



Cite this: DOI: 10.1039/d5nr03065j

## Nanoscale, temporal temperature mapping in AlN/GaN HEMTs via *ab initio* phonon Monte Carlo simulation

Liam Alexis, <sup>a</sup> Gangchen Ren, <sup>a</sup> Samuel Kielar,<sup>b</sup> Jinghang Dai<sup>a</sup> and Zhiting Tian <sup>\*a</sup>

High-electron-mobility-transistor (HEMT) devices, which implement wide-bandgap (WBG) or ultra-wide-bandgap (UWBG) semiconductors, permit the high-voltage and high-frequency regimes necessary for 5G/6G communications and high-power electronics, but are inhibited by the associated adverse temperature rises. A critical barrier to designing effective thermal dissipation architectures is the lack of accurate temperature information provided by current thermal modeling efforts. We develop an *ab initio* phonon Monte Carlo (MC) framework for nanoscale, temporal temperature mapping of HEMT devices that leverages *ab initio* mode- and temperature-dependent phonon properties and combines ballistic phonon transport and interfacial phonon transmission to capture thermal transport across multilayered structures. We then apply it to an AlN/GaN HEMT device and demonstrate that the decreased thermal conductivity with increasing temperature and decreasing size and the phonon reflection at AlN/GaN interfaces lead to a significantly higher temperature in the hot spot area than conventional technology computer-aided design (TCAD) prediction (at  $7.61 \text{ W mm}^{-1}$  and  $8.43 \times 10^5 \text{ W mm}^{-3}$  heat flux, we observe a  $\sim 80 \text{ K}$  difference), introducing secondary heating zones and exacerbating temperature discontinuities across material interfaces. The detailed temperature map with temporal evolution not only advances our understanding of the heat dissipation process in the HEMT devices but also reiterates the necessity of accurate phonon modeling for hotspot temperature prediction. This framework can be widely applied to predict the temperature profiles of various electronic devices, including those beyond HEMTs.

Received 20th July 2025,  
Accepted 27th April 2026

DOI: 10.1039/d5nr03065j

rscl.li/nanoscale

## Introduction

High-electron-mobility transistors (HEMTs) based on AlN/GaN heterostructures have garnered considerable attention in the field of radio frequency (RF) devices due to their high 2D electron gas (2DEG) charge density and high electron mobility.<sup>1–4</sup> Their electrical performance has already been demonstrated.<sup>5,6</sup> During device operation, a strong electric field with a significant potential gradient can form, accelerating electrons to high kinetic energies. This causes them to pass through the highly resistive local depletion region, creating a spatially concentrated region of intense Joule heating. Self-heating generates a hot spot in the device channel region under the drain side of the gate, which can significantly deteriorate the performance and reliability, as well as the lifespan of the device.<sup>7–9</sup> The dissipation and mitigation of intense heating

within this area require a holistic and comprehensive understanding of non-diffusive thermal transport mechanisms in this region<sup>9,10</sup> and subsequent thermal transport beyond this zone.

Technology computer-aided design (TCAD) simulations have been used to simulate this self-heating effect *via* the modeling of the electric fields generated within the materials of an HEMT device with given voltage inputs. Given heat generation, conventional TCAD models produce steady-state temperature mapping by utilizing diffusive heat transport models with constant thermal conductivities and no thermal interface resistances between materials. While these TCAD models provide a starting point for estimating the temperature rise from the self-heating effect, further models require integration of (1) size-dependent thermal conductivity, (2) temperature-dependent thermal conductivity, (3) phonon interface transmission and reflection, and (4) transient temperature evolution to precisely predict the hot spot temperature and temperature gradient across the device.

Therefore, we have developed an *ab initio* phonon Monte Carlo (MC) framework to address the critical needs. By utiliz-

<sup>a</sup>Cornell University Sibley School of Mechanical and Aerospace Engineering, Cornell University, Ithaca, New York 14853, USA. E-mail: zt223@cornell.edu

<sup>b</sup>Cornell University Department of Materials Science and Engineering, Cornell University, Ithaca, New York 14853, USA



ing a phonon MC approach and integrating first-principles mode-dependent phonon inputs, our *ab initio* MC framework directly captures the (1) ballistic phonon transport and, thus, the size effect, (2) temperature-dependent phonon lifetimes and, thus, the temperature-dependence of thermal conductivity, (3) frequency-dependent phonon interface transmission at AlN/GaN interfaces from an advanced atomistic Green's function method, and (4) temperature evolution with time *via* tracking the phonon propagation.

Traditional phonon MC uses empirical phonon properties.<sup>11–13</sup> While *ab initio* phonon inputs have been implemented for MC in recent studies,<sup>14–21</sup> which focused on characterizing the effects of dopants<sup>18,19</sup> and heat spreading<sup>20,21</sup> and analyzing the effectiveness of MC approaches,<sup>19</sup> one major challenge in multi-material *ab initio* MC simulations remains—the accurate description of phonon transmission across interfaces. While other works utilized the diffuse mismatch model<sup>20</sup> and generalized thermal resistance constants<sup>19</sup> to describe phonon transport and thermal resistance across interfaces, we have focused on phonon interfacial transport using the atomistic Green's function (AGF) method with first-principles force constants from density functional theory (DFT) calculations.<sup>22,23</sup> In this work, we integrate temperature- and mode-dependent phonon scattering rates<sup>24,25</sup> and frequency-dependent interface transmission,<sup>22,23</sup> all from first principles, to allow accurate phonon tracking and temperature

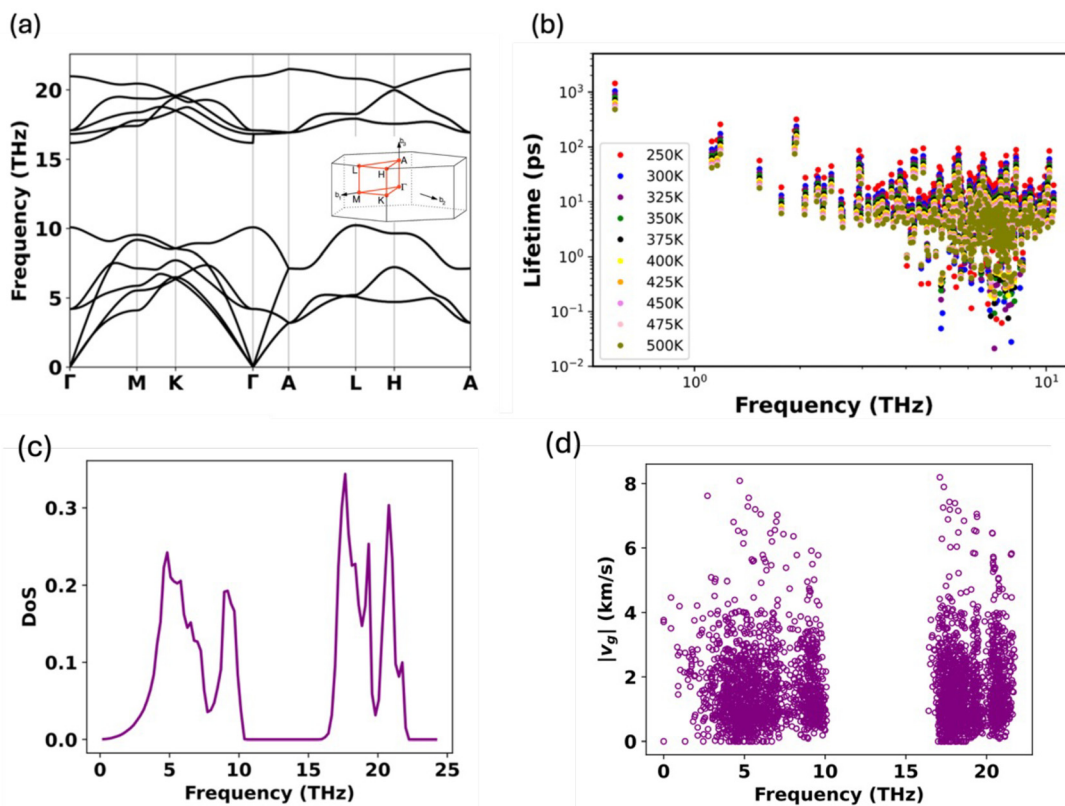
mapping. We use an AlN/GaN HEMT device as an example to show the application of this *ab initio* MC framework. By comparing it with traditional TCAD simulations, we demonstrate the importance of preserving nanoscale phonon physics for accurately predicting hotspot temperatures. Otherwise, the hotspot temperature can be significantly underestimated, thereby affecting thermal management strategies.<sup>19</sup>

