Abstract—The peak power density of GaN high-electron-mobility transistor technology is limited by a hierarchy of thermal resistances from the junction to the ambient. Here, we explore the ultimate or fundamental cooling limits for junction-to-fluid cooling, which are enabled by advanced thermal management technologies, including GaN–diamond composites and nanoengineered heat sinks. Through continued attention to near-junction resistances and extreme flux convection heat sinks, heat fluxes beyond 300 kW/cm$^2$ from individual 2-µm gates and 10 kW/cm$^2$ from the transistor footprint will be feasible. The cooling technologies under discussion here are also applicable to thermal management of 2.5-D and 3-D logic circuits at lower heat fluxes.

Index Terms—Electronics cooling, Gallium Nitride (GaN), high-electron-mobility transistor (HEMT).

I. INTRODUCTION

RECENT progress on electronics thermal management has been motivated in part by the increasing power densities in microprocessors [1], power semiconductor lasers, and radar amplifiers [2], [3]. Gallium nitride high-electron-mobility transistor (GaN HEMT) technology has received special attention, with a focus on limits posed by thermal resistances at the device, substrate, package, and system levels [4]–[7]. Companies will select advanced HEMT cooling technologies for applications in aircraft, satellites, and land vehicles by balancing materials and manufacturing costs against improved performance. The complexity of this balance has overshadowed the recent extensions of the fundamental cooling limits at the device and package scale, i.e., the increasing peak power density that is becoming feasible independent of system level heat rejection and cost concerns.

We here examine the fundamental thermal resistance limits in two parts: 1) conduction/spreading in composite substrates (electron kinetic energy in the transistor is converted to phonons and then spread by conduction) and 2) fluidic convection in advanced heat sinks (from phonons to the working fluid). We consider the state of the art of both these technologies, and project advancements through this decade and beyond. This includes a review of recent progress on the translation of electron kinetic energy in GaN HEMTs to phonons in a high conductivity substrate [6], [7] (Section II), and then to fluid enthalpy through nanoengineered heat sinks (Section III). This trajectory of energy carrying mechanisms dictates the fundamental cooling limit, which may necessarily require a large heat rejection apparatus to the ambient air. The best combination of junction-level and heat-sink level cooling requires optimization considering spreading, and we explore this relationship using finite element simulations.

II. REVIEW OF COMPOSITE SUBSTRATE TECHNOLOGIES FOR HIGH-POWER GaN ELECTRONICS

Any effort to minimize thermal resistance must aggressively spread heat near the junction to minimize the junction temperature. Composite substrates made from diamond can offer higher local thermal conductivities by an order of magnitude compared with silicon (Si) and by a factor of nearly four compared with silicon carbide (SiC) [8]–[10]. However, the much larger lattice mismatch between GaN and diamond (11%) compared with that between GaN and SiC (3.5%) makes diamond integration very challenging. The existing techniques yield large defect concentrations or fully amorphous regions that impair heat conduction cooling of the junction.

References [11] and [12] used AlGaN/GaN heterostructures grown on single-crystal diamond substrates using molecular beam epitaxy (MBE) and metal–organic chemical vapor deposition (MOCVD) with the transition layers of thickness near 1 μm to mitigate the lattice mismatch. Epitaxial transfer [13] was used to attach the MOCVD-grown AlGaN/GaN to diamond using a disordered adhesion film thinner than 50 nm. The interface thermal resistance for this approach is governed by that of the disordered adhesion layer as well as the near-interfacial diamond [14].

Fig. 1 shows the mechanisms responsible for the GaN-substrate interface resistance: 1) phonon scattering at the transition layer boundaries with GaN and the substrate; 2) scattering on point defects, dislocations and other defects within the transition layer; and 3) scattering by near-interfacial disorder in the GaN and substrate. These contributions
are lumped and represented as a single effective interface resistance.

