Phonon black-body radiation limit for heat dissipation in electronics

J. Schleeh, J. Mateos, I. Íñiguez-de-la-Torre, N. Wadefalk, P. A. Nilsson, J. Grahn and A. J. Minnich

Thermal dissipation at the active region of electronic devices is a fundamental process of considerable importance. Inadequate heat dissipation can lead to prohibitively large temperature rises that degrade performance, and intensive efforts are under way to mitigate this self-heating. At room temperature, thermal resistance is due to scattering, often by defects and interfaces in the active region, that impedes the transport of phonons. Here, we demonstrate that heat dissipation in widely used cryogenic electronic devices instead occurs by phonon black-body radiation with the complete absence of scattering, leading to large self-heating at cryogenic temperatures and setting a key limit on the noise floor. Our result has important implications for the many fields that require ultralow-noise electronic devices.

Heat dissipation in transistors occurs in a two-step process in which electrons in the transistor channel emit phonons that then carry the heat away from the active region. The processes that impede phonon transport in these devices are, thus, of critical importance as they determine the self-heating that occurs. In many transistors, such as high-power GaN transistors, thermal resistance is primarily due to interfaces and defects that scatter phonons, reducing the effective thermal conductivity of the near-junction materials by orders of magnitude from their intrinsic values. In principle, the minimum thermal resistance can be achieved in a perfect harmonic lattice in which scattering does not occur and phonons propagate away from the heat generation region at group velocity. This regime was first described by Casimir and is often referred to as phonon black-body radiation, owing to its similarities to thermal radiation by photons.

In practice, entering the phonon radiation regime requires high-quality materials and length scales that are much smaller than the intrinsic mean free paths (MFPs) set by the strength of phonon–phonon scattering. Whereas the necessary length scales at room temperature are small, of the order of 10 nm in typical semiconductors such as Si, this regime could be achieved even at micrometre length scales in cryogenic electronics that operate at around $T \approx 15$ K because phonon MFPs increase exponentially with decreasing temperature. At typical operating temperatures for cryogenic, ultralow-noise amplifiers, in which channel lengths are a few hundred nanometres, intrinsic MFPs are essentially infinite. However, the contribution of interfaces and defects to thermal resistance remains unclear as the heat conduction mechanisms in these devices have been minimally investigated. Accurately determining the magnitude of the self-heating—and, hence, the thermal noise—is of considerable importance, as the origin of the noise determines the relevant strategies to create ultralow-noise devices.

Here, we investigate thermal dissipation mechanisms in cryogenic amplifiers using noise measurements and efficient phonon transport simulations. We show that heat dissipation in these devices occurs by phonon black-body radiation without scattering, that nevertheless results in large self-heating in the limit of zero ambient temperature. Our work demonstrates that this fundamental limit on heat dissipation is a key consideration in the performance of ultralow-noise electronic devices.

We fabricated a three-stage 4–16 GHz monolithic microwave integrated circuit (MMIC) cryogenic amplifier consisting of three InP high electron mobility transistors (HEMTs), as shown in Fig. 1a,b. The HEMTs contain two gate fingers of 100 μm width, as shown in Fig. 1b. We used a standard procedure to measure the noise temperature as a function of the device operating temperature ($T_d$; see Methods and Supplementary Information for details on fabrication and characterization). We plot the results, taken at 12 GHz and with a power input per length along the 100 μm dimension of 29 mW mm$^{-1}$, in Fig. 1c. Consistent with previous reports, the noise temperature decreases linearly with the device operating temperature until around 40 K, below which it deviates from this trend and decreases more slowly until it saturates.