## Methods

Accurate simulation of the thermal transport within the AlN/GaN HEMT device geometry requires individual material properties of GaN and AlN using DFT, interfacial phonon transmission from AGF calculations, and usage of statistical frameworks<sup>11</sup> for MC, integrating all the necessary nanoscale phonon effects. Details of the AGF, DFT, and TCAD calculations are included in the SI.

### GaN and AlN DFT calculations

The resulting phonon dispersion is shown in Fig. 1a and agrees well with other literature calculations. These force constants were used as input for the phonon BTE solver to obtain the phonon density of states (DoS), group velocities, and lifetimes. Given that acoustic modes are the major heat carriers, we tracked only the acoustic modes. The frequency-resolved



**Fig. 1** (a) Calculated phonon dispersion of wurtzite GaN. (b) Calculated frequency-resolved total converged lifetimes from consideration of 3-phon scattering processes in wurtzite GaN. (c) GaN density of states. (d) GaN group velocity.



acoustic phonon lifetimes are shown in Fig. 1b. The DoS, group velocities, and lifetimes produced for both our (1) longitudinal acoustic (LA) and (2) transverse acoustic (TA) polarizations are broken up into equally spaced spectral intervals ( $d\omega$ ) in the case of DoS and discrete frequencies  $\omega$  for the lifetimes and group velocities. Further information on the DFT calculations is available in the SI.

### GaN/AlN interface calculations

AGF is a widely applied approach for studying phonon transport across interfaces. More detailed descriptions can be found in the literature.<sup>26–28</sup> In this work, we used DFT-calculated 2<sup>nd</sup>-order FCs as the only input, and hence, no empirical potential was needed.

For our AGF application, the AlN/GaN structure (Fig. 2) was divided into three parts: the left lead (bulk GaN), the central region (parts of bulk GaN and AlN, AlN/GaN interface), and the right lead (bulk AlN). Here, the 2<sup>nd</sup>-order DFT FCs of the leads and the interface were separately obtained from QE<sup>29</sup> and Phonopy.<sup>30</sup> The details can be found in the SI. We want to highlight that we used FCs of the AlN/GaN interface directly obtained from DFT calculations rather than the commonly

simplified model of using FCs of one material and only changing the atomic mass for the other. Note that the pseudopotentials used for the interfacial force-constant calculations differ from those employed in the GaN and AlN DFT calculations. Nevertheless, rigorous convergence tests confirm that these differences do not affect the results. The calculated phonon transmission is shown in Fig. 3a. The corresponding phonon transmittance (probability,  $\tau$ ) can be calculated by normalizing the transmission with that of the pure material on the incident phonon side, as shown in eqn (1). The results are shown in Fig. 3b, which is the probability we use for a phonon to transmit across an interface.

$$\tau_{12}(\omega) = \frac{\Xi(\omega)}{\Xi_1(\omega)}, \quad \tau_{21}(\omega) = \frac{\Xi(\omega)}{\Xi_2(\omega)} \quad (1)$$

where  $\omega$  is the phonon frequency,  $\tau$  is the transmittance (probability), and  $\Xi$  is the transmission function. The interfacial thermal conductance  $G$  obtained from our AGF calculations is in close agreement with previously reported AGF simulation results,<sup>31</sup> further supporting the reliability and validity of our interface calculations.

### TCAD simulations

We used the thermodynamic model in the Sentaurus Device (SDVICE) module to perform the thermal-related simulations. Sentaurus TCAD simulations solve the fundamental equations listed below to analyze the electrical and thermal characteristics of the HEMT device. The Poisson equation is given by

$$\nabla \cdot (\epsilon \nabla \phi - \bar{P}) = -q(p - n + N_D - N_A) - \rho_{\text{trap}} \quad (2)$$

where  $\epsilon$  is the electrical permittivity,  $\phi$  is the electrostatic potential,  $\bar{P}$  is the ferroelectric polarization,  $q$  is the elementary electronic charge,  $p$  is the hole density,  $n$  is the electron density,  $N_D$  is the concentration of ionized donors,  $N_A$  is the concentration of ionized acceptors, and  $\rho_{\text{trap}}$  is the charge

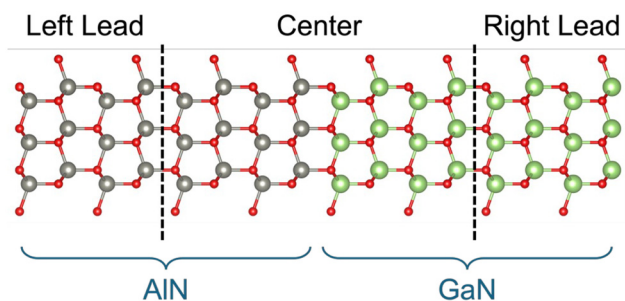


Fig. 2 AlN/GaN interface structure showing left and right leads and the central region, with distinct AlN and GaN bulk layers and the interface between them, respectively.

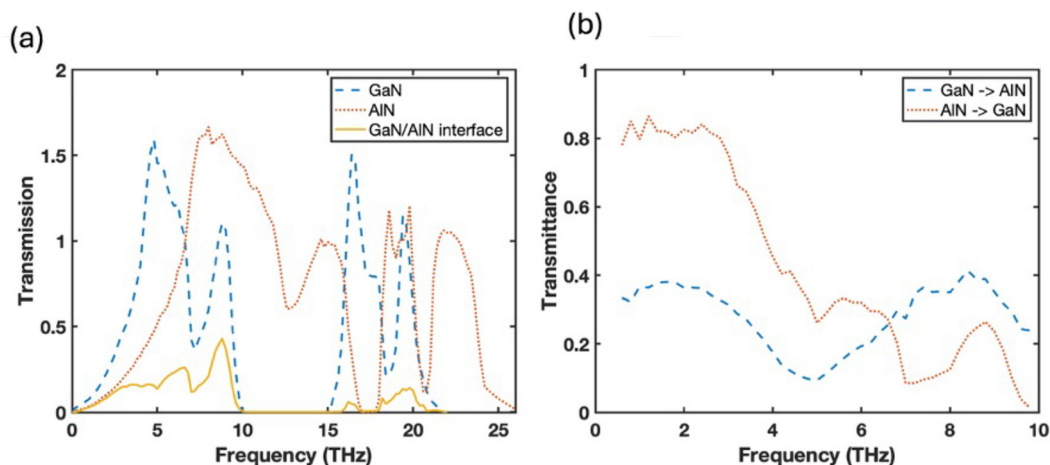


Fig. 3 (a) Calculated GaN, AlN, and GaN/AlN transmission functions using AGF. (b) Phonon transmittance calculated from the transmission function.