Table I and Fig. 2 summarize the interface thermal resistance data for GaN–Si, GaN–SiC, and GaN–diamond composites with theoretical lower limits at room temperature predicted by the diffuse mismatch model [20], [21]. Lateral and/or vertical temperature field data in device obtained with micro-Raman thermometry were used to extract MOCVD GaN–SiC and GaN–Si interface resistances [10], [15], yielding relatively large thermal resistances that varied rapidly with the temperature between 330 and 520 K. Micro-Raman on GaN–diamond device structures that were fabricated by epitaxial transfer [19] estimated the interface resistances of 27 m²·K·GW⁻¹ at 409 K [18] and 18 m²·K·GW⁻¹ at 443 K [19]. A transient interferometric mapping technique was used to estimate the interface resistances of 120 m²·K/GW (±50%) for MOCVD GaN–SiC [22] and 10 m²·K/GW for MBE-grown GaN on single crystalline diamond [17]. These approximate measurements did not extract the GaN thermal conductivity independently.

Time-domain thermoreflectance (TDTR) performed on samples with multiple GaN thicknesses was used to extract both the GaN thermal conductivity and the GaN-substrate thermal interface resistance simultaneously [9], [16], yielding lower resistances (with weaker temperature dependences) than those reported in [15]. Phonon transport and scattering arguments suggest that a combination of point defects within the AlN transition layer and near-interfacial defects around the AlN interfaces—with the adjacent GaN and SiC (or Si)—dominates the resistance [9]. TDTR on two GaN–AlGaN–diamond composites fabricated using epitaxial transfer [13], [19] yielded the interface resistances of 36–47 m²·K·GW⁻¹ at 300 K [14], which included the intrinsic resistance of the AlGaN transition layer. Through complete etching of the AlGaN transition layer, an interface resistance of 29 m²·K/GW was obtained at room temperature using TDTR on two GaN–diamond composites [23]. With continued fabrication effort, the GaN–diamond interface resistance may fall well below 1 m²·K/GW within the next five years, at which point there will be substantive advantages relative to GaN–SiC [23].

### III. Convection Resistances in Advanced Heat Sinks

The last three decades have yielded a dramatic increase in the power densities managed using compact heat exchangers and microheat sinks [29]–[31]. Tuckerman and Pease [32]...
demonstrated a microchannel heat sink with a thermal resistance of 0.9 cm²·K/W using liquid water. Much subsequent effort has examined the possibility of reduced flowrates and potentially higher heat transfer coefficients using convective boiling. Two-phase flow is promising for reducing pumping power requirements, but convection in long channels can introduce or exacerbate instabilities, flow regime oscillations, and dryout [33]–[35].

Substantial effort has been invested in advanced fluid routing and manifolding methodologies that decrease pressure drop for single-phase flow and promise to improve the stability for two-phase flows [27], [36]. The fluid routing challenge was addressed through tree-like or fractal-derived networks [37], [38] or through distributed chip-normal manifold [39]. Several patents feature distributed chip-normal fluid routing to a microfluidic heat exchange region [40], [41], which could take the form of either microchannels, distributed microscale pin fins, or microporous foam. All these studies reduce the axial length over which fluid must traverse the narrowest regions of the heat exchanger.

Nanoengineered structures and surfaces that can be integrated into liquid–vapor phase change systems have revolutionized the design of heat sinks. David et al. [27] demonstrated the vapor extraction through a hydrophobic nanoporous membrane considering both pressure drop and thermal resistance. Thin-layer evaporative cooling devices integrated with nanoporous alumina structures have achieved a heat removal up to 600 W/cm² [28]. Thin-layer nanoporous structures have been used in vapor venting channels in order to reduce the evaporative resistance across the vapor chamber [42]–[45]. Nanostructured surfaces may change the wetting characteristics and potentially enhance the transport phenomena at microscale/nanoscale. Superhydrophobic surfaces [46], hydrophobic–hydrophilic patterns [47], and the use of microengineered/nanoengineered features [48] can promote spontaneous bubble removal and assist heat removal from the surface. Continued attention to nanoengineered surface features is expected to continue the trends to higher heat fluxes and improved stability.

Table II summarizes the convection cooling performance for single- and two-phase flow. Single-phase cooling has been demonstrated up to 1 kW/cm² in microchannels [24] although higher levels are certainly feasible. Two-phase flow boiling has demonstrated a heat removal up to 400 W/cm² in microchannels [25]. Two-phase evaporative cooling has showed a heat removal up to 600 W/cm² integrated with nanoporous alumina membranes [28].

