam deposition. The devices were passivated with a 25nm Al_2O_3 dielectric deposited by atomic layer deposition, and windows were opened in the dielectric by standard photolithography and etching. This fabrication of the GaN transistors followed the standard device process flow for GaN transistors.

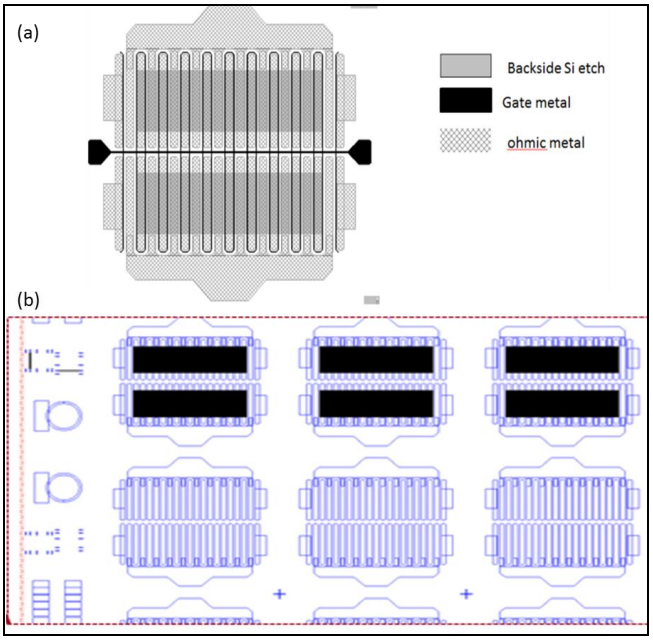

Figure 7: Top-view mask layout of fabricated GaN HEMTs: (a) Single transistor (b) Full mask, the horizontal black bars define the area where the silicon substrate was etched.

Then, the PCM thermal storage structure was fabricated by first defining the region where the PCM would be deposited by photolithography. The silicon substrate was etched through a Bosch dry etch process, and the photoresist was removed by standard cleaning with organics. Prior to silicon etching, it is important to etch away the naturally-grown oxide on the silicon surface by dipping the wafer in buffered HF solution.

Figure 7(a) shows the mask layout of a single transistor with backside silicon etches. Figure 7(b) shows a summary of the overall mask layout used during the device fabrication. A row of transistors with back silicon-etch was followed by a row of transistors without back silicon-etch. This unique design allowed the devices with grooves and without grooves to be tested under the same conditions. The device fabrication was finished with (manual) deposition of the PCM (erythritol) into the silicon grooves by melting the PCM in the grooves. Adequate care was taken to remove the excess PCM lying on the chip outside the groove. The PCM was then encapsulated with a thin In foil. Figure 8 shows the cross-section of a fully fabricated device, as well a top-view and a bottom-view image of the fabricated chip.

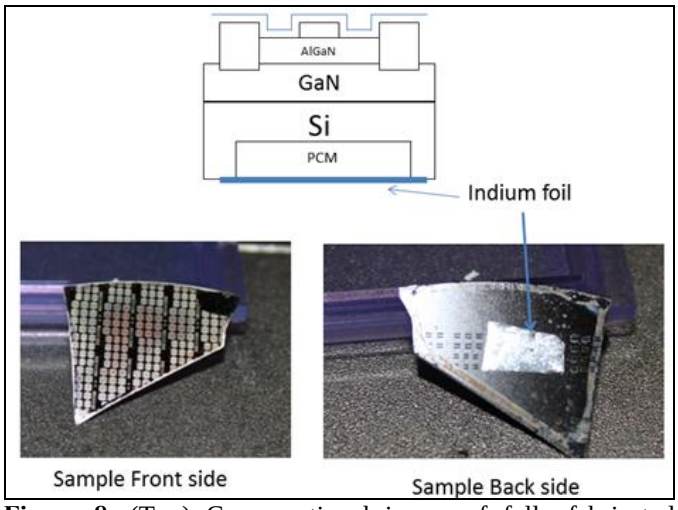

Figure 8: (Top) Cross-sectional image of fully fabricated AlGaN/GaN HEMT with integrated PCM, (bottom-left) Top-view image of a fabricated chip, (bottom-right) Bottom-view image of the fabricated chip showing the In foil encapsulation.

Electrical Testing

The impact of the PCM on the electrical properties and heat dissipation of the fabricated devices was tested through DC and pulsed current-voltage measurements. First, the similar values of breakdown voltage, V_{bk}, for devices with PCM filled grooves (127.5V) and devices without grooves (122.2V) confirmed that under low power conditions the groove fabrication process does not degrade the transistor performance.