density contributed by traps and fixed charges. Continuity equations for electrons and holes are used to describe charge conservation in eqn (3) and (4),

$$\nabla \cdot \bar{J}_n = qR_{\text{net},n} + q\frac{\partial n}{\partial t} \quad (3)$$

$$-\nabla \cdot \bar{J}_p = qR_{\text{net},p} + q\frac{\partial p}{\partial t} \quad (4)$$

where  $\bar{J}_n$  is the electron current density,  $\bar{J}_p$  is the hole current density, and  $R_{\text{net},n}$  and  $R_{\text{net},p}$  are the electron and hole net recombination rates, respectively. The thermodynamic model is activated to analyze the self-heating effect of the device. It is an extension of the drift-diffusion model by including the temperature gradient for current density calculation, as shown in eqn (5) and (6),

$$\bar{J}_n = -nq\mu_n(\nabla\Phi_n + P_n\nabla T) \quad (5)$$

$$\bar{J}_p = -pq\mu_p(\nabla\Phi_p + P_p\nabla T) \quad (6)$$

where  $\mu_n$  and  $\mu_p$  are the electron mobility and hole mobility,  $\Phi_n$  and  $\Phi_p$  are the electron quasi-Fermi potential and hole quasi-Fermi potential,  $P_n$  and  $P_p$  are the absolute thermoelectric powers, respectively, and  $T$  is the lattice temperature. The lattice temperature is computed by solving eqn (7),

$$\begin{aligned} \frac{\partial}{\partial t}(c_L T) - \nabla \cdot (k\nabla T) = & -\nabla \cdot [(P_n T + \Phi_n)\bar{J}_n + (P_p T + \Phi_p)\bar{J}_p] \\ & - \frac{1}{q} \left( E_C + \frac{3}{2}kT \right) (\nabla \cdot \bar{J}_n - qR_{\text{net},n}) \\ & - \frac{1}{q} \left( -E_V + \frac{3}{2}kT \right) (-\nabla \cdot \bar{J}_p - qR_{\text{net},p}) + \hbar\omega G^{\text{opt}} \end{aligned} \quad (7)$$

where  $c_L$  is the lattice heat capacity,  $k$  is the thermal conductivity,  $E_C$  and  $E_V$  are the conduction and valence band energies, respectively,  $G^{\text{opt}}$  is the optical generation rate from photons with frequency  $\omega$ , and  $\hbar$  is Planck's constant divided by  $2\pi$ .

Thermal boundary contact resistance ( $3.9 \times 10^{-5} \text{ cm}^2 \text{ K W}^{-1}$ ) conditions were given by referring to the literature results.<sup>32</sup> The Caughey–Thomas highfield saturation model was used for the mobility. Relevant GaN material physics were included, such as Shockley–Reed–Hall (SRH) and radiative recombination models, piezoelectric polarization, material anisotropy, thermionic emission at heterointerfaces, and interface charge trapping. Also, a nonlocal barrier tunneling model was used for the electrodes to account for the electron tunneling between the GaN valence band and the metal conduction bands.

### Monte Carlo simulations

The MC framework follows the flow outlined in Fig. 4, and the detailed implementation is introduced later. After building the simulated volume and initializing the phonon ensemble, we allow spatial and temporal evolution to track phonon modes, incorporating transport, scattering, and transmission.

### Equilibrium Monte Carlo framework

**Initialization phase.** The initialization utilizes a material-dependent phonon probability density function (PDF) to determine the relative occurrence of various phonon modes and polarizations. Calculations begin with energy and temperature determination. The Bose-Einstein distribution, which relates the equilibrium number of phonons per unit volume,  $\langle n \rangle$ , is related to the frequency,  $\omega$ , and the corresponding DoS value,  $D$ , to produce an equilibrium thermal energy per unit cell,  $E$ . The numerical inversion of  $E$  allows calculation of cell temperature with assumed local thermal equilibrium and is important for determination of the initial phonon ensemble when provided with a target initialization temperature.

$$\langle n \rangle = \frac{1}{\frac{\hbar\omega}{e k_B T} - 1} \quad (8)$$

$$E = \sum_p \int_{\omega} \langle n \rangle \hbar\omega D(\omega) d\omega \quad (9)$$

The PDF is then produced by first finding  $N_i$ , the equilibrium number of phonons per unit volume in the  $i$ th spectral interval for each polarization, which is then combined to produce our  $\omega$  parametrized PDF,  $F_i$ , and our polarization statistical function  $P_i$ .

$$N(p) = \sum_{i=1}^{N_b} \langle n(\omega_{0,i}, p) \rangle D(\omega_{0,i}, p) \Delta\omega_i \quad (10)$$

$$F_i = \frac{\sum_{k=1}^i N_k}{\sum_{k=1}^{N_b} N_k} \quad (11)$$

$$P_i \left( \frac{\text{LA}}{\text{TA}} \right) = \frac{N_i(\text{LA})}{N_i(\text{TA1}) + N_i(\text{TA2}) + N_i(\text{LA})} \quad (12)$$

The simulated volume is broken up into cells of volume,  $V_c$ , with side length,  $x_c$ , which are related to our lattice unit cell volume,  $V_u$ , via the scale factor,  $\text{sf} = \frac{V_u}{V_c}$ . In combination with eqn (8) and (10), we use this sf to determine our initial phonon population at our desired temperature and scale it to our volume of increased temperature. Phonons are then randomly drawn with frequency and polarization determined by  $F_i$  and  $P_i$ . The phonon group velocity and lifetime are then assigned according to phonon frequency and temperature.

**Running phase.** With a fully defined initial phonon ensemble, we transition into the temporal and spatial evolution phase. Phonon movement is mediated by a discrete timestep  $\Delta t$ . Each of these phonons experiences drift and has a chance of undergoing a scattering event.

Phonon drift is characterized by linear ballistic steps of  $\Delta x$  calculated using