To better understand the origin of the noise, we use a noise model to interpret the measurements. In this model, noise is described by assigning equivalent temperatures to resistive elements on the input and output. Typically, the input is assumed to be at the device operating temperature whereas the output has an elevated temperature that is attributed to ambient temperature-dependent hot-electron noise. The input noise is, hence, directly dependent on the intrinsic physical temperature, whereas the output is indirectly dependent. We extract the equivalent drain temperature $T_d$ using our measurements and we assume that all other temperatures are set at the device operating temperature. The result is shown in Fig. 1d. $T_d$ follows the prediction by the circuit model until about 40 K, below which it saturates to a constant value.

The observation of a saturation of drain temperature is unexpected because $T_d$ should continue to decrease steadily throughout the temperature range if the channel were isothermal with the ambient, with the intercept at $T_i = 0$ yielding the hot-electron noise contribution when thermal noise is absent. To interpret these results, we postulate that the intrinsic temperature $T_i$ may be different from $T_d$ owing to self-heating. We estimate the contribution of non-thermal noise by fitting $T_d$ with a line for $T_d > 40$ K and extrapolating to zero temperature, as shown in Fig. 1d.
To explain this apparent self-heating, we must examine heat dissipation by phonons in the transistor channel. Phonon transport in the device is governed by the phonon Boltzmann transport equation (BTE):

\[ \frac{\partial e_\omega}{\partial t} + \mathbf{v} \cdot \nabla e_\omega = -\frac{e_\omega - e_\omega^0}{\tau_\omega} \]

where \( e_\omega \) is the desired distribution function, \( \omega \) is the phonon frequency, \( e_\omega^0 \) is the local equilibrium distribution function, \( \mathbf{v} \) is the group velocity, \( \tau_\omega \) is the frequency-dependent relaxation time, \( t \) is time and \( \mathbf{r} \) is the spatial coordinate. The BTE is challenging to solve because of its high dimensionality, particularly in a large computational domain with multiple spatial dimensions. Previous works have solved the multidimensional BTE using a number of numerical techniques, such as discrete ordinates\(^{2,3}\), Monte Carlo (MC) methods\(^{2}\), and a two-flux method with a coupled electron solver\(^ {23}\). However, for the semi-infinite domain here, these methods are either prohibitively computationally costly or contain simplifying approximations that may not be accurate.

To overcome this challenge, we solve the frequency-dependent BTE using a novel deviational MC method\(^ {24}\). MC techniques solve the BTE by simulating the advection and scattering of computational particles, denoted as phonon bundles, which represent some number of actual phonons\(^ {25}\). Variance reduction in deviational MC methods is achieved by simulating only the deviation from a known Bose–Einstein distribution, thereby incorporating deterministic information and reducing the variance. These algorithms enable the BTE to be rigorously solved many orders of magnitude faster than other numerical approaches, with little stochastic noise and minimal memory requirements (see Methods).

The steady-state temperature profile of the transistor with this heating distribution at different device operating temperatures for the same input power used in the measurements is shown in Fig. 2. Temperature variations in the direction perpendicular to the plane of the figure are neglected, as they are small compared to the in-plane variations (Supplementary Information). Note that in this context temperature is defined in terms of equivalent equilibrium temperature\(^ {26}\), or the temperature of a thermal distribution with the same energy density as the actual distribution, owing to the strong non-equilibrium in the absence of scattering. The phonon emission region corresponding to Joule heating is modelled as a volumetric generation term, with the size and location determined by electrical requirements (see Methods).

At 300 K (Fig. 2a) the temperature in the channel is almost the same as the ambient. The temperature rise in the channel is still

