Journal of Applied Physics | ARTICLE | scitation.org/journal/jap

Thermal management strategies for gallium oxide vertical trench-fin MOSFETs

Cite as: J. Appl. Phys. **129**, 085301 (2021); doi: 10.1063/5.0033001
Submitted: 12 October 2020 · Accepted: 22 January 2021 ·
Published Online: 23 February 2021

View Online

Export Citation

CrossMark

Robert H. Montgomery,[1] Yuewei Zhang,[2] Chao Yuan,[1] Samuel Kim,[1] Jingjing Shi,[1] Takeki Itoh,[2] Akhil Mauze,[2] Satish Kumar,[1] James Speck,[2] and Samuel Graham[1,3,a)]

AFFILIATIONS

[1]George W. Woodruff School of Mechanical Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA
[2]Materials Department, University of California, Santa Barbara, Santa Barbara, California 93106, USA
[3]School of Materials Science and Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA

[a)]**Author to whom correspondence should be addressed:** sgraham@gatech.edu

ABSTRACT

Trench-fin MOSFETs, with their near-surface heat generation and the higher-surface area afforded by their geometry for thermal management, represent a promising solution to the thermal problems frequently encountered in lateral β-Ga_2O_3 devices. Here, we investigate potential thermal-management strategies for a vertical β-Ga_2O_3 trench-fin MOSFET through parametric analysis, offering recommendations on how best to design a device for maximal current density and excellent thermal performance. Primarily, by using a thermally conductive dielectric over the MOSFET structure, significant improvements to device power density may be achieved, aided by thermal spreading. Additionally, we find that by bonding thermal spreaders to its topside can yield significant improvements in thermal performance.

Published under license by AIP Publishing. https://doi.org/10.1063/5.0033001

I. INTRODUCTION

β-Gallium oxide (β-Ga_2O_3), an ultra-wide bandgap semiconducting material, has recently received a great deal of attention for its predicted high-performance metrics[1,2] and theoretical low-cost of production. β-Ga_2O_3 has a bandgap of 4.8 eV, leading to a high breakdown field of 8 MV/cm[3] theoretically. β-Ga_2O_3 offers lower on-resistance for a given blocking voltage than silicon carbide (SiC) or gallium nitride (GaN), the current industry standards. However, β-Ga_2O_3 suffers from a critical drawback—it possesses low and anisotropic thermal conductivity (10.7 $Wm^{-1}K^{-1}$, 23.4 $Wm^{-1}K^{-1}$, and 13.7 in the [100], [010], and [001] directions,[4] an order of magnitude smaller than that of GaN[5]). Much work has been done on the electrical characterization and integration of β-Ga_2O_3 into lateral MESFETs (metal–semiconductor field-effect transistors),[1,6] lateral MOSFETs (metal–oxide–semiconductor field-effect transistors),[7–9] and most appealingly, the vertical MOSFET configuration, which offers higher current densities, greater breakdown voltages, and lower drift resistances.[10,11] The thermal management of lateral devices has been studied to some extent.[10–19] However, little work has been done on understanding the effect of thermal environment on peak temperature of vertical devices, and no guidelines for best-practices in the implementation of thermal management for vertical β-Ga_2O_3 devices are available in the literature. Vertical MOSFETs offer superior thermal management,[20,21] as the volumetric nature of heat generation in their electronic design, coupled with the six-sides provided by their rectangular-prism shape for thermal management, means that there are more options and area available for heat dissipation. This work investigates the thermal management of a vertical β-Ga_2O_3 trench-fin MOSFET (also known as a FinFET) as a general example for vertical β-Ga_2O_3 devices, seeking to set out guidelines on how to successfully implement efficient thermal management when designing these devices.

In 2017, the first vertical FETs fabricated from β-Ga_2O_3 were demonstrated by Wong *et al.*,[22] who fabricated the devices in a vertical CAVET (current aperture vertical electron transistor) configuration. That same year, Sasaki *et al.*[23] fabricated the first β-Ga_2O_3 trench MOSFETs, utilizing a combination of n− and n+ doping to offer improved on-state resistance. Kotecha *et al.*[24] built off the vertical trench-fin MOSFET architecture designed to offer similar improvement in gallium nitride[25,26] and fabricated a vertical trench-fin MOSFET from β-Ga_2O_3, offering both improved on-state resistance and helping to circumvent the inability of β-Ga_2O_3 to accept p-type doping. Through a combination of an optimized geometry, high breakdown voltage, superior affordability,

Published under license by AIP Publishing.

and enhanced thermal management, β-Ga_2O_3 devices are ideally suited for applications in energy storage and transformation, offering innovations in high frequency motor controllers, high frequency inverters, and power-grid transmission systems.[27] Vertical β-Ga_2O_3 offers compelling reasons for its research and adoption; however, there are some drawbacks that must be overcome prior to commercialization of this material in power electronics. Placing the thermal-management solution closer to the heat generation region leads to marked reductions in thermal resistance and, therefore, device temperature, as shown in Fig. 1, where the heat in the lateral device on the left must pass through 1–100 μm of β-Ga_2O_3 prior to reaching the heatsink; for vertical devices, the heat can be pulled out laterally or through a thinner (in comparison with lateral) drift region.

As previously mentioned, β-Ga_2O_3 possesses low and anisotropic thermal conductivity, leading to the large intrinsic thermal resistance for devices composed of it. This large thermal resistance, in part, leaves β-Ga_2O_3 devices susceptible to heat-induced reliability problems.[28] Effective thermal management has thus become crucial. To better understand the need for thermal management, Kumar *et al.*[10,11] used the TCAD simulations of Choi *et al.*[29] to map the Joule heating present within lateral β-Ga_2O_3 devices. Oh *et al.*[19] utilized Silvaco Atlas to investigate lateral β-Ga_2O_3-on-diamond architectures and characterize the self-heating effects. Thermal-management solutions, such as the integration of Ga_2O_3 thin-membrane onto high thermal conductivity substrates,[14–18] have been successfully applied for β-Ga_2O_3 lateral devices. Through Finite Element Analysis (FEA),[12,13] it was proposed that additional thermal-management approaches for lateral devices, such as thinning Ga_2O_3, flip-chip bonding, and overgrowth of a high thermal conductivity dielectric atop the epitaxial layer for heat spreading could be effective. Cheng *et al.* recently utilized techniques to improve the thermal boundary conductivity (TBC) between β-Ga_2O_3 and high thermal conductivity substrates through modern bonding methods.[30] Figure 1 offers a helpful reference to understand why some of these techniques, which are designed for the planar geometry of lateral device, will be difficult to incorporate into the complex 3D design of vertical devices. The additional complexity of vertical devices, compared to lateral, brings additional challenges in understanding how the various device components play a role in thermal resistance. Vertical devices offer more surface area near the drift region that is available for thermal management. Placing the thermal-management solution closer to the heat generation region may lead to marked reductions in thermal resistance and, therefore, device temperature.

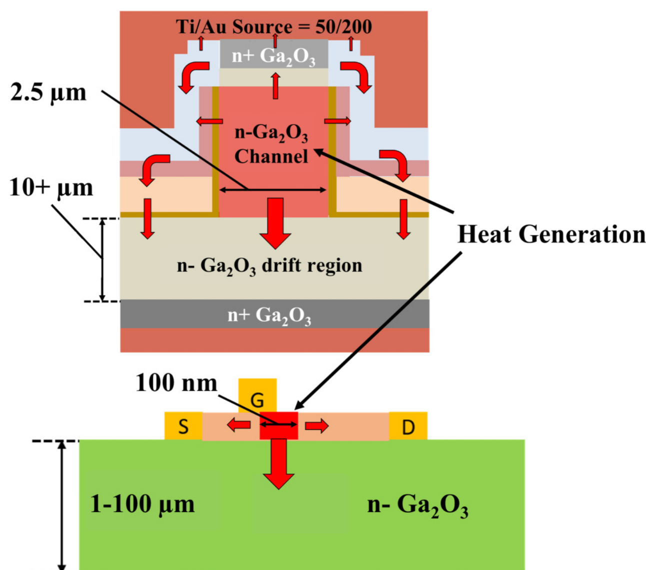


FIG. 1. Vertical vs lateral device length scales—red rectangles are regions of heat generation, whereas red arrows represent paths for heat to leave these regions and are sized proportionate to how much heat may leave the device through that direction.