$$\Delta x = \Delta t v_i \quad (13)$$



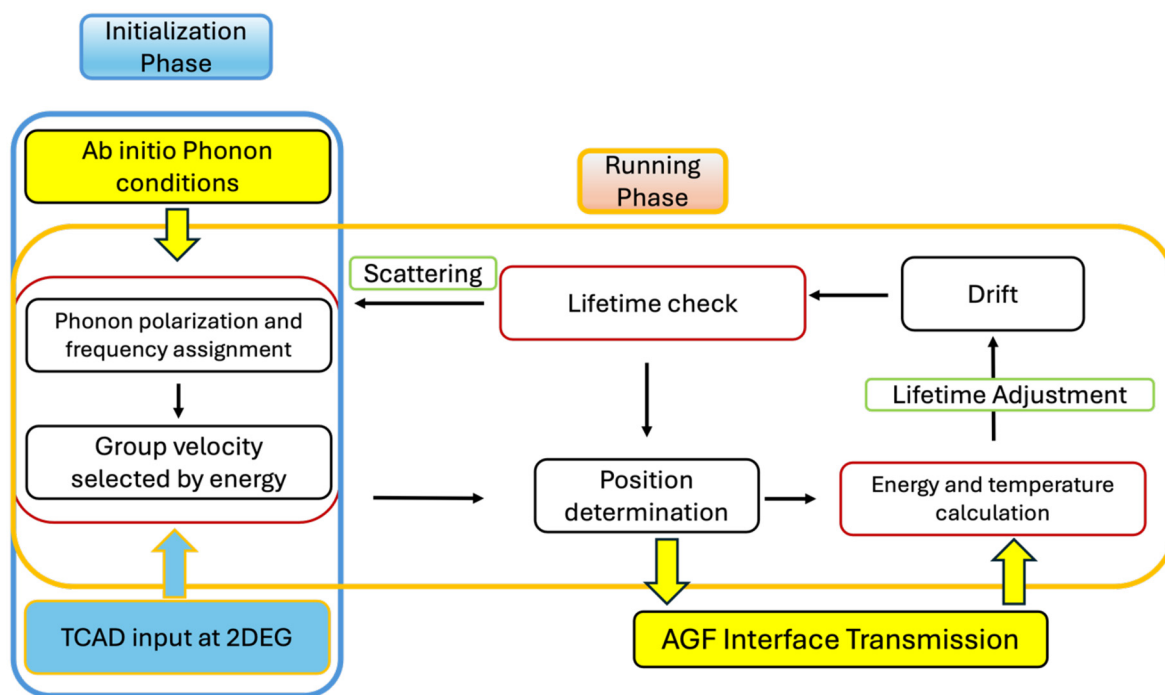


Fig. 4 Monte Carlo workflow. The initialization phase is outlined in a blue box, and the running phase is outlined in orange. *Ab initio* phonon inputs and AGF interface transmission are highlighted in yellow, and TCAD inputs are highlighted in blue.

where  $v_i$  is the phonon group velocity. Phonons may move to new cells, so the energy in each cell is calculated and updated. Assuming thermal equilibrium within each cell, this energy is then used to calculate the local cell temperature using eqn (9) and update the lifetimes of phonons in that cell, producing our spatial temperature map.

Two three-phonon scattering events are of interest to our model – normal and Umklapp scattering; however, the intricacies of each lie outside the scope of this analysis, and the statistics of both are factored into the DFT calculated lifetime and DoS. Scattering is handled computationally by resetting the frequency of our scattered phonon, resampling a frequency from our PDF, and reassigning lifetime and group velocity based on frequency and temperature. Random sampling of a new frequency can result in a violation of energy conservation rules. Therefore, we add or delete phonons at the end of each timestep until energy conservation is maintained.

Our main tool in determining which phonons scatter is the lifetime. The lifetime values provide us with the average time between collisions, and according to that, a time of existence scheme is implemented to determine the age (time since initialization),  $t_{\text{age}}$ , of each phonon with each step. This is then related to the scattering probability,  $P_s$ , and the lifetime,  $\tau$ , via

$$P_s = 1 - \frac{1}{e^{\tau - t_{\text{age}}} \Delta t} \quad (14)$$

This scattering probability is a modified version of the scattering function defined in the literature<sup>11</sup> and takes into con-

sideration the increasing probability of a scattering event occurring as the  $t_{\text{age}}$  increases. As this  $t_{\text{age}}$  approaches the lifetime, the overall scattering probability will increase as  $e^{\tau - t_{\text{age}}}$  approaches infinity.

The final runtime consideration necessary to holistically simulate our device geometry is transmission. The transmittance is presented as a normalized probability dependent on the phonon frequency. As phonons approach the material interface, a random number is drawn and compared with their transmittance. If a higher number is drawn, the phonon is reflected; if a lower number is drawn, the phonon is transmitted. Upon transmission, the frequency is maintained, but the group velocity and lifetime are changed according to the dynamics of the new material. More detailed descriptions can be found in the literature.<sup>11</sup>

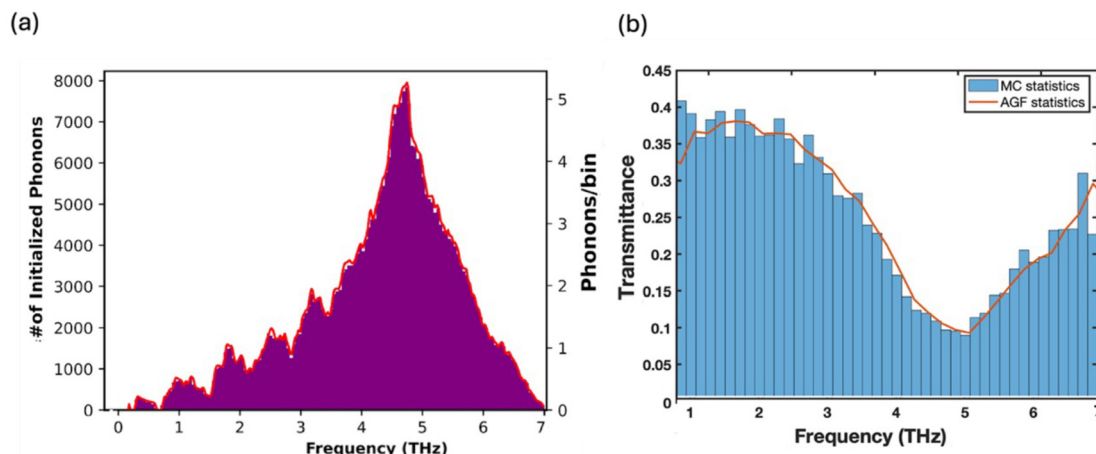
Adherence to both the expected distribution as determined by the DoS and BE distribution and the interfacial transmittance statistics as defined by AGF calculations is shown in Fig. 5, with the expected statistical trendlines being closely followed by the histograms of initialized and transmitted phonons, respectively.

## Results and discussion

### TCAD

The lattice temperature for the HEMT device is calculated by utilizing the thermodynamics model from Sentaurus TCAD. Following the conventional procedure of using the heat





**Fig. 5** (a) The purple histogram, which represents the Monte Carlo mediated random distribution of phonons, closely follows the red line representing the DoS and Bose-Einstein distribution predicted phonon spectral bin occupation. (b) The tracked phonon transmission event statistics (histogram) from GaN to AlN follow the red line, which represents the AGF-predicted phonon transmission from GaN to AlN as a function of phonon energy.

diffusive conduction model, which assumes a constant bulk thermal conductivity value and no thermal interface resistance in between the layers, the temperature map for the device is shown in Fig. 6b. The simulated power dissipation is  $7.61 \text{ W mm}^{-1}$  from the drain voltage of 4 V and the corresponding drain current of  $1.903 \times 10^{-3} \text{ A } \mu\text{m}^{-1}$ . A hot spot is located at the drain side of the gate of the channel region, where the electric field potential is the greatest, leading to a more significant self-heating effect. To have a fair comparison with our MC simulation, heat flux distribution from the lattice temperature calculation *via* the thermodynamics model is extrapolated as presented in Fig. 6b. For this 2D TCAD calculation, we have probed both the  $x$ - $z$  and  $y$ - $z$  plane heat flux distributions. As the heat flux from both planes is very close to each other, a uniform heat dissipation around the hot spot region is observed. This volumetric heat flux of  $8.43 \times 10^5 \text{ W mm}^{-3}$  from the self-heating effect calculation acts as the initial heat generation for the later MC calculation. The temperature of the hot spot in our TCAD model attains a maximum temperature of 354 K.

