

Cooling of Semiconductor Devices via Quantum Tunneling

Amin Reihani^{1,§}, Zheng Li^{2,§}, Jian Guan¹, Yuxuan Luan¹, Shen Yan¹, Jin Xue,² Edgar Meyhofer^{1,*},
Pramod Reddy^{1,3,†} and Rajeev J. Ram^{2,‡}

¹*Department of Mechanical Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA*

²*Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA*

³*Department of Materials Science and Engineering, University of Michigan, Ann Arbor, Michigan 48109, USA*



(Received 27 June 2024; revised 12 October 2024; accepted 19 November 2024; published 27 December 2024)

Classical transport of electrons and holes in nanoscale devices leads to heating that severely limits performance, reliability, and efficiency. In contrast, recent theory suggests that interband quantum tunneling and subsequent thermalization of carriers with the lattice results in local cooling of devices. However, internal cooling in nanoscale devices is largely unexplored. Here, using a novel scanning thermal microscopy technique with millikelvin temperature resolution and nanometer spatial resolution, we directly record the cross-sectional temperature in functional InGaAs tunnel diodes. Our measurements reveal large, localized cooling of 2–3 W/cm² at the tunnel junction, which is in quantitative agreement with the bipolar Peltier process associated with interband tunneling. These advances hold significant potential for integration into electronic and energy conversion devices and improving their performance.

DOI: 10.1103/PhysRevLett.133.266301

Heat generation in semiconductor devices [1,2] poses major challenges to their performance and reliability [3,4]. To mitigate detrimental heating effects, researchers have sought to integrate external cooling devices such as heterostructured [5] or thin-film thermoelectric [6] coolers. These approaches are useful in some circumstances. However, when heating is localized to nanoscale regions, internal cooling via quantum tunneling can offer significant advantages [7,8]. Specifically, recent computational work [7] has suggested that such local cooling can be achieved in tunnel diodes [9] via interband tunneling.

In addition to high speed oscillators and amplifiers [10], tunnel diodes are now being actively studied for applications in low-leakage tunnel field-effect transistors [11–14], multijunction (tandem) photovoltaics [15–17], and photoelectrochemical cells [18,19]. Although many of the above examples utilize tunnel diodes for energy applications, there has been little research on the energy conversion processes in tunnel diodes. Understanding these processes is critical as interband quantum tunneling in diodes generates spatially separated Peltier cooling and heating, which can potentially lead to localized cooling of the lattice [7]. While the local heating and cooling effects associated with quantum mechanical tunneling processes have been observed in some unipolar devices [20–24], internal cooling has not been observed in any bipolar device, including tunnel diodes.

In this Letter, we leverage recent advances in scanning thermal microscopy [25,26] and directly record temperature fields with nanometer spatial and millikelvin temperature resolution in cross sections of cleaved tunnel diodes [Figs. 1(a) and 1(b)] using custom-fabricated scanning thermal probes [STPs, Fig. 1(d)]. This allowed us to systematically evaluate the hypothesis that reverse-biased tunnel diodes feature electronic cooling and suppress temperature rises in the semiconductor lattice.

To understand the origin of the localized cooling, it is instructive to consider the current-voltage (I-V) characteristics of the tunnel diode employed in this work [Fig. 1(c)]. A key feature of the I-V curve is the negative differential resistance at a bias of ~0.35 V and the operating regions of interest for this device are labeled *i*, *ii*, *iii* in Fig. 1(c). In these diagrams, V is the applied bias ($V > 0$ is forward bias), q is the elementary charge, $E_{f,n/p}$ are the quasi-Fermi levels, and $\Pi_{n/p}$ are the Peltier coefficients of *n* and *p* regions. In regime *i*, where the diode is forward-biased [see Fig. 1(e)], the average energy of the carriers is higher on the *n* side than that on the *p* side. Therefore, the transported carriers deposit heat into the lattice at the junction. Conversely, under a small reverse bias [region *ii* in Fig. 1(c)], the average energy of charge carriers on the *p* side is smaller than that on the *n* side. Therefore, carriers tunneling from the *p* to the *n* side are expected to absorb energy from the lattice at the junction [see Fig. 1(f)]. The thermal energy removed from the lattice leads to local cooling at the junction and is later deposited as Peltier heat on the metal contacts or in the bulk region as indicated by red arrows in Fig. 1(f). Finally, in regime *iii* [Fig. 1(c)], the junction is again heated [Fig. 1(g)] as the average energy of

*Contact author: meyhofe@umich.edu

†Contact author: pramodr@umich.edu

‡Contact author: rajeev@mit.edu

§These authors contributed equally to this work.