---

**Figure 1** Optical micrographs and noise measurements of a low-noise amplifier integrated circuit. **a**, Optical microscopy image showing part of the 4–16 GHz MMIC amplifier. In the image are the three InP HEMTs, microstrip lines, NiCr thin-film resistors (TFRs) and metal–insulator–metal (MIM) capacitors. **b**, Scanning electron micrograph image of an InP HEMT. Inset: schematic view of the HEMT cross-section. The HEMT is grown on an InP substrate with an InAlAs buffer for lattice matching, an InGaAs channel, and a doped InGaAs barrier to provide ohmic contacts. **c**, Measured noise temperature (symbols) and predicted linear scaling (dashed line) versus low-noise amplifier (LNA) chassis temperature of the device at 12 GHz. The measured noise temperature decreases linearly with a decrease in the chassis temperature until about 40 K, below which it saturates, deviating from the ideal prediction. **d**, Extracted drain temperature (symbols) and extrapolated linear scaling (dashed line) versus LNA chassis temperature. The measured drain temperature exhibits the same saturation for chassis temperatures below 40 K. The error bars are obtained by propagating the uncertainty in measurements from the noise figure analyser through the circuit model used to determine the noise temperature.
The observation of strong self-heating at cryogenic temperatures may be initially puzzling if diffusion theory is used to model the heat dissipation. For example, diffusion theory based on the bulk thermal conductivity of InP predicts the temperature rise at $T_e = 10\,\text{K}$ to be around $0.01\,\text{K}$, orders of magnitude smaller than the actual values in Fig. 3b. Further, assuming a smaller effective thermal conductivity $T_e$, further cooling has not decreased the channel temperature.

We plot the calculated peak channel temperature versus the device operating temperature for different power inputs in Fig. 3a. The figure shows that the temperature of the channel saturates when the device operating temperature decreases below approximately $30\,\text{K}$ for every non-zero power input, a result that is consistent with noise measurements. We also compare the measured intrinsic temperature rises with those calculated from our MC simulations. To obtain the measured intrinsic temperature rises, we use the noise model to determine the increase in intrinsic temperature $\Delta T = T_e - T_a$ that is necessary to explain the additional noise above the ideal prediction. The MC temperature rises are taken to be the difference between the peak channel and the device operating temperatures. These two results are plotted in Fig. 3b, demonstrating excellent agreement and showing that the measured trend in intrinsic temperature rise is accurately captured by our MC simulations without any adjustable parameters.

With this insight, the self-heating can be explained as a consequence of heat conduction by phonon black-body radiation. In the radiation regime, the heat flux is largely determined by the number of available heat conduction channels. At sufficiently low temperatures, the number of available channels decreases rapidly, as the occupied thermal phonon population is restricted to low-energy modes following the Bose–Einstein distribution. Macroscopically, this restriction in phonon occupation corresponds to a decrease in specific heat as $T^3$, according to the Debye model. Therefore, the total amount of heat that can be dissipated by the occupied phonon states decreases rapidly with temperature. Below a device operating temperature of approximately $30\,\text{K}$, the available low-energy phonons are unable to dissipate the heat of the electron
NATURE MATERIALS

The sound velocities are the sound velocities. The theoretical and... for InP (ref.28), where \( T \) is the dissipated heat flux per unit length, \( \sigma \) is the equivalent of the Stefan–Boltzmann constant for phonons in InP, \( q = 29 \text{ mW mm}^{-1} \) is the dissipated heat flux per unit length, and \( T_c \) is the device operating temperature. The average velocity \( v_{29} = \sum v_i/3 \approx 3,000 \text{ m s}^{-1} \), where \( v_i \) are the sound velocities of the various phonon branches \( i \) for InP (ref. 27). We obtain \( w \) by computing the weighted area of the phonon emission region from the electrical MC simulations, where the weight describes the relative fraction of emission events occurring in a certain volume, and equating this area with \( \pi R^2 \), where \( R \) is the radius of a circular emission region in the plane of the device. The emitting area per unit length is then \( w = 2\pi R = 74 \text{ nm} \). We plot this estimate of \( \Delta T \) along with the measured and MC temperature rises in Fig. 3b and observe good agreement, confirming the origin of the self-heating.