### Monte Carlo

Our simulation geometry mimics Fig. 6a with the initial heat generation area around the 20 nm zone and the simulated TCAD geometry Fig. 6b focusing on the  $50 \text{ nm} \times 50 \text{ nm} \times 50 \text{ nm}$  volume encompassing the entire thickness of the upper AlN layer and GaN layers and 12 nm into the lower AlN layer (shown as the dashed boxes in Fig. 6a and b). The AlN/GaN interfaces are located at  $Z = 20 \text{ nm}$  (z20) and  $Z = -12 \text{ nm}$  (z12). The TCAD simulation heat flux density provides time- and volumetrically parametrized energy, which is translated into a phonon injection function that adds phonons to the 2DEG area at each timestep.

This modeled device has a background temperature of 300 K, which represents the environment in which the device

is expected to operate. Instead of increasing the number of phonons or reducing the ratio of the simulated *vs.* real number of phonons to simulate higher temperature, we treat 300 K as the background and introduce thermal phonons on top of the background energy. We intentionally keep a large ratio of the simulated *vs.* real number of phonons (1 : 10) to best represent the actual dynamics. If we were to directly simulate the 300 K phonon ensemble within our volume, we would deal with the order of  $10^9$  simulated phonons, which only increases with heat injection.

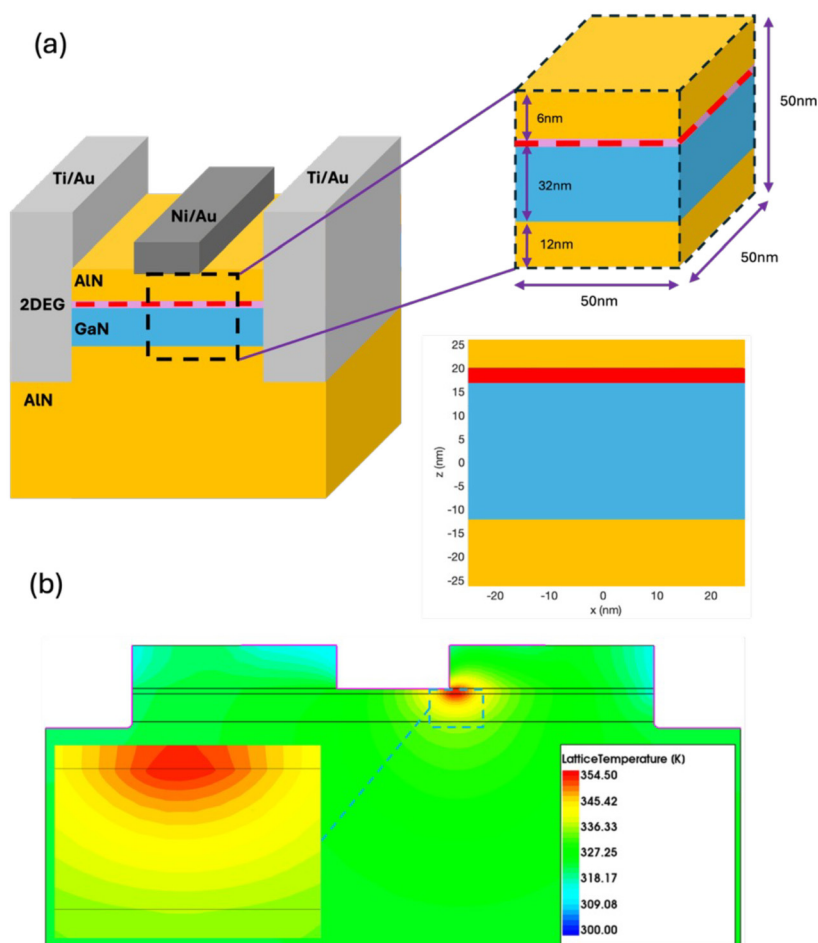
Deviated density of states (DoS) and scattering functions were utilized in our simulations, which allow us to build our phonon ensemble on top of an equilibrium background phonon flux maintained at 300 K to reduce the computational cost. This relies on adjusted probability distributions and scattering mechanics obtained *via* the calculation of the deviated phonon occupation  $\langle n_d \rangle$ .<sup>33</sup>

$$\langle n_d \rangle = \frac{1}{e^{k_B T_1} - 1} - \frac{1}{e^{k_B T_{eq}} - 1}$$

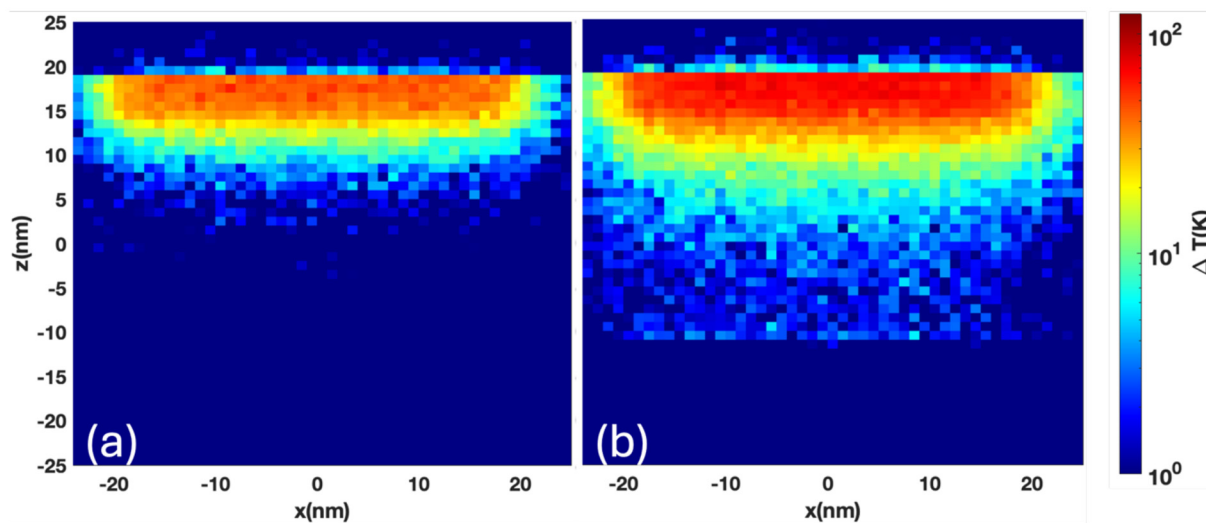
where  $T_1$  is a local temperature used in calculating the deviation, and  $T_{eq}$  is 300 K. Equilibrium *vs.* deviated PDF examples are displayed in Fig. S6.

Two simulation modes are utilized within our study. The first mode focuses on the time interval from heat initialization at 0 ps to 30 ps. This simulated space is broken into  $1 \text{ nm}^3$  cells and cycles through 400 iterations of  $7.5 \times 10^{-2} \text{ ps}$  each. This ensures that phonons are allowed to begin impinging upon each AlN/GaN interface, focusing on the initial effects around the z20 interface. The second mode focuses on longer-term interactions up to 200 ps. This simulation space is broken into  $8 \text{ nm}^3$  cells and cycles through 1300 iterations of  $1.55 \times 10^{-1} \text{ ps}$  each. This allows us to observe longer-scale temporal effects and the development of thermal gradients that





**Fig. 6** (a) Geometry of the simulated HEMT device outlining the simulated volume of interest encompassing the GaN layer, upper AlN layer, and 12 nm into the lower AlN layer, and 2D slice utilized for temperature profile diagrams. (b) TCAD temperature profile in the simulation geometry of a 2D HEMT device, highlighting a 50 nm × 50 nm slice encompassing the same area as displayed in MC simulation results.