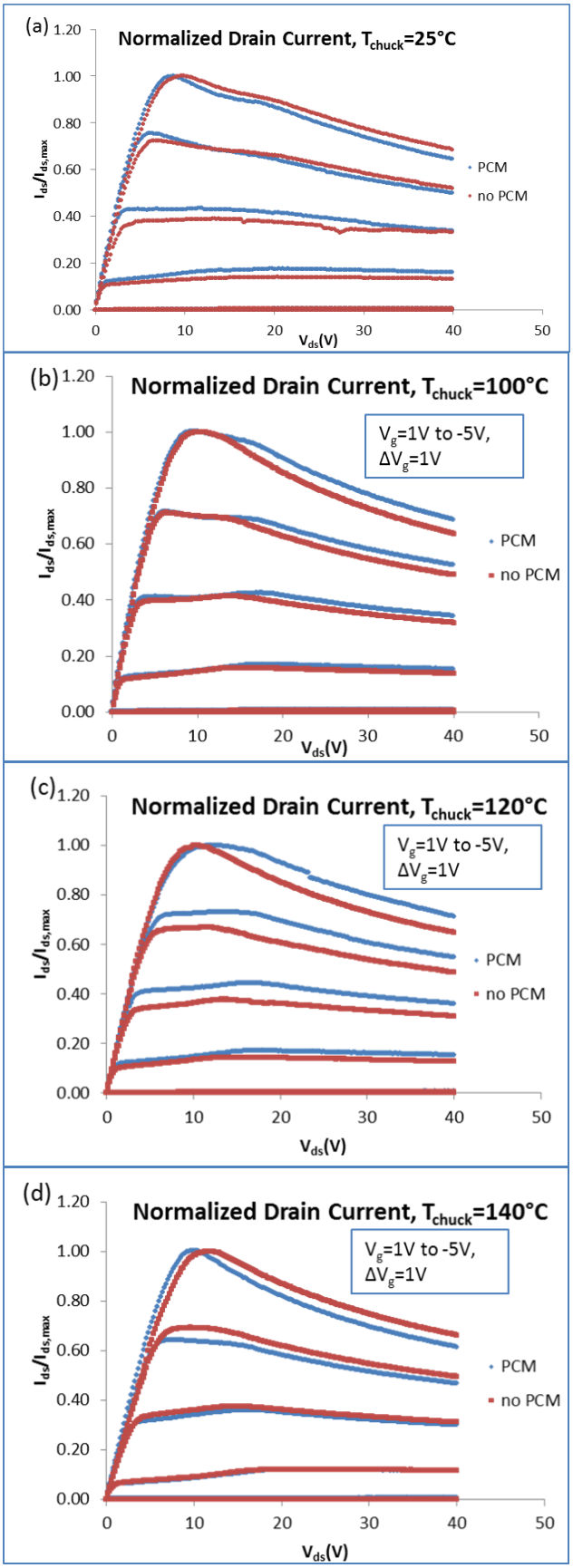

Figure 9: Normalized values of drain current ($I_{ds}/I_{ds,max}$) as a function of drain-to-source voltage in a transistor with PCM cooling solution (red) and in a conventional transistor (blue).

Figure 9 shows the DC current-voltage (I-V) characteristics of PCM filled grooved devices (blue) compared to devices without grooves (red). At room temperature, the normalized values of drain current ($I_{ds}/I_{ds,max}$) for the PCM filled devices are very similar to the devices without grooves. Performance of these devices at high temperatures was simulated by raising the chuck temperature. With increasing chuck temperature (100°C and 120°C), the difference between the red and blue lines also increased. This indicated that the difference in percent decrease of relative drain current between PCM and non PCM devices is larger at higher temperatures. The increase in gate voltage, V_g, results in increased self-heating and higher junction temperature. The current density in the PCM-enabled devices is higher than the value in the reference devices; this discrepancy is associated with the role of PCM inducing enhanced thermal management and reducing the self-heating. The role of PCM is confirmed when the chuck temperature is increased to 140°C which is well above the melting temperature of PCM. As expected, the I-V curve characteristics are similar to those at room temperature, because the melted PCM no longer stores heat.

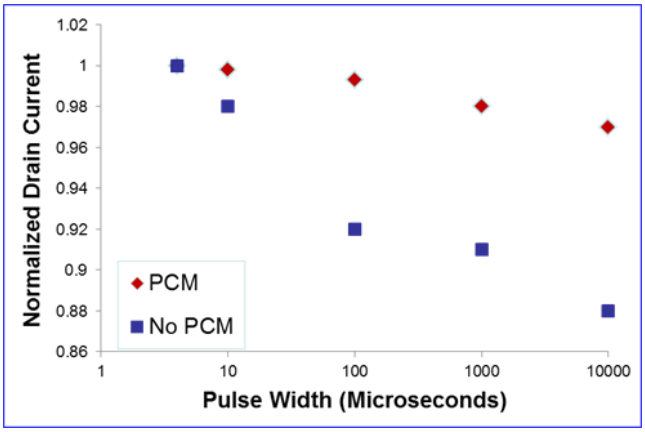

Figure 10: Normalized drain current in the fabricated devices as a function of the pulse width. The maximum drain current in the device with PCM is much more stable with pulse width than in the standard device.