This work proposed three typical cooling strategies for vertical β-Ga_2O_3 trench-fin MOSFET devices, as seen in Fig. 2. The first of these was bottom-sided cooling, in which the die attach, dielectric spacer, and heat spreader were placed beneath the drain electrode of the device—this schema is pictured in Fig. 2(a). Top-side cooling placed the die attach, dielectric spacer, and heat spreader atop the device's source electrode, as seen in Fig. 2(b), a configuration designed to pull heat out of the device more efficiently. In this configuration, heat has a shorter and more thermally conductive path out of the device. Double-sided cooling was the final geometric configuration utilized for these experiments and used two die attaches, dielectric spacers, and heat spreaders, one on each side of the device—this configuration is pictured in Fig. 2(c). For all configurations, the thickness of the die attach was set to 50 μm, whereas that of the heat spreader was set to 0.5 mm—whereas this represented a fairly aggressive cooling setup, the goal was to understand the trends in thermal management and improvements that might be realized rather than discern concrete knowledge of exact performance specifications. To investigate the ramifications that cooling strategies have on the performance of gallium oxide devices, we have developed a finite element model within COMSOL Multiphysics[31] suite. Using this model, we have performed comprehensive parametric studies of the device to determine which components of the thermal-management system had the highest impact on device performance. We sought to develop guidelines and recommendations on device thermal-management through this work by examining trends across various thermal-management techniques. Through the coordination of these thermal-management strategies, we show that a single vertical device with our architecture in an optimized thermal-management configuration is able to dissipate roughly four times more power than an unoptimized device of the same architecture without exceeding the 200 °C metric, which has been established as the safe operating temperature.[13,32] We finally show that by adding multiple devices, the total maximum power output may be increased by roughly three times without incurring large increases in the maximum channel temperature—likely even more when a combination of additional fins and thermal crosstalk minimizing strategies are implemented. Through this, we hope to inform those seeking to design vertical β-Ga_2O_3 power electronics as to their material choice—how a better thermal-management solution results in higher power designs. We hope to accelerate the

Published under license by AIP Publishing.

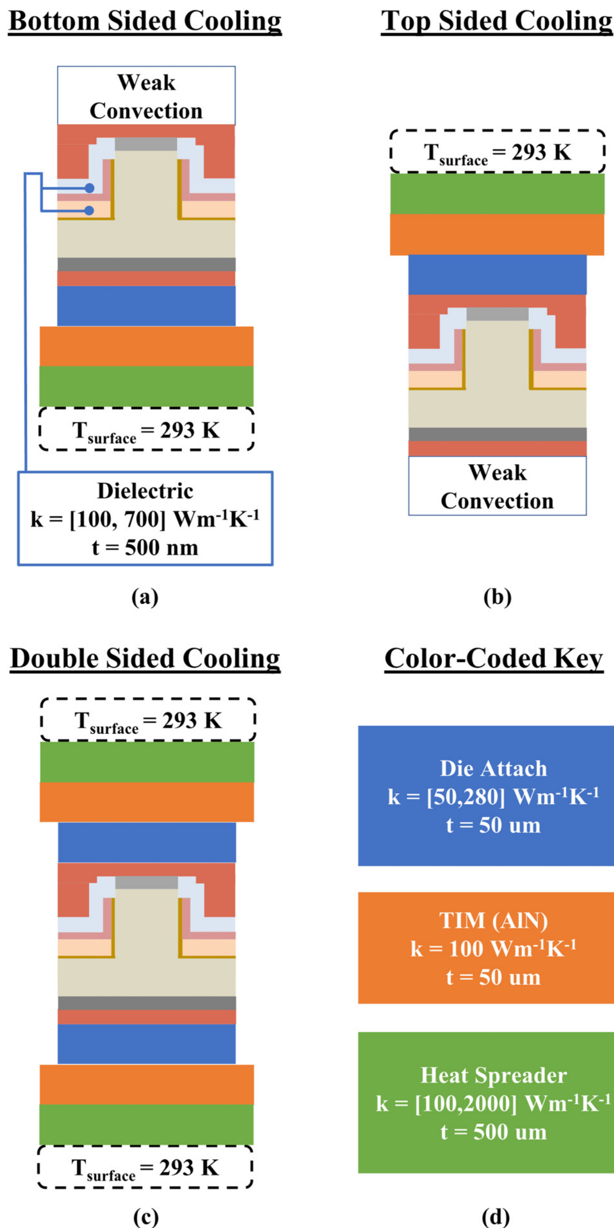


FIG. 2. Cooling configurations used in experiments. (a) Bottom-sided cooling configuration, placing thermal management at the bottom of the die, with convective cooling at the top of the die; (b) a top-sided cooling configuration, placing thermal management at top, with convective cooling at the bottom of the die; and (c) a double-sided cooling configuration, which places thermal management at both the top and bottom of the die, leaving convective cooling only on the sides of the die.

development of β-Ga_2O_3 by setting out thermal-management strategies than can circumvent the conventional drawbacks of the material and enable market-ready vertical β-Ga_2O_3 devices ahead of schedule.

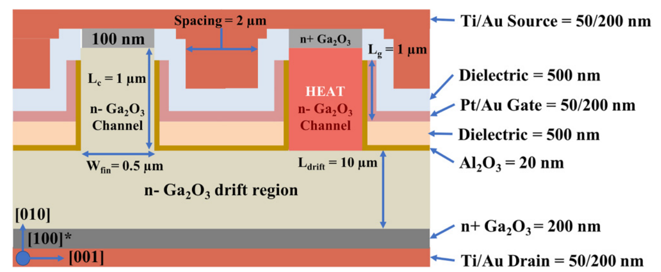


FIG. 3. Layout of materials and thicknesses within a two-fin device, depicting one region of heat generation with a red square on right-hand fin.

II. METHODOLOGY

A vertical trench-fin MOSFET, the geometry of which is detailed in Fig. 3, was constructed within the COMSOL Multiphysics environment as a quarter-symmetry model. For each material within the device, the thermal conductivity, density, and specific heat of each material were specified—materials that were varied are specified in Table I. Titanium, gold, and platinum were used for the contacts (arranged in layers), and β-Ga_2O_3 was used for the channel and drift regions. Given the high anisotropy of β-Ga_2O_3, temperature-dependent thermal conductivities, measured by Guo *et al.*,[4] were used for all principal crystalline directions (Fig. 4). The device was split into four quadrants to reduce computational expense, solving the equation with respect to a quarter-symmetry model—this is depicted graphically in Fig. 5(a). Out of the multiple regions that comprised the device, three regions had their thermal conductivities selected as free variables for use in parametric variation. These three regions were the dielectric layers included in the semiconductor stack itself, the die attach soldering the device to the thermal-interface material (TIM) and heat spreader, and the heat spreader itself; these regions are labeled i, ii, and iv in Fig. 5(b), respectively. The dielectric region is shown in detail in Fig. 5(c), where it is labeled i. The ranges of thermal conductivities used are listed in Table I. Values were also chosen for baseline parameters; the baseline thermal conductivity of the dielectric was selected to match that of polycrystalline aluminum nitride (AlN)[33] ($k = 100\ Wm^{-1}K^{-1}$); that of the die attach, to match that of sintered Ag ($k = 280\ Wm^{-1}K^{-1}$); finally, the

TABLE I. Thermal conductivities considered.

	$k_{heat\ spreader}$ ($Wm^{-1}K^{-1}$)	$k_{dielectric}$ ($Wm^{-1}K^{-1}$)	$k_{die\ attach}$ ($Wm^{-1}K^{-1}$)
Range	100–2000	10–700	50–280
Baseline	400	100	280
Worst case	100	10	50
Best case	2000	700	280
Materials considered	Copper	AlN	AuSn
	Diamond	Diamond	SnAuCu (SAC)
	AlN	Al_2O_3	sintered Ag

Published under license by AIP Publishing.

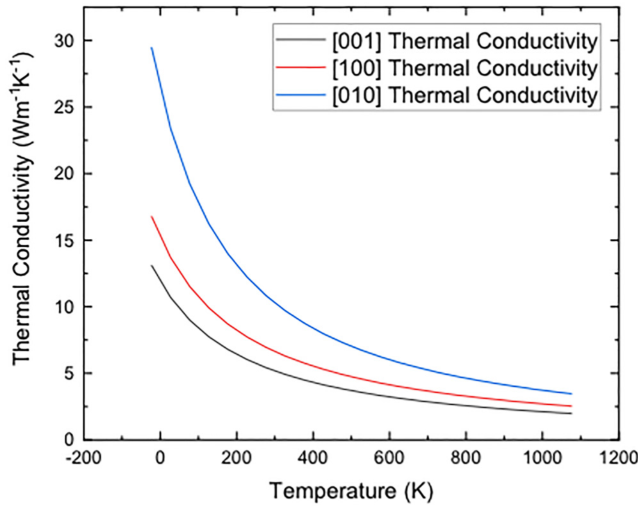


FIG. 4. Temperature-dependent thermal conductivity curves for β-Ga_2O_3.

thermal conductivity of the heat spreader was matched to copper (k = 400 $Wm^{-1}K^{-1}$) for the baseline case. Region iii in Fig. 5(b) is the thermal-interface material (TIM) containing the traces and the circuit layer—it is composed of polycrystalline AlN (k = 100 $Wm^{-1}K^{-1}$) and is 50 μm thick; the thermal conductivity of this region was held as constant. For each cooling configuration, best and worst-case scenarios were also designated, and the values for all of these possible combinations are tabulated in Table I.