**Fig. 7** Shorter time, 1 nm<sup>3</sup> resolution analysis. (a) 15 ps and (b) 30 ps showing the temperature distribution with our three material layers with intense thermal resistance observed across the z20 interface with minimal phonon transmittance.

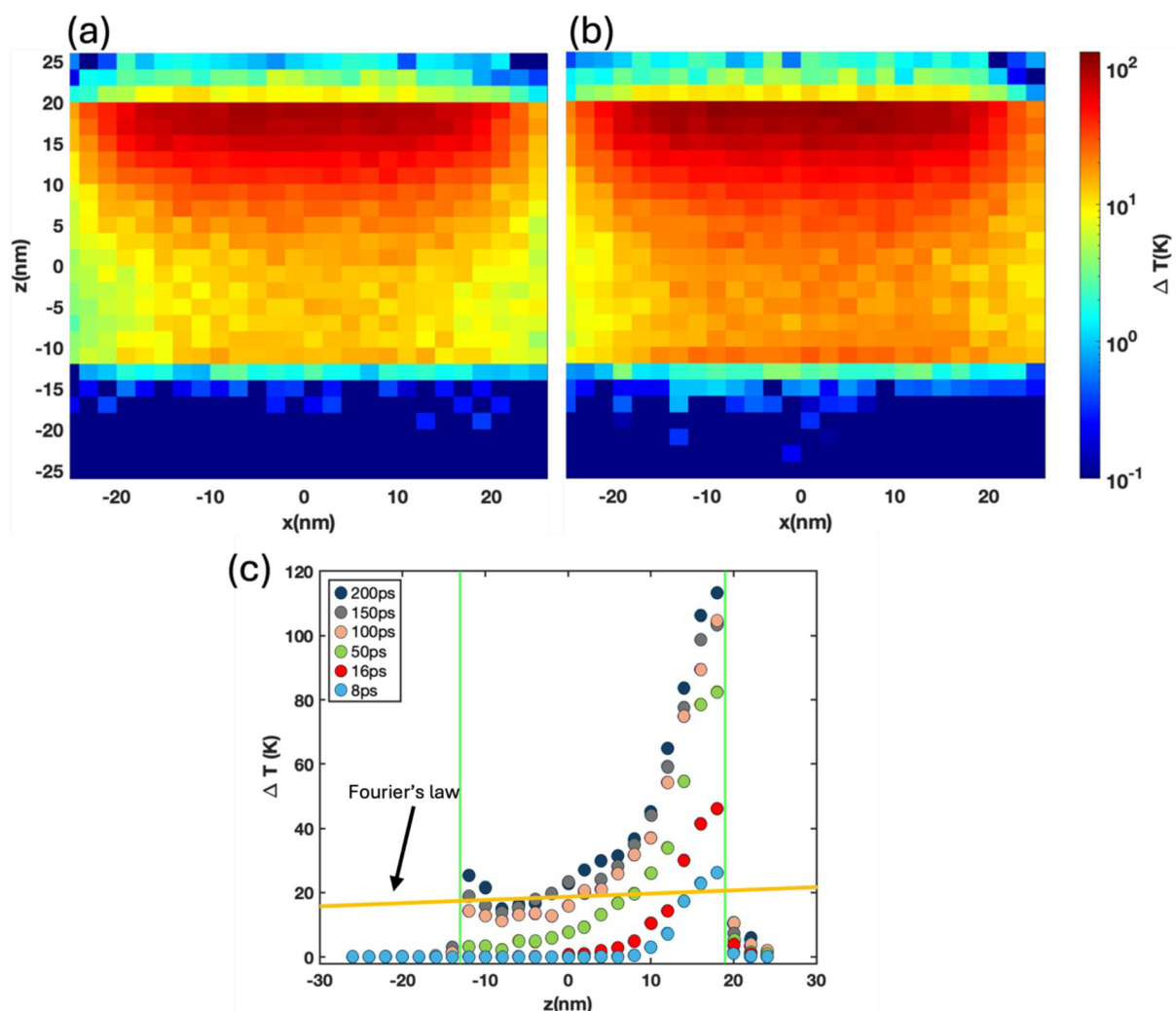


encompass the z12 interface. We place the same heat flux volume on the GaN side of the z20 interface as TCAD.

For the shorter time simulation using the  $1 \text{ nm}^3$  resolution, our MC attains a temperature of 358 K at 15 ps (Fig. 7a) and a maximum temperature of 382 K at 30 ps (Fig. 7b). The hot spot in the MC simulation attains a higher temperature than the steady-state hot spot temperature from TCAD of 354 K due to the influence of temperature and size on scattering rates and thermal conductivity, as well as the thermal interface resistance. Scattering rates increase with increasing temperature due to the Umklapp phonon scattering, meaning that using a constant thermal conductivity value at room temperature in TCAD overestimates thermal conductivity and underestimates the temperature rise. In addition, the nm-thick layers strongly suppress the thermal conductivity from the bulk value due to the boundary scattering (size effect). Using a bulk value also contributes to overestimating thermal conductivity and under-

estimating temperature rise in TCAD. Moreover, there is thermal resistance at the material interfaces, and one should expect a temperature discontinuity at the interface accordingly. A temperature drop was not seen in TCAD because the thermal interface resistance was not considered. In contrast, our MC framework integrates the first-principles phonon interface transmission for the first time and properly captures the temperature drop at the interface. With these statistical effects, as phonons hit the z20 interface from GaN to AlN, only 15–20% of phonons are transmitted, with the other 80–85% being reflected back into the GaN layer; not only is AlN heating less due to receiving only a small portion of phonons, but heating within GaN is augmented due to losing a small portion of phonons to the AlN layer, further increasing the temperature drop across the z20 interface.

As time progresses to 30 ps (Fig. 7b) the interfacial transmittance trend continues with the reflection of phonons back



**Fig. 8** Longer time,  $8 \text{ nm}^3$  analysis (a) 100 ps image showing a well-defined secondary heating zone at the z12 interface. (b) 200 ps image showing a well-defined secondary heating zone at the z12 interface with an increased temperature in the zone between the central hot spot and the z12 interface. (c)  $x = 0$  temperature sweep showing interfacial thermal gradients due to interfaces with secondary heating effects at the z12 interface apparent at 100 ps with significant deviation from Fourier's law.



into the GaN layer at the z20 interface, further increasing the GaN temperature and increasing the temperature drop. The interfacial effects also increase the size of the central hot spot region and broaden the surrounding excited region on the GaN side, with minimal heating still observed on the AlN side.

Looking further past the hot spot area into the lower range (below  $z = 0$ ), a runtime of 30 ps also begins to show the onset of temperature drops at the z12 interface, this being totally attributed to the 15%–20% phonon transmission across the interface. The  $z = 0$  to  $z = -12$  range shows lower temperatures directly adjacent to the hot spot area because phonons generated in the hot spot area have been slowed down by the increased temperature *via* the increased scattering, impeding their movement. The reflected phonons from the z20 interface influence faster heating on the GaN side of the z20 interface, increasing scattering rates, reducing thermal conductivity, and producing a positive feedback temperature loop. As phonons are reflected, the hot spot temperature increases. Phonons generated in the hot region are more likely to scatter, and with each scattering, the velocity randomizes, delaying their movement toward the z12 interface and necessitating analysis at later times.