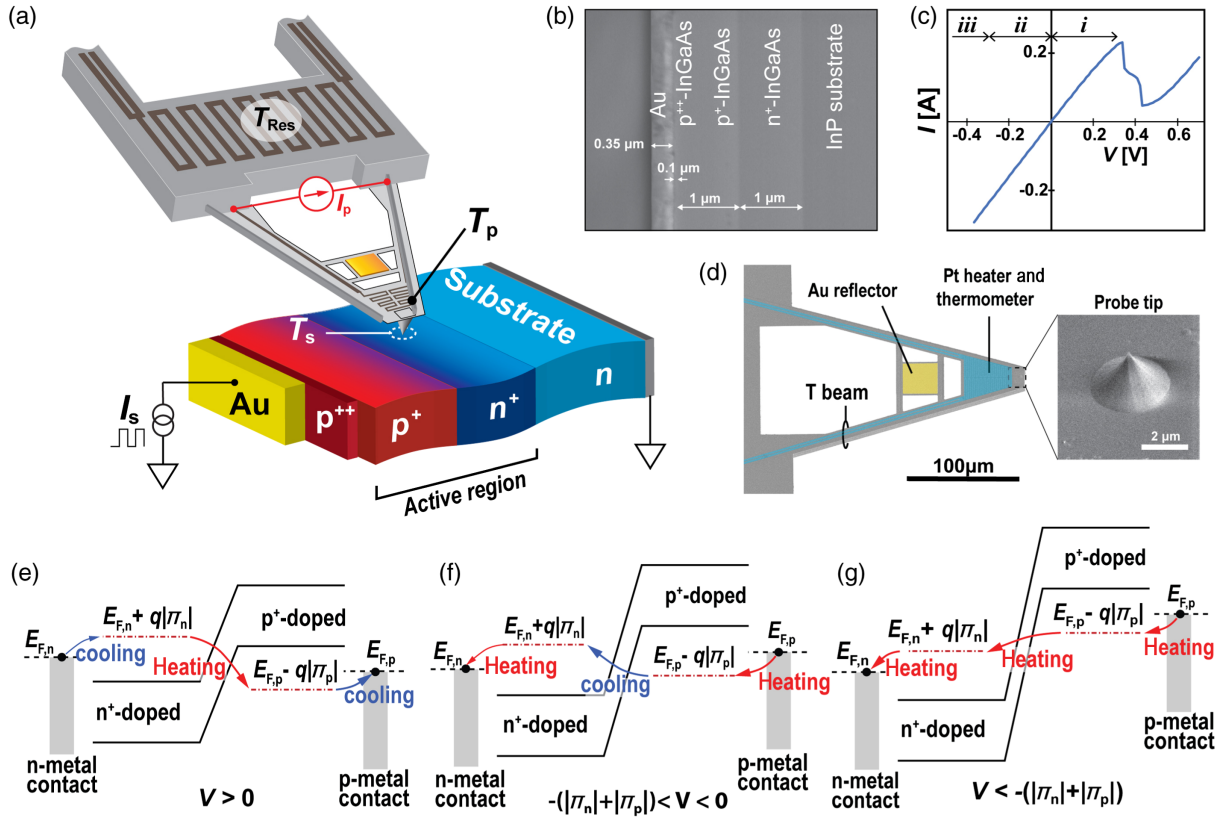


FIG. 1. Experimental setup and physical principle of energy transfer via interband tunneling. (a) Schematic description of the scanning thermal microscopy measurements on the cross section of a cleaved functional InGaAs tunnel diode with a smooth cross section. A square wave of current is supplied to the tunnel diode and the lattice temperature as a function of time is measured at different locations using scanning thermal microscopy. (b) Scanning electron microscope image of the cleaved cross section of the tunnel diode. (c) Current-voltage characteristic of the tunnel diode used in the experiments. Various operating regimes of interest are labeled *i*, *ii*, *iii* and discussed below. (d) Scanning electron microscope image of the scanning thermal probe (Pt heater and Au reflector are false colored). (e)–(g) Schematic description of energy transport due to interband tunneling in the operating regimes identified in panel (c). Solid black lines indicate the conduction and valence bands, black dashed lines indicate the quasi-Fermi levels, and the red dashed-dotted lines indicate the mean energy of charge carriers in the bulk. (e) Band structure in regime *i*, where forward bias causes junction heating. (f) Regime *ii*, small reverse bias results in junction cooling. (g) Regime *iii*, large reverse bias causing junction heating. In (e)–(g), V is the applied bias, q is the charge, $E_{f,n/p}$ are the quasi-Fermi levels, Π_n and Π_p are the Peltier coefficients of the *n* and *p* regions, respectively.