The radiation mechanism for thermal resistance identified here is distinctly different from those identified previously in electronic devices. In many devices, thermal resistance is due to phonon scattering by defects, such as point defects and dislocations, and interfaces separating different materials, such as the interface between an epitaxial layer and a substrate, followed by a high-mobility channel of InGaAs for electron transport. The HEMT is fabricated on a 4-inch InP wafer. First, a buffer layer of lattice-matched InAlAs is grown on the substrate, followed by a high-mobility channel of InGaAs for electron transport. The mobility is 12,000 cm\(^2\) V\(^{-1}\) s\(^{-1}\) at room temperature and increases to 60,000 cm\(^2\) V\(^{-1}\) s\(^{-1}\) when the device is cooled to 20 K. The channel electrons originate from a delta doping layer embedded in a barrier layer grown on top of the channel. A highly doped InGaAs layer is formed at the contact areas to create low-resistance ohmic contacts to the device. The ohmic contacts are formed from a Ni/Ge/Au stack, and an annealing step provides additional doping from the Ge

**Methods**

**InP HEMT fabrication and characterization.** The HEMT is fabricated on a 4-inch InP wafer. First, a buffer layer of lattice-matched InAlAs is grown on the substrate, followed by a high-mobility channel of InGaAs for electron transport. The mobility is 12,000 cm\(^2\) V\(^{-1}\) s\(^{-1}\) at room temperature and increases to 60,000 cm\(^2\) V\(^{-1}\) s\(^{-1}\) when the device is cooled to 20 K. The channel electrons originate from a delta doping layer embedded in a barrier layer grown on top of the channel. A highly doped InGaAs layer is formed at the contact areas to create low-resistance ohmic contacts to the device. The ohmic contacts are formed from a Ni/Ge/Au stack, and an annealing step provides additional doping from the Ge
to further reduce the ohmic contact resistance to 0.04 Ω mm. The cap layer is etched away at the gate area, and a 130 nm mushroom-shaped gate is deposited on top of the barrier layer. The gate stack is Ti/Pt/Au, and the mushroom shape gives a large area and is passivated with a 60 nm SiN layer. Air bridges and contact pads are formed by gold plating. Finally, the InP substrate is thinned down to 100 μm to facilitate device dicing. The total size of a HEMT chip is 420 μm × 430 μm.

The InP HEMT process was integrated into a monolithic microwave integrated circuit (MMIC). Added to the InP HEMT process is the formation of microstrip lines, NiCr thin-film resistors (TFRs) of 50 Ω sq−1, metal-insulator-metal (MIM) capacitors with 390 pF mm−2, via-hole grounding, and a back-side gold-plated ground plane. The wafer thickness, after thinning, is reduced to 75 μm to allow thinner microstrip lines. The via-holes were dry etched to a diameter of 45 μm. The size of the MMIC low-noise amplifier (LNA) is 0.75 mm × 2 mm. The MMIC was mounted together with an external input matching network, bias circuit and stabilizing capacitors and resistors in a gold-plated module with SMA connectors. To avoid oscillations, the microstrip and MMIC cavities were dimensioned to have cutoff frequencies far above the highest MMIC gain frequency.

The MMIC LNA was mounted in a cryogenic measurement setup to measure the cryogenic noise temperature dependence. Temperature sensors were mounted on both the attenuator and LNA chassis. At 10 K the measurement uncertainty, using a cold attenuator, was about 0.3 K (ref. 32). While sweeping the temperature, the Agilent figure Analyzer was set to CW mode at 12 GHz to record noise temperature as fast as possible and avoid temperature drift during measurement. The optimum low-noise bias for the InP HEMTs in this LNA was 0.87 V and 33.3 mA mm−1, yielding a power input of 29 mW mm−1 (referenced to the transistor terminals). Along with the recorded noise temperature, the physical temperature of the LNA and cold attenuator was monitored and stored both before and after each noise measurement. The noise parameter extraction was performed as described in ref. 19. Additional details are available in the Supplementary Information.