The COMSOL Heat Transfer module was implemented to solve the heat equation (1) at a steady state. The inner surfaces with the aforementioned quarter symmetry had adiabatic boundary conditions applied to them according to (2), where "n" is the normal vector to any given surface. The horizontal exposed surface of each (if more than one was present) heat spreader was subjected to an isothermal boundary condition according to (3). All remaining external boundaries were subjected to convection boundary conditions, with the air temperature far away from the chip (T_∞) assumed to be 20 °C and the convection coefficient h = 5 $Wm^{-2}K^{-1}$ chosen to simulate natural convection (4),

$$\nabla \cdot (k\nabla T) + \dot{q} = 0, \tag{1}$$

$$\frac{\partial T}{\partial n}(x, y, z) = 0, \tag{2}$$

$$T(x, y, z) = T_{surface} = 293\,\text{K}, \tag{3}$$

$$h\,(T(x, y, z) - T_\infty) = -k\frac{\partial T}{\partial n}(x, y, z). \tag{4}$$

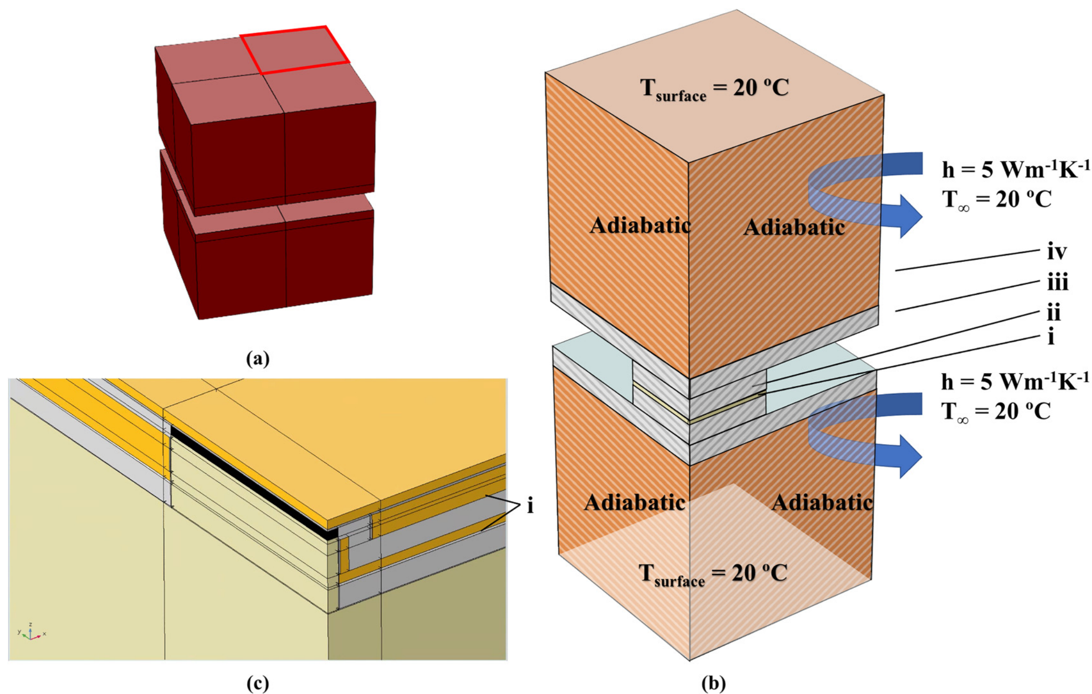


FIG. 5. Geometric layout of device. (a) Depiction of what a full device looks like, with slice taken for a quarter-symmetry model outlined in red, (b) model of a device with heat spreader, die attach, and die in a double-sided configuration, and (c) close-up of fin within the COMSOL model in a bottom-sided configuration.

Published under license by AIP Publishing.

Utilizing the Silvaco Atlas electrical simulation suite, a Joule heating map was obtained for the device (the right side of Fig. 5). Physically, this map represented the heat dissipated due to the on-state resistance of the device; Joule heating occurs whenever heat passes through a conductor with finite resistance. Even in the on-state, transistors possess a not-negligible resistance, known as the on-state resistance, which leads to heat generation when current flows through the device. Whereas another physical mechanism for heat generation exists in switching-losses, where power is dissipated as the device's state is changed, these steady-state simulations did not consider that mechanism due to its transient nature. Using the map of this heat generation as a generalized guide to heating trends within the device, the channel region was chosen to act as a volumetric heat source within the device. The power was determined through (5) and distributed over the entire volume of the channel region. As previously mentioned, 200 °C was chosen as the safe operating temperature,[13,32] above which device failure would be anticipated,

$$\text{Power} = \text{Cross Sectional Power}\left(\frac{\text{MW}}{\text{cm}^2}\right) \times A_{\text{cross-section}}. \quad (5)$$

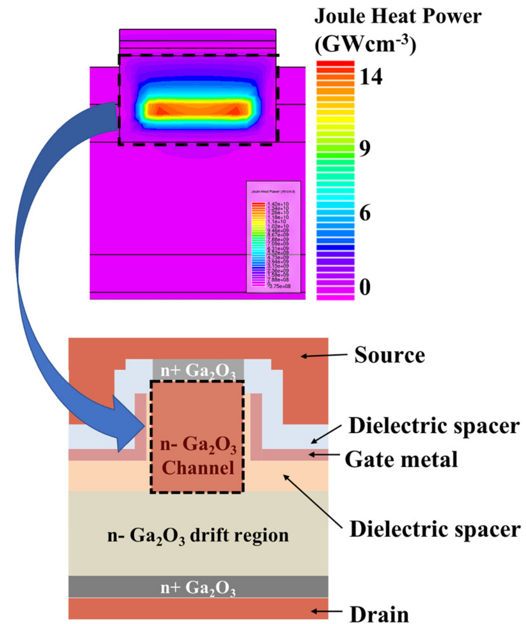


FIG. 6. Heat generation within the Silvaco ATLAS simulation was mapped to a specified volumetric region of the model.

III. RESULTS AND DISCUSSION

A. Bottom-sided cooling

To evaluate the priority of what regions within the device were most in need of thermal optimization in each geometric configuration, three standard sweeps were established, varying the thermal conductivity of the dielectric, the heat spreader, and the die attach between the values listed in Table I. Figure 7(a) shows the channel maximum T (T_{max}) as a function of the heat spreader thermal conductivity. Varying the thermal conductivity of the heat spreader produced a stronger effect on T_{max} in devices arranged for bottom-sided cooling, as all of the heat generated by these devices had to leave solely through the heat spreader—that said, the effect was still barely noticeable, even in this configuration (from 224 °C to 222 °C at 15 MWcm^{-2}). Considering that even less significant trends were observed in correlation with heat spreader thermal conductivity in all other configurations (3 °C difference for top-sided cooling and 1 °C for double-sided cooling, both at 15 MWcm^{-2}), these results have been omitted from all further studies under consideration. Also seen in Fig. 6(a), gallium oxide's poor through-plane thermal conductivity makes it impossible to run baseline devices at 15 MWcm^{-2} while remaining below safe operating temperatures (200 °C), as all heat had to pass through the β-Ga_2O_3 drift region to leave the device. The thermal bottleneck of this β-Ga_2O_3 is further underscored by the sweep over the die attach thermal conductivity [Fig. 7(b)], which shows very little improvement in device performance (from 230 °C to 223 °C for 15 MWcm^{-2}), even when quadrupling the thermal conductivity of this layer. Again, this lack of improvement is primarily due to the large thermal resistance of β-Ga_2O_3 compared to the die attach and the heat spreader. When placed in series, the thermal resistance of β-Ga_2O_3 dominates. All of this is well-exemplified by Fig. 8(a), which shows the paths that heat can take as it leaves the device in a top-sided cooling configuration. The thermal bottleneck of the drift region is clearly detailed, as the vast majority of heat must ultimately pass through this layer. The dielectric sweep [Fig. 7(c)] indicates strongly that these high temperatures are a function of the device's geometry, showing that increasing dielectric thermal conductivity results in significantly reduced peak temperature in the device. As the dielectric is a lateral layer throughout the chip, this indicates that thermal spreading is the mechanism responsible for the increase in performance. As improving the dielectric conductivity provides a 57% decrease in T_{max} for bottom-sided cooling (328 °C–141 °C for 15 MWcm^{-2}), as related by (6), it is strongly recommended that any parties constrained to this thermal-management architecture opt for polycrystalline diamond, which has an ultrahigh thermal conductivity of roughly 700 $\text{Wm}^{-1}\,\text{K}^{-1}$,[34] as their dielectric of choice (cost permitting). Al_2O_3 offers the worst performance examined, with a thermal conductivity of roughly 10 $\text{Wm}^{-1}\,\text{K}^{-1}$—it offers acceptable, but not remarkable, performance at lower power levels. At higher power levels, the ability of Al_2O_3 to effectively remove heat is significantly diminished and the device exceeds the safe-temperature threshold of 200 °C. Of particular note should be that using a low-quality dielectric, such as SiO_2 ($k = 1.4\ \text{Wm}^{-1}\,\text{K}^{-1}$), at 15 MWcm^{-2} results in a T_{max} of 800 °C—thus, for high power operations, we recommend that SiO_2 be avoided for use as a dielectric in these devices. A sweep to establish acceptable levels for SiO_2 is presented further into this investigation,