charge carriers in the *p* side becomes larger than those on the *n* side. We also note that in regime *ii* the heat deposited per charge carrier ($\Delta\Pi_{p-elec}$) at the ohmic contact between the electrode and the *p*-type region equals the difference in Peltier coefficients of the *p* region (Π_p) and the contact electrode (Π_{elec}), i.e., $\Delta\Pi_{p-elec} = \Pi_p - \Pi_{elec}$.

We began our experiments by cleaving a tunnel diode made of $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ (see the Supplemental Material [27] for fabrication details) to access the cross section of the device [see Fig. 1(a)]. The cross section of the device [obtained using a scanning electron microscope, Fig. 1(b)] is flat and smooth with a root mean square roughness of ~ 0.6 nm (see Sec. S2 of [27]). The width of the depletion region, measured via density phase-contrast transmission electron microscopy, was < 20 nm (see Sec. S3 of [27]). To conduct detailed thermal measurements, the STP and the tunnel diode were incorporated into an ultrahigh vacuum

scanning probe microscope and carefully aligned with respect to each other [see Fig. 1(a)].

To identify the active region of the device [the *p*⁺ and *n*⁺ regions of the device, see Fig. 1(a)] and the location of the *p*-*n* junction, we first measured (Sec. S6 of [27]) both the sample topography [Fig. 2(a)] and tip sample thermal contact resistance [R_{TS} , Fig. 2(b)], which reflects the local thermal conductivity. The topographical data clearly delineate the electrode (elevated region) and edge of the electrode. Also, the cross section is flat and no contrast is observed between different device layers. However, the R_{TS} map clearly shows contrast between the InP substrate and the InGaAs active region due to their different thermal conductivities ($5 \text{ Wm}^{-1} \text{ K}^{-1}$ for $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ [36] and $68 \text{ Wm}^{-1} \text{ K}^{-1}$ for InP [37] at room temperature). Therefore, the R_{TS} and topography maps combined with the known device dimensions reveal the active region and location of the *p*-*n* junction (Sec. 7 of [27]).