Sections II and III summarize the progress on reducing the thermal resistances due to the conduction in the substrate and convection into the working fluid. While progress on these two mechanisms has been achieved separately, the peak cooling performance relies on their combination and interactive optimization for a given lateral cooling area.
Fig. 3. (a) Gate dimensions used in the simulation that mimic CREE-CGH-60120D. The quarter symmetry region is modeled using symmetry boundary conditions. A constant heat flux of 200 kW/cm$^2$ to the gates gives a total power of 100 W to the system and 25 W to the quarter. (b) Example of the conduction simulation result using 3-D COMSOL Multiphysics. A sensitivity analysis was performed to determine the minimum tetrahedral mesh size of 1 µm with a number of 700,000 for the entire structure. The color map shows the temperature distribution of generation 3 with a power to the system of 160 W and a heat transfer coefficient of 150 kW/m$^2$·K.

IV. COMPUTATIONAL METHODOLOGY

This paper aims to determine the long-term heat flux feasibility of integrating advanced conduction materials and heat sink technologies. The challenge is that the optimal configurations will involve combinations of dimensions and geometry that are not necessarily the subject of this paper. To help with this effort, finite element calculations are performed (COMSOL Multiphysics) to account for the thermal resistances associated with the GaN and transition layer and conduction within the substrate (SiC or diamond), spreader, and heat exchanger. These calculations determine the temperature field by solving the steady-state heat conduction equation [49] based on the material properties and boundary conditions. All the simulation models are based on the gate dimensions of the CREE CGH-60120D chip, as shown in Fig. 3. We perform the simulation starting with the baseline that resembles contemporary practice and then, for future projections, extend the calculations to the designs that contain advanced materials and heat sink technologies. The calculations include copper fins (generation 1) or a microporous copper structure for generations 2 and 3. The most advanced technologies include GaN–diamond technology (generation 3).

The simulation models include multiple layers between the junction and the convecting fluid, as shown in Figs. 4(a) and 5. A representative system has the GaN HEMT on the device layer (SiC or diamond), interface material 1, spreading layer, interface material 2, and heat exchangers (i.e., a copper block, copper fins, or a porous copper structure). At boundary conditions, heat fluxes are given to the gates (2 µm × 160 µm × 40). A range of heat transfer coefficients associated with channel size and phase status is imposed to the fin walls. Adiabatic boundary conditions are applied to the quarter geometry to reduce the computational requirements. The simulation accounts for the material properties and device dimensions for each layer to solve the conduction equations. The thermal resistance of the transition layer between GaN–SiC and GaN–diamond composites is suggested as 20 m$^2$·K/GW, as discussed in Section II. The thermal conductivity of the interface material used for models is 60 W/m·K, which approximates that of a favorable solder bond. The thermal conductivities of SiC and GaN are expressed as the functions of temperature. Device dimensions are varied to optimize the junction-to-ambient thermal resistance. For example, a thicker heat spreader can increase the thermal resistance while improving the thermal spreading performance. To obtain a balance between these competing effects, a 100-µm thick spreader is used.