$$\text{Percent Decrease} = \frac{T_{\text{Worst-case}} - T_{\text{Best-case}}}{T_{\text{Worst-case}}} \times 100\%\ (\text{all in } ^\circ\text{C}). \quad (6)$$

Published under license by AIP Publishing.

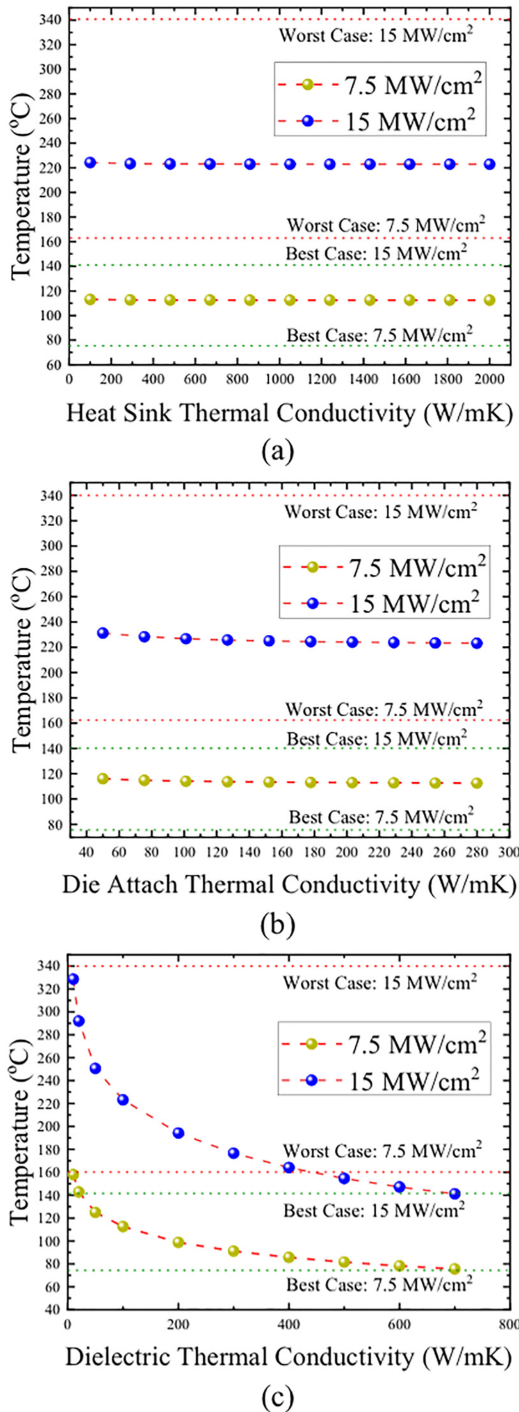


FIG. 7. Sweeps of thermal conductivity for a device with bottom-sided cooling, displaying the maximum temperature of the device vs the various thermal conductivities swept: (a) heat spreader sweep, (b) die attach sweep, and (c) dielectric sweep. Worst-case and best-case performance (using the values tabulated in Table I) are indicated with green and red lines at each power level used.

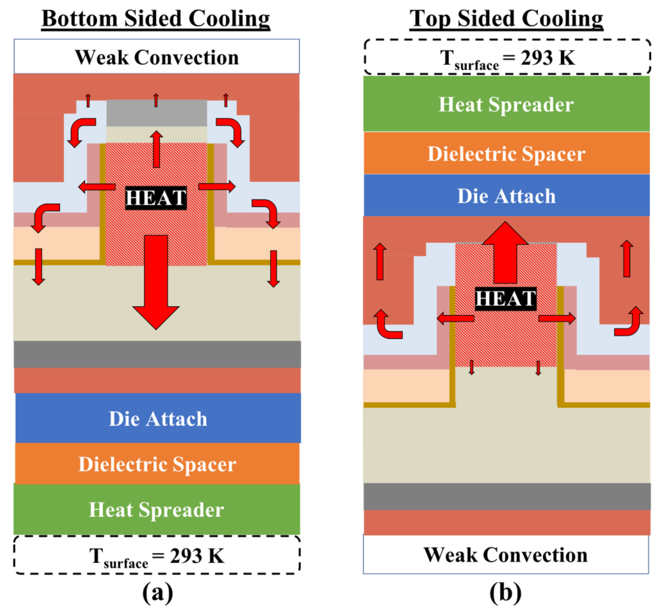
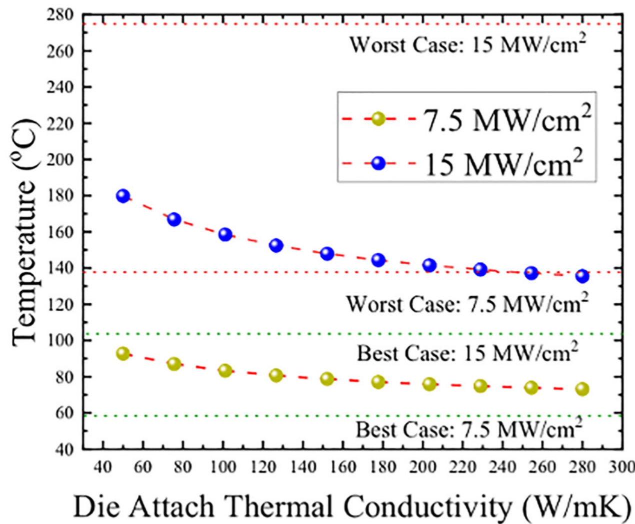


FIG. 8. Paths for heat to exit the device in (a) bottom-sided cooling and (b) top-sided cooling.

B. Top-sided cooling

Top-sided cooling showed far better performance at removing heat from the channel region than bottom-sided cooling did. As the die attach in this scheme was placed directly on top of the source, heat generated in the channel had only to pass through two thin-layers to reach the die attach, providing a much less thermally resistive path out of the device than bottom-sided cooling. This is detailed in Fig. 8(b), which shows how the majority of the heat can easily leave the device through the source electrode. As the heat source was far closer to the die attach in this model, varying the thermal conductivity [Fig. 9(a)] of that component proved effective at improving the device's performance. Moving from the worst possible thermal conductivity for the die attach (50 $Wm^{-1}K^{-1}$) to the best (280 $Wm^{-1}K^{-1}$) provided a 25% decrease in absolute device maximum temperature (from 179 °C to 135 °C at 15 $MWcm^{-2}$), a significant improvement in performance. This indicates that for devices implementing top-sided cooling, optimizing the thermal conductivity of the die attach should be a significant goal, as it becomes a thermal bottleneck for top-sided cooling. The dielectric layer [Fig. 9(b)] continues to play a significant role in heat spreading within this implementation, allowing for a 55% decrease in device temperatures between a low conductivity and a high conductivity dielectric (from 154 °C to 104 °C at 15 $MWcm^{-2}$). As observed previously, the dielectric allows for heat generated in the channel region to spread out across the entirety of the device before flowing out of the device—in this case, however, there is far less thermal resistance along the path for heat to leave the device, allowing for significantly lower temperatures with the same dielectrics (223 °C for baseline 15 $MWcm^{-2}$ bottom-sided cooling vs 135 °C for top-sided cooling).