Pulsed current-voltage (I-V) measurements were performed on the devices with PCM-filled grooves and devices without grooves. These measurements were performed with the chuck temperature set to 95°C to be able to operate the small devices that were fabricated under thermal conditions similar to what scaled devices encounter. The geometrical parameters of these devices were: gate-to-drain length, L_{gd}=3µm, gate length, L_g=2µm, gate-to-source length, L_{gs}=1.5µm.

At every pulse width, the PCM filled groove measurement was followed by no groove measurements, before going to the next pulse width. This procedure allowed for enough time (in the range of minutes) for the re-solidification of the melted PCM. Each data point in Figure 10 represents the maximum drain current at the respective pulse width divided by the overall maximum drain current for a single pulse. This measurement provides the current decrease with increasing pulse width and differences between PCM devices and

conventional devices. As the pulse width decreases, the maximum current through the device increases due to reduced device self-heating. As evident from Figure 10, the maximum drain current in the device with PCM is much more stable under pulsed conditions than in the standard device. The device performance in terms of current density shows 10% improvement due to enhanced thermal management. This result provides notable evidence that the PCM thermal storage effectively reduces the potential temperature increase in the device.

Summary & Conclusions

A novel micro-scale thermal storage design to remove the heat generated in the active regions of a pulsed mode semiconductor device was modeled, fabricated, and tested. Wetting angle measurements were performed on possible PCM candidates and erythritol was chosen as the primary candidate. Groove fabrication procedure was established successfully without any mechanical deterioration of the transistor device. The transistor performance under low power conditions, where the junction level heating is not enough to start the PCM melting process, was found to be identical for devices with and without PCM filled grooves. This confirmed that the PCM filling or groove making process does not affect the low power transistor operation. From the current-voltage characteristics at different chuck temperatures ranging from 25°C to 120°C, it was found that at higher temperatures, the current density in the PCM-enabled device was larger than in the reference device, due to the enhanced thermal management. The role of PCM was further confirmed when measurements at chuck temperatures well above the melting temperature of PCM did not show any signs of increase in the current density. The maximum current density (electrical performance) in the device with PCM material was found to be much more stable under pulsed conditions than in the standard device. The device electrical performance, in terms of current density, showed over 10% improvement due to enhanced thermal management. Additional work is needed to establish precise control on groove fabrication and groove dimension/location optimization for effective performance of the technology.

The optimization of the groove dimension is very critical. In the case of organic PCMs, if the quantity of PCM used is more than the required amount for thermal storage, then this additional PCM will not participate in thermal storage and rather act as a thermal insulator. Moreover, leakage of melted PCM in the devices is a concern. Hence, a detailed understanding on the role of PCM is needed via extensive characterization of the PCM filled devices in real application conditions to predict the true merit of the concept.

Acknowledgments

The work was supported by the NSF Small Business Innovation Research Grant (SBIR), under Award No. 1047111. We thank Dr. Jose Pascual-Guiterrez for his input during the project. It should be noted that Dr. Tapan Desai and Mr. Daniel Piedra contributed equally to this paper. A patent application (United States Patent Application 13/316,906) has been submitted based on this concept titled “Semiconductor having integrally-formed enhanced thermal management”.

References

[1] U. K. Mishra et.al., “GaN-Based RF Power Devices and Amplifiers”, Proc. of the IEEE vol. 96 ,no. 2, pp. 287-305, Feb. 2008

[2] Han Wang et.al., “High Performance InAlN/GaN HEMTs on SiC Substrate”, CS MANTECH Conference, Portland, Oregon, USA, 2010

[3] J. G. Felbinger et.al., “Comparison of GaN HEMTs on Diamond and SiC Substrates”, IEEE Electron Device Letters, vol. 28, pp. 948, 2007

[4] D. Francis et.al., “GaN-HEMT Epilayers on Diamond Substrates: Recent Progress”, CS MANTECH Conference, Austin, Texas, USA, 2007

[5] S. Boslet and R. M. Morales, “Thermal Modeling Approaches for GaAs Semiconductors”, Electronics Cooling, February 2008

[6] S. Lee, T.F. Lemczyk, and M. M. Yovanovich, “Analysis of Thermal Vias in High Density Interconnect Technology,” in 8th IEEE Semi-Therm Symposium, pp. 55-61, Feb. 1992

[7] R.S. Li, “Optimization of Thermal Via Design Parameters Based on an Analytical Thermal Resistance Model,” in Thermal and Thermomechanical Phenomena in Electronic Systems, pp. 475-480, 1998.

[8] M. P. Renavikar et.al., Intel Technology Journal, vol. 12, pp. 1, 2008

[9] J. Kim et.al., “Effect of oxidation on Indium Solderability”, Journal of Electronic Materials, vol. 37, pp. 483, 2008