The details of the proposed designs are listed in Table III. The baseline design includes the GaN (1 µm), transition layer, SiC (100 µm), interface material 1 (15 µm), copper molybdenum (CuMo) spreader (100 µm), and interface material 2 (15 µm) that are integrated on a 1 cm × 1 cm × 0.5 cm copper block. A heat transfer coefficient $h$ of 700 W/m$^2$·K for air cooling removes a total power of 5 W when the junction temperature is maintained at 100 °C. Generation 1 utilizes a copper substrate (100 µm) with 200-µm copper fins. A single-phase heat transfer coefficient can be estimated using a Nusselt number equation [$h \sim \text{Nuk}/(w/2)$ for laminar flow]. This suggests $h$ of 2 and 6 kW/m$^2$·K for single- and two-phases, respectively, with 20-µm-width channels.
Generation 1 removes 60 and 100 W for single- and two-phase flow when the junction temperature is maintained at 100 °C. Generation 2 replaces the copper spreader and fins found in generation 1 with a conductive, microporous copper liquid delivery layer (LDL) and removes 90 and 140 W for single- and two-phase flow [50]. A microporous copper LDL is fabricated using the template-assisted electrodeposition technology. This technology should be addressed with a 3-D manifolding system to minimize the pressure drop due to the porous structure. We can obtain the 3× larger surface area to volume (SA/V) by replacing the fins with porous materials while the hydraulic diameters of channel and pore are same. Thus, high effective coefficients are estimated from the large SA/V of the microporous structure due to the areal effect. In order to estimate the performance of 20-μm pores, we have imposed the 2.5× higher effective high h of 50 and 150 kW/m²·K for single- and two-phase flow to the 20-μm-width fins developed for generation 2. Furthermore, the integration of the microporous copper LDL and nanoporous capping layer can significantly decrease the pumping power requirements by separating liquid and vapor phases. As a result, the integration of a thin, conductive, microporous copper LDL with diamond substrate in generation 3 promises to increase the maximum heat removal (in Fig. 4b). In this structure, GaN–diamond composites are used to raise the cooling limit by utilizing the high thermal conductivity of diamond. The combined technology of LDL and diamond spreading removes 100 and 160 W for single- and two-phase flow. A heat input of 160 W is equivalent to 300 kW/cm² to the footprint. This heat flux can be calculated as 6.3 W/mm, which allows the output power level to exceed 15 W/mm of gate periphery with 60% performance efficiency. Our simulation models show the potential effects to decrease the total thermal resistance and enhance the thermal performance compared with the baseline model.

We investigate the contributions of the various components to the total temperature rise in Fig. 6. We repeat
TABLE III
SUMMARY OF THE PROPOSED ARCHITECTURES

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Material</th>
<th>Thickness</th>
<th>Thickness</th>
<th>Thickness</th>
<th>Thickness</th>
<th>Thermal Property</th>
</tr>
</thead>
<tbody>
<tr>
<td>HEMT Transition Device</td>
<td>GaN</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>GaN(T)</td>
</tr>
<tr>
<td>Interface 1</td>
<td>SiC</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>SiC(T)</td>
</tr>
<tr>
<td>Interface 2</td>
<td>CuMo</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>15</td>
<td>60</td>
</tr>
<tr>
<td>Exchanger</td>
<td>Cu</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td>270</td>
</tr>
<tr>
<td></td>
<td>Diamond</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>400</td>
</tr>
<tr>
<td></td>
<td>Cu block</td>
<td>15</td>
<td>20000</td>
<td>200</td>
<td>400</td>
<td>60</td>
</tr>
<tr>
<td></td>
<td>Porous Cu (20 µm)</td>
<td>200</td>
<td>200</td>
<td>400</td>
<td>400</td>
<td></td>
</tr>
</tbody>
</table>

*Transition layer has the 20 m²K/GW thermal resistance.
*GaN(T) = -0.1623 T + 214.17, T (in K).
*SiC(T) = 0.0038 T² - 4.1734 T + 1259, T (in K) < 600 K.

Fig. 6. COMSOL Multiphysics solid conduction simulations for structures of generation 1 (black solid line), generation 2 (blue dashed line), and generation 3 (red solid line). Temperature at the y-axis indicates the temperature along the centerline below the center gate when we impose the 100-W system power and 50-kW/m²·K heat transfer coefficient. The contributions of the various components to the total temperature rise are plotted, which are decided by material properties and dimensions. For example, the increase in thermal conductivity will proportionally decrease the thermal resistance, resulting in the temperature rise of each component. The increase in thickness will proportionally increase the thermal resistance.

Single-phase convection with high heat transfer coefficients requires higher pumping powers. However, two-phase convection within porous structure remains a challenge for basic research, for example, through micromachining and other approaches to stabilize the flow and manage the vapor phase. The use of two-phase flow aims to achieve improved heat transfer performance by obtaining a stable, ultrathin evaporating liquid film at the walls of the heat exchanger that remains wetted even at extreme heat fluxes. Also, the use of the nanoporous capping layer aims to decrease the overall pressure drop by separating the vapor phase. Therefore, two-phase heat exchanges can operate at a higher coefficient of performance (COP), defined as the ratio between the heat flux and pumping power. Since it is important to compare COP to determine the performance of heat exchangers, the details of pressure drop, COP calculation associated with pumping power should be investigated carefully in the future.