To allow for a longer simulation time, we switched to the analysis with the  $8 \text{ nm}^3$  cell size. Fig. 8a and b display the temperature profiles at 100 ps and 200 ps, respectively, where the interfacial effects are even more profound with more drastic temperature drops than the short runtimes (Fig. 7). Fig. 8c shows the temperature profile along the  $x = 0$  line at various times. The hot spot temperature increases with time and starts to plateau at 200 ps. The maximum temperature is observed at the interface. The z20 interfacial temperature drop is also on the order of  $\sim 100 \text{ K}$ , with a comparable temperature drop in the GaN layer from the max temperature zone not seen until around 20 nm away from the z20 interface. Unlike the linear temperature profile one would expect within the GaN

layer from Fourier's law, the nonlinearity in Fig. 8c highlights the complex nature of varying thermal conductivity with temperature, constant heat flux input, and transient conditions, combined with the influence of the interfaces.

The hot spot size continues to grow compared to the  $1 \text{ nm}^3$  case at shorter times, as expected. Between 100 ps and 200 ps, the central hot spot area ( $>400 \text{ K}$ ) extends marginally, with the outside area expanding  $\sim 10 \text{ nm}$ . At the z12 interface, we observe a local maximum *via* secondary heating events influenced by the reflection effects of the z12 interface. The area between the hot spot and the z12 interface experiences a relatively even temperature distribution due to a balancing effect from the hot spot phonons and the reflected z12 phonons. This secondary heating zone is not observed at all if there is no interface.

The maximum temperature within this hot spot is 430 K, representing a  $\sim 80 \text{ K}$  increase from the TCAD-determined conditions. This area primarily encompasses a  $30 \text{ nm} \times 5 \text{ nm}$  area on the GaN side of the z20 interface. Surrounding this is a region of 400 K that encapsulates a  $40 \text{ nm} \times 10 \text{ nm}$  area, which leads to the 10 K–100 K excitation region extending down to the z20 boundary.

For a more direct comparison of the effect of size and temperature effects within MC to the TCAD simulations, the MC simulation was run without the frequency-based statistical transmittance of the material interfaces. All phonons that reach interfaces (with velocity vectors pointing across the interfaces) are allowed to cross over and adopt the dynamics of the new material. Fig. 9 shows the temperature distribution at two timesteps with our  $1 \text{ nm}$  resolution. A maximum temperature of 380 K is obtained from this simulation with no indication of dominant interfacial resistance or secondary heating zones, with more even heating occurring above and below the central hot spot region and a slightly quicker drop seen at z20 due to the higher AlN thermal conductivity. This represents a  $\sim 30 \text{ K}$

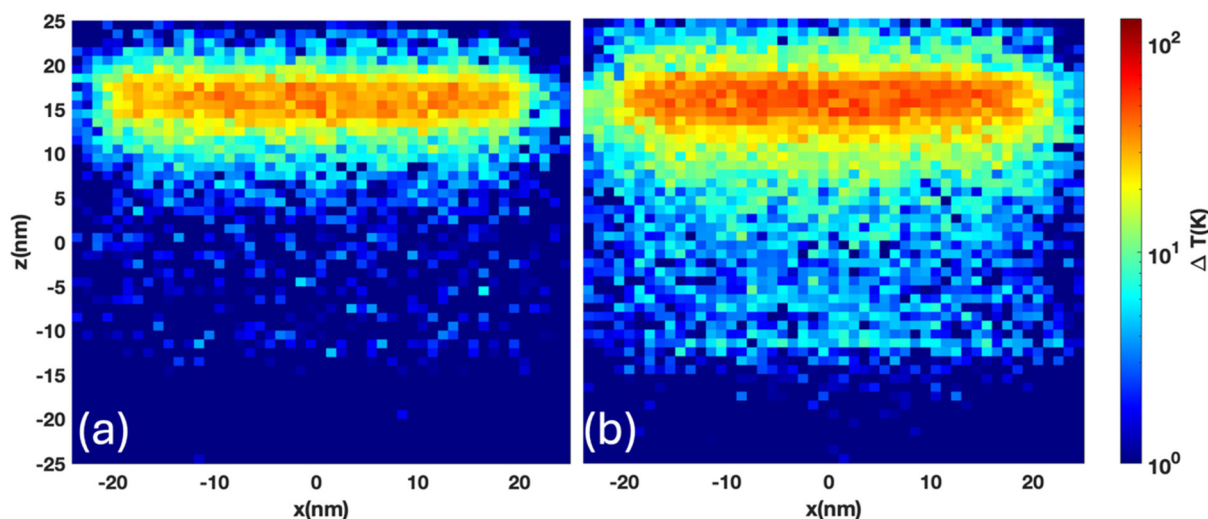


Fig. 9 Shorter time,  $1 \text{ nm}^3$  resolution analysis. (a) 15 ps and (b) 30 ps showing the temperature distribution with our three material layers, with no statistical interface treatment showing no interfacial resistance or secondary heating zones.



increase in maximum temperature compared to the TCAD simulation and a decrease of  $\sim 50$  K compared to the MC with interfaces. This increase in temperature compared to the TCAD case is only due to the size and temperature effects and allows us to isolate the influence of interfaces on the thermal evolution within the multilayered architecture. When combined, these interfacial, size and temperature-dependent effects build upon each other, resulting in an 80 K increase in temperature for the MC case with interfaces.

Another TCAD case and MC simulation were run with a 6 V drain voltage, producing a maximum TCAD temperature of 387 K and the corresponding maximum MC temperature of 480 K. Figures outlining these results are included in the SI (Fig. S4) and show a similar secondary heating zone at the lower interface with higher temperatures in the MC case.

The consolidation of the probabilistic methods utilized to capture both intra- and inter-material phonon movement within this MC simulation provides significant insight into the effect of interfaces on the temporal and spatial propagation of various phonon frequencies. Our framework preserves the accuracy of *ab initio* calculations and the statistical methods typically seen in MC simulations while allowing isolation of the interfacial effects, allowing us to directly observe the resultant differences in temperature evolution and distribution, the importance of temperature-dependent thermal effects within thermal analysis and the influence of material interfaces on all these effects.

## Summary

We have developed a comprehensive MC framework that enables accurate temperature mapping in layered multi-material systems by integrating *ab initio* phonon properties, temperature-dependent scattering rates, and interface phonon transmission. By utilizing TCAD-calculated heat flux inputs, we are able to apply these simulations to model hot spots in HEMT devices. Through the integration of these probabilistic phonon effects, we show that the temperatures seen within the TCAD simulations are underestimated by  $\sim 80$  K at  $7.61 \text{ W mm}^{-1}$ , in addition to showing a large drop in temperature distribution across material interfaces. This demonstrates the importance of factoring in temperature, size, and interface-dependent effects in considering thermal transport in these systems with multiple materials and demonstrates the potential functionality of multi-material *ab initio* simulations in modeling HEMT and microelectronics systems in general.

## Conflicts of interest

There are no conflicts to declare.

## Data availability

Data supporting this article have been included in the figures and tables of the main text. All equations, coding packages,

and methods have also been outlined within the main text. The dimensions of the AlN/GaN HEMT device were provided by Prof. Sukwon Choi at Penn State.

Supplementary information (SI) is available. See DOI: <https://doi.org/10.1039/d5nr03065j>.