Published under license by AIP Publishing.



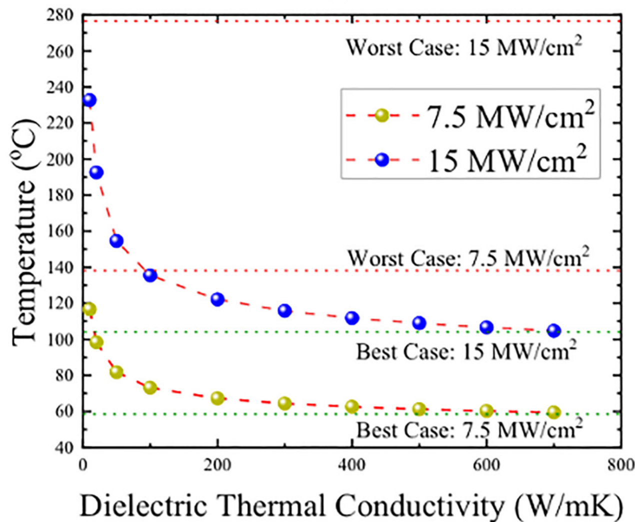
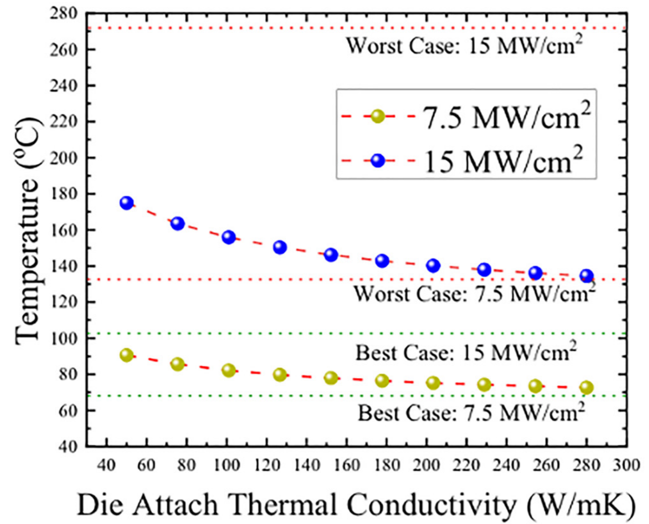


FIG. 9. Sweeps of thermal conductivity for a device with top-sided cooling vs the maximum temperatures achieved at each thermal conductivity. (a) Die attach sweep and (b) dielectric sweep. As discussed in Fig. 6, worst-case and best-case configurations are calculated from the values in Table I. Heat spreader sweeps are excluded as improving their thermal conductivity shows no significant improvements in device performance.



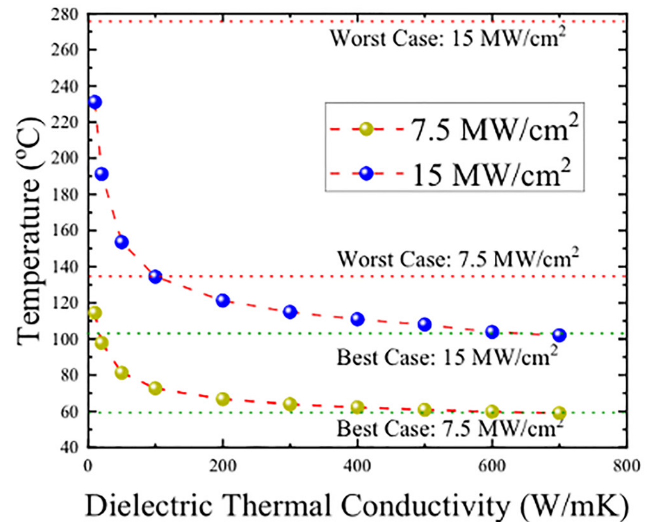


FIG. 10. Sweeps of thermal conductivity for a device with double-sided cooling vs the maximum temperatures achieved across the entire device at each set of conditions. (a) Die attach sweep and (b) dielectric sweep. As previously discussed, best-case and worst-case scenarios are designated with the red and green lines for each of the two power levels surveyed.

C. Double-sided cooling

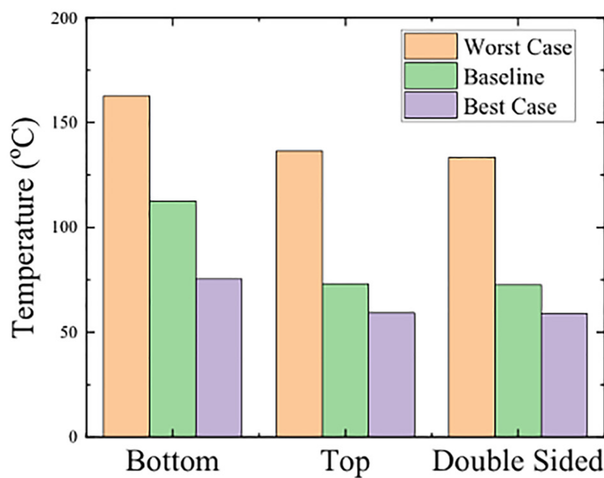
Double-sided cooling showed very good improvements over bottom-sided but was only marginally more effective than top-sided cooling, improving from 135 °C baseline at 15 $MWcm^{-2}$ for top-sided cooling to 134 °C for double-sided cooling, both thoroughly beating the 223 °C baseline for bottom-sided cooling. Sweeping the thermal conductivity of the die attach [Fig. 10(a)] produced a roughly 23% decrease in absolute temperature T_{max} within the device (from 175 °C to 134 °C at 15 $MWcm^{-2}$), leading to the conclusion that most of the heat generated in double-sided cooling is leaving via the same channels as that of top-sided cooling. This nearly identical performance continues when sweeping the dielectric layer

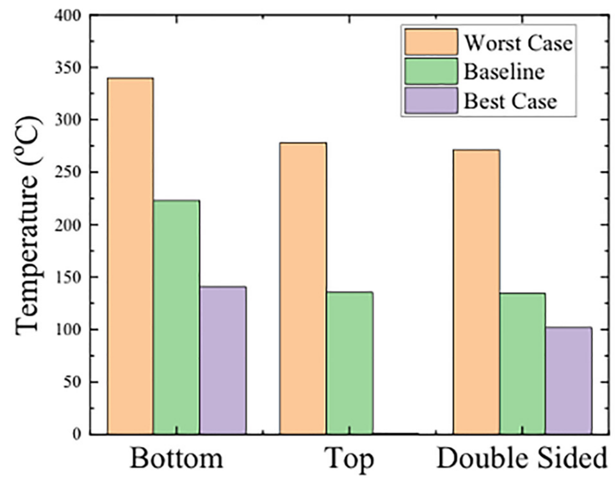
Published under license by AIP Publishing.

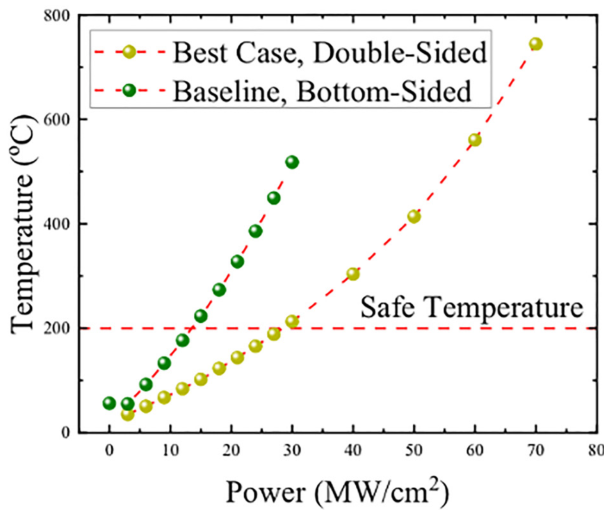
[Fig. 10(b)], which shows a 56% decrease in device temperature from worst to best (from 153 °C to 102 °C at 15 $MWcm^{-2}$). Whereas, clearly, some heat is leaving the device via the bottom heatsink, it is only enough to result in a few degrees Celsius improvement over top-sided cooling (at best-case conditions, we observe an improvement of 7 °C). However, as all of these trends have been derived under constant temperature boundary conditions, this alone is not enough to completely discount double-sided cooling.