V. CONCLUSION

This paper explores the limits of high heat flux cooling technologies, which combine composite substrates based on diamond with micromachined heat sink technologies and project advancements that will facilitate GaN HEMT cooling in the next five years. GaN–diamond composites will enhance the near-junction cooling owing to recent reductions in the GaN–diamond interface resistance and improvement in the quality of near-interfacial diamond. Microstructured and nanostructured heat sinks will improve convection resistance by leveraging a variety of technologies including a microporous/nanoporous phase change media. The rapid development of both conduction and convection technologies will raise the cooling limits of GaN HEMT devices to 100 kW/cm² within the next few years and the eventual possibility of 300+ kW/cm² to the gates with 2-µm length.
The recent focus on so-called 3-D packaging to leverage CMOS performance has inherent limitation when stacking high-power devices such as logic and memory. Some of the proposed cooling technology may also be eventually applicable to both 2.5-D as well as 3-D packaging. Realization of these extreme levels of performance will rely on progress in ambient heat rejection, which is highly application specific and not addressed here. The novel solutions studied here pose challenges for mechanical reliability, which will need to be addressed to avoid mechanical failures.

REFERENCES


Yoonjin Won received the B.S. degree in mechanical engineering and the B.B.A. degree from Seoul National University, Seoul, Korea, in 2005, and the M.S. and Ph.D. degrees in mechanical engineering from Stanford University, Stanford, CA, USA, in 2007 and 2011, respectively.

She is currently a Research Scientist with the NanoHeat Laboratory, Stanford University. She will be an Assistant Professor with the Department of Mechanical and Aerospace Engineering, the University of California at Irvine, Irvine, CA, USA, in 2015. Her primary research is in the characterization and design of nanostructured materials for energy conversion and electronics cooling technology. She is studying the two-phase heat transport by separating phases on nanoengineered structures and surfaces.

Jungwan Cho received the B.S. degree in mechanical engineering from Seoul National University, Seoul, Korea, in 2008, and the M.S. and Ph.D. degrees in mechanical engineering from Stanford University, Stanford, CA, USA, in 2010 and 2014, respectively.

He is currently a Post-Doctoral Scholar with Prof. Goodson. His current research interests include understanding nanoscale transport phenomena in thin films and at interfaces between various materials. He is currently involved in characterizing thermal properties of AlGaN/GaN high electron mobility transistors, using nanosecond and picosecond thermoreflectance techniques.

Mr. Cho is a recipient of Kwanjeong Educational Foundation Fellowship and Samsung Scholarship to support the M.S. and Ph.D. studies, respectively.

Kenneth E. Goodson is the Robert Bosch Chairmanship of the Mechanical Engineering Department and the Davies Family Provostial Professor at Stanford. He is a heat transfer specialist with interests ranging from electronics cooling to vehicle waste heat recovery. Goodson brings fundamental science to applications in heat management and energy conversion. His lab pioneered phonon free path measurements using silicon nanolayers and has highly-cited papers on diamond, carbon nanotubes, phase change memory, and two-phase microfluidics.

He co-founded Cooligy, which developed heat sinks for the Apple G5 and was acquired by Emerson in 2006. Goodson’s 40 PhD alumni include 17 professors at Stanford, MIT, UC Berkeley, and other schools. He is a Fellow with ASME, IEEE, APS, and AAAS. Recognition includes the Kraus Medal, the Heat Transfer Memorial Award for Science, the SRC Technical Excellence Award, the Thermo Award, and plenary lectures at IThERM, InterPack, Therminic, SemiTherm and PHONONS. Goodson studied at MIT and worked with the materials group at Daimler Benz.

Damen Agonafer received the Ph.D. degree in mechanical science and engineering from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2012.

He is currently involved in phase separation for low surface energy liquids utilizing engineered porous media for extreme heat transfer applications. In addition to his main focus in nanoscale-interfacial transport, he has also worked in the design of efficient micrototal analysis systems (µTAS). His current research interests include the areas of nanofluidics and nanoscale interfacial transport phenomena, which have several applications including chemical sensing and two-phase cooling technology. Dr. Agonafer was the recipient of the Alfred P. Sloan Fellowship Award.