Phonon BTE simulations. We solve the phonon BTE using variance-reduced Monte Carlo algorithms that simulate the advection and scattering of computational particles in a domain33. In these algorithms a computational particle represents a fixed amount of deviational energy from an equilibrium Bose–Einstein distribution rather than a constant number of phonons as in previous algorithms. The algorithm directly solves for the steady-state solution of the BTE by sequentially advecting and scattering the phonons in the domain. The temperature and heat flux from the phonons is obtained by spatially discretizing the domain and recording the time spent and the velocity of each phonon in a particular cell. Performing an average over many phonons yields the steady-state heat flux and temperature profile of the device.

The structure is taken to be two dimensional, with temperature variations in the perpendicular direction neglected. On top of the device, we enforce an adiabatic condition in which phonons are reflected back with equal probability in the hemisphere defined by the boundary normal because the gate and passivation layer are isothermal with the top of the device layer (Supplementary Information). The InP substrate is treated as semi-infinite, with phonons that have travelled sufficiently far away from the channel being deleted. We verified that the precise value of the cutoff radius has negligible impact on the simulation results. We assume that the entire domain is composed of InP, neglecting all variations in material composition and interfaces. The heat generation in the channel is incorporated as a volumetric heat generation process. Phonons are generated in a specified location and volume determined from the electrical MC simulations with random velocity directions on a unit sphere and an equilibrium distribution in frequency. The number of generated phonons N is related to the volumetric heat generation q by q = EN, where E is the specified effective deviational energy of a computational particle.

We model the dispersion for the InP substrate with a Born–von-Karman model which includes a single polarization of the form ω = γω0 sin(πq/a0), where ω0 and γ are the phonon frequency and wavevector, respectively. Here q0 = (6πn0)1/3 is the Debye wavevector to enforce the correct number of phonon modes, n0 = 4πa2 is the density of primitive cells, and n0 = 5.8687 nm is the lattice constant of InP and ω0 = 2πν0/τ, where the average sound velocity is υ0 = 3,000 m s−1. The relaxation time τ is of the form τ = 4πa2/3AT exp(−βT), where T is the absolute temperature, and AT = 4 × 10−6 °K−1 and θ = 210 K are chosen to approximately match the temperature-dependent thermal conductivity of InP.

Electrical MC simulations. The electron and the consequent phonon emissions have been simulated using a two-dimensional semiclassical model. This model correctly reproduces the a.c., d.c. and noise behaviour of InP HEMTs both at 300 K (ref. 33) and at cryogenic temperatures34–36. We obtain the spatial distribution of dissipated power and, hence, phonon generation for the phonon BTE simulations by recording the energy balance of the electrons at each mesh point over 0.1 ns. We find that for low gate bias most of the phonon emissions take place in the channel at the drain edge of the recess, consistent with previous reports37. We performed the electrical simulations at temperatures of 300 K and 30 K. We find the distribution of dissipated power to be constant in the low-temperature range below 150 K, where electron transport is completely ballistic under the gate, so that a self-consistent approach with the phonon BTE is not necessary under these conditions.

Received 11 June 2014; accepted 3 October 2014; published online 10 November 2014

References

Acknowledgements
The authors thank S. Weinreb for useful discussions. I.I-d-l-T. and J.M. were partially supported by the Spanish MINECO through project TEC2013-41640-R and by the Consejería de Educación de la Junta de Castilla y León through project SA052U13. J.S., N.W., P.A.N. and J.G. were supported by the GigaHertz Centre in a joint research project financed by the Swedish Governmental Agency of Innovation Systems (VINNOVA), Chalmers University of Technology, Omnisys Instruments AB, Wasa Millimeter Wave, Low Noise Factory and SP Technical Research Institute of Sweden. A.J.M. was supported by a Caltech startup fund and by the National Science Foundation under Grant no. CAREER CBET 1254213.

Additional information
Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to A.J.M.

Competing financial interests
The authors declare no competing financial interests.