D. Understanding the limits of power density

To further explore the efficiency of each cooling scheme, best and worst cases for each geometric configuration were established. The values of thermal conductivity were set to their lowest values for the worst-case scenario and their highest values for the best-case scenario, as previously designated in Table I. Comparisons of maximum channel temperatures for best, worst, and baseline cases for both 7.5 $MWcm^{-2}$ [Fig. 11(a)] and 15 $MWcm^{-2}$ [Fig. 11(b)] reinforce once again the utility of top-sided and double-sided







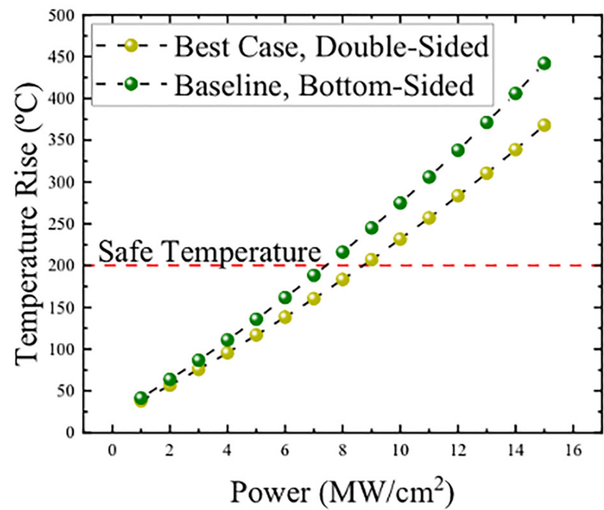


FIG. 11. Results of best/worst-case scenarios: (a) 7.5 $MWcm^{-2}$, (b) 15 $MWcm^{-2}$, (c) power sweeps to determine upper limits, and (d) power sweep with an SiO_2 dielectric. Temperatures plotted are the maximum values across the entire device. The best jump in performance is realized in making the switch from bottom-sided cooling to top-sided cooling.

Published under license by AIP Publishing.

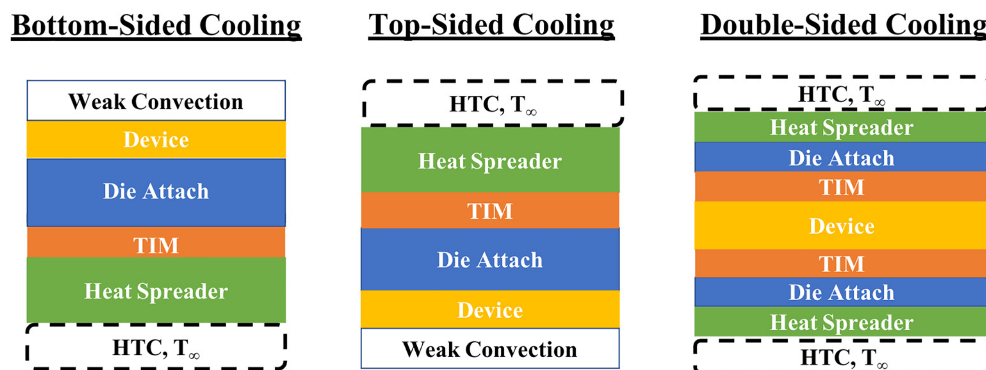


FIG. 12. Boundary conditions imposed on devices for forced convection cooling.

cooling over bottom-sided cooling. Across all three schemes of worst-case cooling, poor thermal spreading inhibits any gains made by adjusting the cooling configuration. However, when the thermal spreading is increased to both the average and the best-case scenarios, significant performance increases are realized by adjusting the cooling schema used. Interestingly, the best-case scenario for bottom-sided cooling is not significantly worse than that of top-sided and double-sided cooling (141 °C, 104 °C, and 101 °C, respectively); this is again due to the excellent thermal-spreading provided by a high-quality dielectric. Between the best case for double-sided and top-sided cooling, there is barely any advantage at all (101 °C and 104 °C, respectively); given the logistical difficulty in employing double-sided cooling, this seems to suggest that top-sided cooling is the best balance between economy and performance. With this best-case designated, the power being input to the devices was swept for the bottom-sided baseline and compared to a power sweep performed across the best-case double-sided [Fig. 11(c)]. With 200 °C established as a safe baseline temperature, the bottom-sided baseline was only able to safely operate at roughly 12.5 $MWcm^{-2}$—by comparison, double-sided cooling with best-case thermal conductivities was able to operate at roughly 30 $MWcm^{-2}$, achieving over double the performance when compared to an inefficiently cooled counterpart. To investigate the case in which SiO_2 was an acceptable dielectric material, the dielectric was set to SiO_2 and power levels were swept to establish safe-temperature margins for both best-case double-sided cooling and baseline bottom-sided cooling. The upper and lower bounds for power density were found to be 9 $MWcm^{-2}$ and 7.5 $MWcm^{-2}$, respectively, as shown through Fig. 11(d). If SiO_2 must be utilized in the device architecture, then thermally optimized configurations examined in this paper are unlikely to offer significant improvements in device performance. If additional fins are added in parallel, power density can be further increased beyond these limits on single-fin operation.

E. The effects of realistic cooling conditions

To investigate the heat transfer coefficients (HTCs) that would be needed to properly cool these devices, the constant temperature boundary conditions were replaced with convection boundary conditions in all three geometric configurations (Fig. 12). Heat transfer coefficients ranging logarithmically from 10^3 to 10^6 were considered (Fig. 13), and the fluid temperature was fixed at 25 °C. The ranges of forced convection for liquids were taken as 10^2–2×10^4 $Wm^{-2}K^{-1}$ and those of phase-change convection were taken as 10^4–10^5 $Wm^{-2}K^{-1}$. Sweeping these values for bottom-sided cooling [Fig. 14(a)] demonstrated that for both 15 $MWcm^{-2}$ and 7.5 $MWcm^{-2}$ power densities, only 7.5 $MWcm^{-2}$ could be cooled to below the safe cut-off temperature, and any HTC was acceptable for this job. For top-sided cooling, the results [Fig. 14(a)] were roughly the same for 7.5 $MWcm^{-2}$, whereas for 15 $MWcm^{-2}$ became possible to realize, only requiring an HTC of 2500 $Wm^{-2}K^{-1}$, easily realizable with liquid cooling. Figure 14(b) shows the 15 $MWcm^{-2}$ sweeps in higher detail, and it quickly becomes clear that the main distinction between top-sided and bottom-sided cooling is again how close to the source of heat the thermal management is positioned. With top-sided cooling, the heat can very quickly be removed from the device, whereas with bottom-sided cooling, it is allowed to accumulate and drives device temperature beyond safe levels. Double-sided cooling showed similarly excellent performance

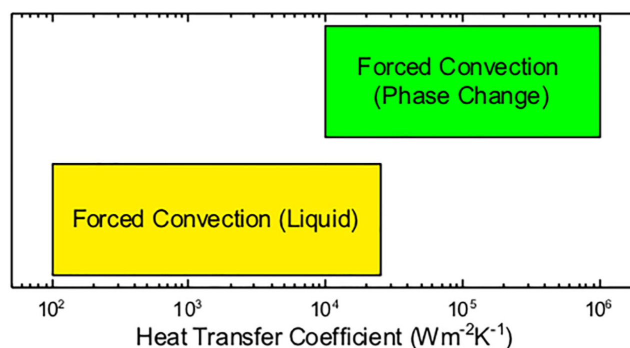
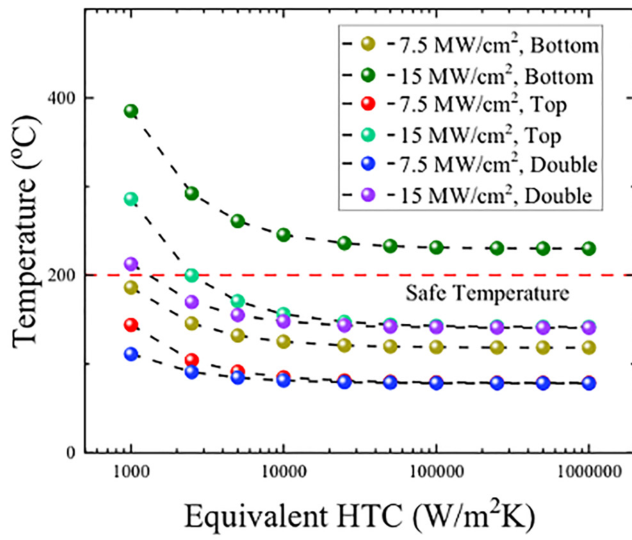


FIG. 13. Ranges of a heat transfer coefficient corresponding to each method considered.