## Acknowledgements

This work is supported by DARPA YFA (D23AP00159-00).

## References

- 1 T. Zimmermann, *et al.*, AlN/GaN insulated-gate HEMTs with 2.3 A/mm output current and 480 mS/mm transconductance, *IEEE Electron Device Lett.*, 2008, **29**, 661–664.
- 2 F. Medjdoub, *et al.*, Low-noise microwave performance of AlN/GaN HEMTs grown on silicon substrate, *IEEE Electron Device Lett.*, 2011, **32**, 1230–1232.
- 3 M. Higashiwaki, T. Mimura and T. Matsui, AlN/GaN insulated-gate HFETs using Cat-CVD SiN, *IEEE Electron Device Lett.*, 2006, **27**, 719–721.
- 4 L. Yang, *et al.*, AlN/GaN HEMTs with  $f_{\text{max}}$  Exceeding 300 GHz by Using Ge-Doped n++GaN Ohmic Contacts, *ACS Appl. Electron. Mater.*, 2023, **5**, 4786–4791.
- 5 H. Li, *et al.*, High Performance AlN/GaN HEMT for Millimeter-Wave Low-Voltage Applications Fabricated Using Low-Damage Etching, *IEEE Electron Device Lett.*, 2024, **45**, 2327–2330.
- 6 K. D. Chabak, *et al.*, High-performance AlN/GaN HEMTs on sapphire substrate with an oxidized gate insulator, *IEEE Electron Device Lett.*, 2011, **32**, 1677–1679.
- 7 Y. Qin, *et al.*, Study of Self-Heating and High-Power Microwave Effects for Enhancement-Mode p-Gate GaN HEMT, *Micromachines*, 2022, **13**, 106.
- 8 M. Gonschorek, J. F. Carlin, E. Feltn, M. A. Py and N. Grandjean, Self heating in AlInN/AlN/GaN high power devices: Origin and impact on contact breakdown and IV characteristics, *J. Appl. Phys.*, 2011, **109**, 063720.
- 9 B. Chatterjee, *et al.*, Nanoscale electro-thermal interactions in AlGaIn/GaN high electron mobility transistors, *J. Appl. Phys.*, 2020, **127**, 044502.
- 10 S. Choi, E. R. Heller, D. Dorsey, R. Vetry and S. Graham, The impact of bias conditions on self-heating in AlGaIn/GaN HEMTs, *IEEE Trans. Electron Devices*, 2013, **60**, 159–162.
- 11 S. Mazumder and A. Majumdar, Monte Carlo study of phonon transport in solid thin films including dispersion and polarization, *J. Heat Transfer*, 2001, **123**, 749–759.
- 12 M.-S. Jeng, R. Yang and G. Chen, *Monte Carlo Simulation of Thermoelectric Properties in Nanocomposites*, 2005, <https://nanoengineering.mit.edu>.
- 13 J. P. M. Péraud and N. G. Hadjiconstantinou, Efficient simulation of multidimensional phonon transport using



- energy-based variance-reduced Monte Carlo formulations, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **84**, 205331.
- 14 G. Chen, B. Hu, Z. Wang and D. Tang, Coupled nonequilibrium Monte Carlo simulations of thermal transport mediated by nanoscale hotspot in GaN transistors, *Int. J. Therm. Sci.*, 2023, **194**, 108592.
  - 15 B. Hu, Z. Wang, K. Xu and D. Tang, Hotspot and nonequilibrium thermal transport in AlGaIn/GaN FinFET: A coupled electron-phonon Monte Carlo simulation study, *Int. J. Heat Mass Transfer*, 2025, **241**, 126679.
  - 16 X. Xie, H. Chen, Z. Wang and K. Xu, Temperature-corrected full-band Monte Carlo simulation of phonon transport mechanism in 2D GaN, *Int. J. Therm. Sci.*, 2025, **210**, 109648.
  - 17 Y. Shen, H.-A. Yang and B.-Y. Cao, Near-junction phonon thermal spreading in GaN HEMTs: A comparative study of simulation techniques by full-band phonon Monte Carlo method, *Int. J. Heat Mass Transfer*, 2023, **211**, 124284.
  - 18 J. Xu, Y. Sheng, Z. Wang and H. Bao, Isovalent Doping Reduces Channel Temperature Within Nanoscale Devices, *IEEE Trans. Electron Devices*, 2025, **72**, 4333–4339.
  - 19 Y. Sheng, *et al.*, Multiscale Thermal Simulation for GAAFET With First-Principles-Based Boltzmann Transport Equation, *IEEE Trans. Electron Devices*, 2025, **72**, 4700–4707.
  - 20 Y. Shen, *et al.*, Bias Dependence of Non-Fourier Heat Spreading in GaN HEMTs, *IEEE Trans. Electron Devices*, 2023, **70**, 409–417.
  - 21 Y. C. Hua, H. L. Li and B. Y. Cao, Thermal Spreading Resistance in Ballistic-Diffusive Regime for GaN HEMTs, *IEEE Trans. Electron Devices*, 2019, **66**, 3296–3301.
  - 22 Z. Tian, K. Esfarjani and G. Chen, Enhancing phonon transmission across a Si/Ge interface by atomic roughness: First-principles study with the Green's function method, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 235304.
  - 23 Z. Tian, K. Esfarjani and G. Chen, Green's function studies of phonon transport across Si/Ge superlattices, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2014, **89**, 235307.
  - 24 S. Kielar, *et al.*, Anomalous lattice thermal conductivity increase with temperature in cubic GeTe correlated with strengthening of second-nearest neighbor bonds, *Nat. Commun.*, 2024, **15**, 6981.
  - 25 Z. Tian, *et al.*, Phonon conduction in PbSe, PbTe, and PbTe<sub>1-x</sub>Se<sub>x</sub> from first-principles calculations, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **85**, 184303.
  - 26 N. Mingo and L. Yang, Phonon transport in nanowires coated with an amorphous material: An atomistic Green's function approach, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2003, **68**, 2454061–24540612.
  - 27 Z. Tian, K. Esfarjani and G. Chen, Enhancing phonon transmission across a Si/Ge interface by atomic roughness: First-principles study with the Green's function method, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2012, **86**, 235304.
  - 28 J. Dai and Z. Tian, Rigorous formalism of anharmonic atomistic Green's function for three-dimensional interfaces, *Phys. Rev. B*, 2020, **101**, 041301.
  - 29 P. Giannozzi, *et al.*, QUANTUM ESPRESSO: A modular and open-source software project for quantum simulations of materials, *J. Phys.: Condens. Matter*, 2009, **21**, 395502.
  - 30 A. Togo and I. Tanaka, First principles phonon calculations in materials science, *Scr. Mater.*, 2015, **108**, 1–5.
  - 31 C. A. Polanco and L. Lindsay, Phonon thermal conductance across GaN-AlN interfaces from first principles, *Phys. Rev. B*, 2019, **99**, 075202.
  - 32 B. F. Donovan, *et al.*, Thermal boundary conductance across metal-gallium nitride interfaces from 80 to 450 K, *Appl. Phys. Lett.*, 2014, **105**, 203502.
  - 33 J.-P. M. Péraud and N. G. Hadjiconstantinou, Efficient simulation of multidimensional phonon transport using energy-based variance-reduced Monte Carlo formulations, *Phys. Rev. B: Condens. Matter Mater. Phys.*, 2011, **84**, 205331.