Published under license by AIP Publishing.



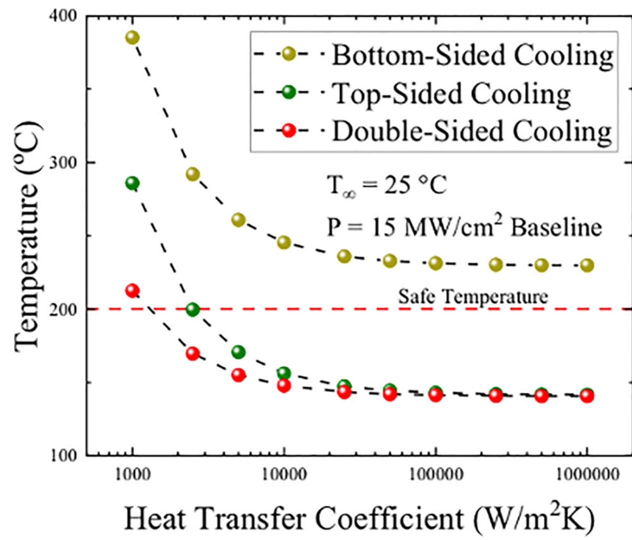


FIG. 14. Sweeps of an equivalent heat transfer coefficient for baseline device configurations vs maximum device temperature in (a) both power levels across all three geometric configurations and (b) detail on 15 $\mathrm{MWcm^{-2}}$, where geometric choice matters. Bottom-sided cooling is not at all safe for cooling these devices when configured at higher power levels but may find some use at lower ones. Top-sided and double-sided cooling both offer good performance at easily attained values of HTC.

[Fig. 14(b)], requiring roughly half the HTC value that top-sided cooling requires to reach safe temperatures at 15 $\mathrm{MWcm^{-2}}$. This is likely a result of double-sided cooling having two heatsinks and thus twice the surface area for heat to exit the device. Examining Fig. 14(b) further, it becomes apparent that at higher HTCs, the advantage that double-sided cooling has over top-sided cooling evaporates. However, as even marginally decent liquid cooling can bring top-sided cooling into the region of safe operation, there is no striking reason to go with double-sided cooling in most situations. Given that bottom-sided cooling (at 15 $\mathrm{MWcm^{-2}}$) never reaches safe operation temperatures at the highest possible HTC (corresponding to phase-change cooling), it is not recommended for use in this class of device without thorough tuning of power dissipation.

F. Multi-fin devices

Through the addition of multiple fins, the overall power dissipated throughout the entire device increases linearly, allowing for the investigation of the effects of thermal crosstalk between devices. Whereas more devices will obviously increase the overall wafer temperature, the effects caused by the interactions of adjacent fins may be minimized by increasing the spacing between devices, thus decreasing thermal crosstalk. To determine at what pitch adjacent fins would experience minimal thermal crosstalk with their neighbors, sweeps were performed on a device composed of three fins ($n = 3$), with best-case thermal conductivities and double-sided cooling. The effects of adding two additional fins to dissipate power are immediately apparent (Fig. 15), as even with the smallest tested pitch ($S = 2\,\mu$m) and at the highest power (10 $\mathrm{MWcm^{-2}}$), the device is nearly able to operate at a safe temperature. Whereas the truly asymptotic value of spacing is not reached within the scope of this study, spacing effects on fins show a trend across all three power values tested—beyond a spacing of roughly 32 μm, additional spacing between devices will only offer diminishing returns in decreasing thermal crosstalk. Reasonable improvements in device temperature cease after setting spacing to roughly 8 μm.

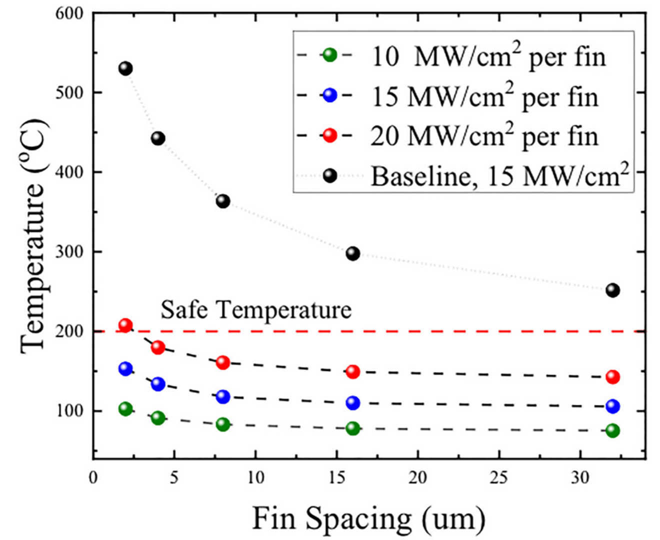


FIG. 15. Effects of increased fin spacing allow for both lower maximum operating temperatures and higher operating powers.

Published under license by AIP Publishing.

A baseline study was also performed for spacing at 15 MWcm^{-2}, with baseline thermal conductivities and heatsink being utilized (Fig. 15). As bottom-sided cooling lacks the ability to carry heat away from the dielectric as well as top-sided or double-sided do, any devices constructed in this method will necessarily require larger spacing to function, something demonstrated by increased steepness of the baseline's slope compared to the curves for ideally managed devices. This increased steepness is indicative that multiple devices implemented with baseline conditions will likely require spacings in excess of 64 μm to be reasonably free from the influence of adjacent devices. Even when spacing has not been optimized, adding more fins continues to be an ideal method for decreasing temperatures within the device.

G. Effects of increasing numbers of fins

To quantify the effect of additional fins in the absence of increased fin spacing, the number of fins and heat generation regions within the die was varied from 3 to 7, with a baseline spacing of 2 μm. All of these studies were performed with optimal thermal conductivity values and double-sided cooling. Increasing the number of fins without increasing the spacing resulted in device temperature increasing semi-linearly (Fig. 16), indicating that thermal crosstalk between the additional fins was beginning to saturate the device's thermal capacity. The concentration of heat at higher power approached the limits of the dielectric's and die attach's ability to remove heat effectively from the devices when only three devices were present—increased spacing would likely lower the effects of this thermal crosstalk. Of particular interest is that additional power seems to scale at a higher rate than the number of fins. This implies that devices are able to dissipate more overall power, if parallel operation is desired, by running more fins in parallel at lower power than is possible by simply pumping that amount of power through one fin. Logically, this makes sense, as additional fins provide access to more material for near-heat source thermal management—spreading out the heat generation over a larger overall volume and giving more paths for the heat to take out of the device.

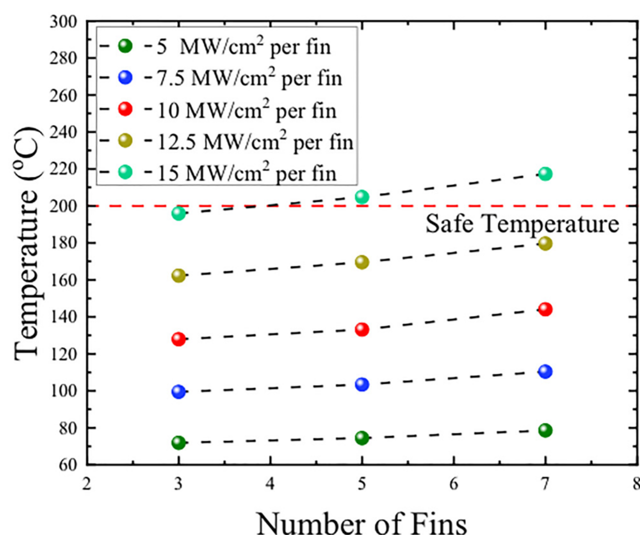


FIG. 16. Impact of increasing number of fins on performance and maximum temperature.

IV. CONCLUSION

Through parametric analysis of the design space, we have determined improvements that can be made to cooling strategies for the thermal management of a vertical gallium oxide fin field-effect transistor. We have studied the effects of three cooling configurations on maximum device temperature, along with the effects that material choice has on these temperatures. We have further investigated the heat transfer coefficients needed to manage the heat produced by devices in this configuration and come to several conclusions regarding the trends we have observed in these idealized geometries. Whereas double-sided cooling strategies are the superior option since they offer more surface area for convective cooling strategies to act upon (and, therefore, allow for lower HTCs), they are also far more difficult to implement from an engineering perspective. For this reason, we recommend that those seeking to efficiently cool gallium oxide chips of this design simply invert the chip and bond the source electrode to their heat spreader, utilizing wire-bonding to make electrical contact with the drain electrode. Additionally, we cannot stress enough the importance of a high-quality dielectric material; as the dielectric is in direct contact with the gate electrode and, therefore, the region of heat generation, it offers a thick and deep region for heat to spread within. Utilizing a low-thermal conductivity dielectric such as SiO_2 is not recommended when designing the thermal management for these devices. When higher electrical performance is desired, a careful balance between the number of devices desired and device spacing must be struck—in general, fin spacing should be kept greater than 8 μm and fin number should be tailored to the application. The parametric studies carried out in this research show that with the proper choice of thermal-management materials and strategies, the drawbacks introduced by the high thermal resistivity of gallium oxide are possible to overcome.

ACKNOWLEDGMENTS

R.H.M., Y.Z., C.Y., S.K., J.S., T.I., A.M., S.K., J.S., and S.G. acknowledge funding from the Air Force Office of Scientific Research (AFOSR) Gallium Oxide Materials Science and Engineering (GAME) Multidisciplinary University Research Initiatives (MURI) under Grant No. FA9550-18-1-0479 under program manager Dr. A. Sayir.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request and with the permission of the Air Force Office of Scientific Research.

Published under license by AIP Publishing.

REFERENCES

[1] M. Higashiwaki, K. Sasaki, A. Kuramata, T. Masui, and S. Yamakoshi, Appl. Phys. Lett. **100**, 013504 (2012).
[2] S. J. Pearton, J. Yang, P. H. Cary, F. Ren, J. Kim, M. J. Tadjer, and M. A. Mastro, Appl. Phys. Rev. **5**, 011301 (2018).
[3] M. Higashiwaki, K. Sasaki, H. Murakami, Y. Kumagai, A. Koukitu, A. Kuramata, T. Masui, and S. Yamakoshi, Semicond. Sci. Technol. **31**, 034001 (2016).
[4] Z. Guo, A. Verma, X. Wu, F. Sun, A. Hickman, T. Masui, A. Kuramata, M. Higashiwaki, D. Jena, and T. Luo, Appl. Phys. Lett. **106**, 111909 (2015).
[5] K. R. Bagnall, S.M. thesis (Massachusetts Institute of Technology, 2013).
[6] Z. Xia, C. Joishi, S. Krishnamoorthy, S. Bajaj, Y. Zhang, M. Brenner, S. Lodha, and S. Rajan, IEEE Electron Device Lett. **39**, 568 (2018).
[7] M. H. Wong, Y. Nakata, A. Kuramata, S. Yamakoshi, and M. Higashiwaki, Appl. Phys. Express **10**, 041101 (2017).
[8] A. J. Green, K. D. Chabak, E. R. Heller, R. C. Fitch, M. Baldini, A. Fiedler, K. Irmscher, G. Wagner, Z. Galazka, S. E. Tetlak, A. Crespo, K. Leedy, and G. H. Jessen, IEEE Electron Device Lett. **37**, 902 (2016).
[9] K. D. Chabak, N. Moser, A. J. Green, D. E. Walker, S. E. Tetlak, E. Heller, A. Crespo, R. Fitch, J. P. McCandless, K. Leedy, M. Baldini, G. Wagner, Z. Galazka, X. Li, and G. Jessen, Appl. Phys. Lett. **109**, 213501 (2016).
[10] N. Kumar, C. Joishi, Z. Xia, S. Rajan, and S. Kumar, IEEE Trans. Electron Devices **66**, 5360 (2019).
[11] N. Kumar, D. Vaca, C. Joishi, Z. Xia, S. Rajan, and S. Kumar, IEEE Electron Device Lett. **41**, 641 (2020).
[12] J. Pomeroy, C. Middleton, M. Singh, S. Dalcanale, M. Uren, M. Wong, K. Sasaki, A. Kuramata, S. Yamakoshi, M. Higashiwaki, and M. Kuball, IEEE Electron Device Lett. **40**, 189–192 (2018).
[13] B. Chatterjee, K. Zeng, C. D. Nordquist, U. Singisetti, and S. Choi, IEEE Trans. Compon. Packaging Manuf. Technol. **9**, 2352–2365 (2019).
[14] Z. Cheng, V. D. Wheeler, T. Bai, J. Shi, M. J. Tadjer, T. Feygelson, K. D. Hobart, M. S. Goorsky, and S. Graham, Appl. Phys. Lett. **116**, 062105 (2020).
[15] H. Zhou, K. Maize, G. Qiu, A. Shakouri, and P. D. Ye, Appl. Phys. Lett. **111**, 092102 (2017).
[16] H. Zhou, M. Si, S. Alghamdi, G. Qiu, L. Yang, and P. D. Ye, IEEE Electron Device Lett. **38**, 103 (2017).
[17] J. Noh, M. Si, H. Zhou, M. Tadjer, and P. Ye, "The impact of substrates on the performance of top-gate p-Ga_2O_3 field-effect transistors: Record high drain current of 980 mA/mm on diamond," in *2018 76th Device Research Conference (DRC)* (IEEE, 2018).
[18] C.-H. Lin, N. Hatta, K. Konishi, S. Watanabe, A. Kuramata, K. Yagi, and M. Higashiwaki, Appl. Phys. Lett. **114**, 032103 (2019).
[19] J. Oh, J. Ma, and G. Yoo, Results Phys. **13**, 102151 (2019).
[20] Y. Zhang, M. Sun, D. Piedra, J. Hu, Z. Liu, Y. Lin, X. Gao, K. Shepard, and T. Palacios, "1200 V GaN vertical fin power field-effect transistors," in *2017 IEEE International Electron Devices Meeting (IEDM)* (IEEE, 2017).
[21] Y. Zhang, M. Sun, Z. Liu, D. Piedra, H.-S. Lee, F. Gao, T. Fujishima, and T. Palacios, IEEE Trans. Electron Devices **60**, 2224 (2013).
[22] M. H. Wong, K. Goto, A. Kuramata, S. Yamakoshi, H. Murakami, Y. Kumagai, and M. Higashiwaki, "First demonstration of vertical Ga_2O_3 MOSFET: Planar structure with a current aperture," in *2017 75th Annual Device Research Conference (DRC)* (IEEE, 2017), p. 1.
[23] K. Sasaki, Q. T. Thieu, D. Wakimoto, Y. Koishikawa, A. Kuramata, and S. Yamakoshi, Appl. Phys. Express **10**, 124201 (2017).
[24] R. Kotecha, W. Metzger, B. Mather, S. Narumanchi, and A. Zakutayev, ECS J. Solid State Sci. Technol. **8**, Q3202 (2019).
[25] M. Sun, M. Pan, X. Gao, and T. Palacios, "Vertical GaN power FET on bulk GaN substrate," in *2016 74th Annual Device Research Conference (DRC)* (IEEE, 2016), p. 1.
[26] Y. Zhang, M. Sun, J. Perozek, Z. Liu, A. Zubair, D. Piedra, N. Chowdhury, X. Gao, K. Shepard, and T. Palacios, IEEE Electron Device Lett. **40**, 75–78 (2018).
[27] S. B. Reese, T. Remo, J. Green, and A. Zakutayev, Joule **3**, 903 (2019).
[28] A. J. Green, K. D. Chabak, M. Baldini, N. Moser, R. Gilbert, R. C. Fitch, G. Wagner, Z. Galazka, J. McCandless, A. Crespo, K. Leedy, and G. H. Jessen, IEEE Electron Device Lett. **38**, 790 (2017).
[29] S. Choi, E. Heller, D. Dorsey, R. Vetury, and S. Graham, J. Appl. Phys. **114**, 164501 (2013).
[30] Z. Cheng, L. Yates, J. Shi, M. J. Tadjer, K. D. Hobart, and S. Graham, APL Mater. **7**, 031118 (2019).
[31] COMSOL Multiphysics® v. 5.4. www.comsol.com. COMSOL AB, Stockholm, Sweden.
[32] C. Yuan, Y. Zhang, R. Montgomery, S. Kim, J. Shi, A. Mauze, T. Itoh, J. S. Speck, and S. Graham, J. Appl. Phys. **127**, 154502 (2020).
[33] A. Franco Júnior and D. Shanafield, Ceramica **50**, 247 (2004).
[34] J. Anaya, H. Sun, J. Pomeroy, and M. Kuball, "Thermal management of GaN-on-diamond high electron mobility transistors: Effect of the nanostructure in the diamond near nucleation region," in *2016 15th IEEE Intersociety Conference on Thermal and Thermomechanical Phenomena in Electronic Systems (ITherm)* (IEEE, 2016), p. 1558.

Published under license by AIP Publishing.