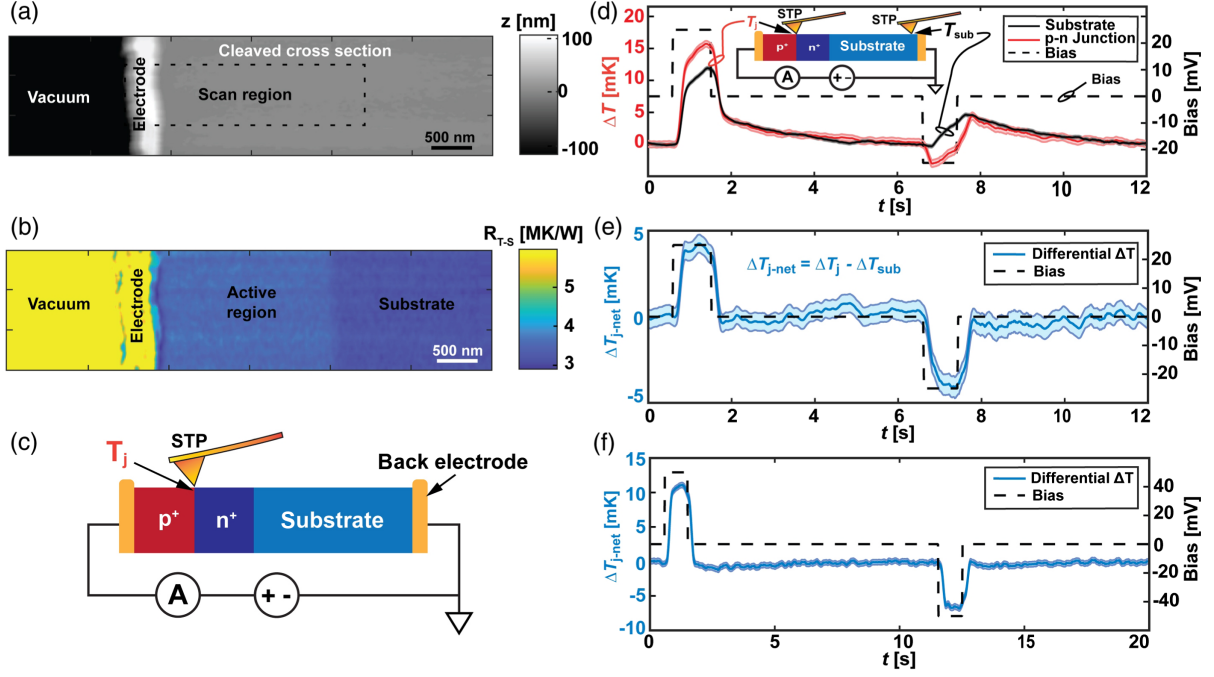


FIG. 2. Junction temperature measurement under three-level bias pulse. (a) Measured topography of the cleaved cross section of the tunnel diode. The active region of the diode (to the right of the electrode) features a flat smooth surface. (b) The tip sample thermal resistance (R_{TS}) map obtained from scanning thermal microscopy scan of the device cross section under zero bias. The contrast in R_{TS} clearly demarcates the InGaAs active region (low thermal conductivity) from the InP substrate (high thermal conductivity). (c) Schematic of the measurement procedure shows how the junction temperature (T_j) is recorded by placing the probe at the tunnel junction. (d) Junction (T_j) and substrate (T_{sub}) temperatures for a ± 25 mV pulse (inset shows location where substrate temperature is measured). The plots were obtained by averaging data from over 10 000 cycles. The solid line indicates the mean value of the signal and the shaded area indicates the 95% confidence interval of the mean. As can be seen, both the junction and substrate temperatures feature slow common-mode temperature changes due to device carrier heating. (e), (f) Net temperature change of the junction obtained by subtracting the substrate temperature from the measured junction temperature for (e) ± 25 mV pulse and (f) ± 50 mV pulse.

To explore the hypothesized bias-dependent heating and cooling we first placed the STP tip [Fig. 2(c)] in close proximity (± 25 nm) to the p-n junction and applied a three-level bias pulse to the tunnel diode (this pulse scheme attenuates low frequency noise) consisting of voltages $+V_B$, 0, and $-V_B$. Here, the bias magnitude was chosen to be $V_B = 25$ mV with a pulse duration of 1 s and a zero bias for a duration of 5 s in the first experiment, and $V_B = 50$ mV with a pulse duration of 1 s and a zero bias duration of 10 s in the second experiment, so as to cover the regimes *i* and *ii* of the I-V relationship discussed above. We note that the zero bias duration was chosen sufficiently long to allow the device and carrier temperatures to relax to their initial values after each pulse (see Sec. 8 of [27]). While applying the bias voltage, the junction temperature (T_j) was measured by first monitoring the temperature of the STP tip (T_P) using the Pt thermometer integrated into the STP and by relating T_P to T_j via the approach described in Refs. [25,26] (see Sec. S6 of [27]).

To improve the temperature resolution of our measurements, we averaged the time-domain temperature (T_j) obtained from many pulse cycles. Data obtained by averaging over $\sim 10\,000$ cycles is shown in Fig. 2(d) in

red (data for varying number of cycles is shown in Fig. S13 of [27]). These data show that the temperature of the junction increases upon applying a forward bias and decreases (i.e., the region is cooled) upon applying a reverse bias. However, the observed cooling is less pronounced than the heating. We hypothesized that this asymmetry is likely due to Joule heating in the electrodes and electrical contacts to the chip carrier (see Sec. S4 of [27]). In order to systematically account for the effects of chip carrier heating, we performed an experiment where we measured the electrode temperature of the device [labeled T_{sub} in the inset of Fig. 2(d)] by placing the STP tip near the back electrode [see Fig. 2(d) and Sec. S4 of [27]]. In this experiment, three-level pulses identical to those used in obtaining the T_j data in Fig. 2(d) were applied. The black solid line in Fig. 2(d) shows results from time-domain measurements of T_{sub} (obtained by averaging over 10 000 cycles). Next, we subtracted T_{sub} from T_j to remove the effects of temperature rise of the chip-carrier boundary (T_{sub}) and plotted the net temperature change ($\Delta T_{net} = T_j - T_{sub}$) in Fig. 2(e). These data reveal a clear heating and cooling signal, corresponding to forward- and reverse-bias conditions. Further, the magnitude of cooling

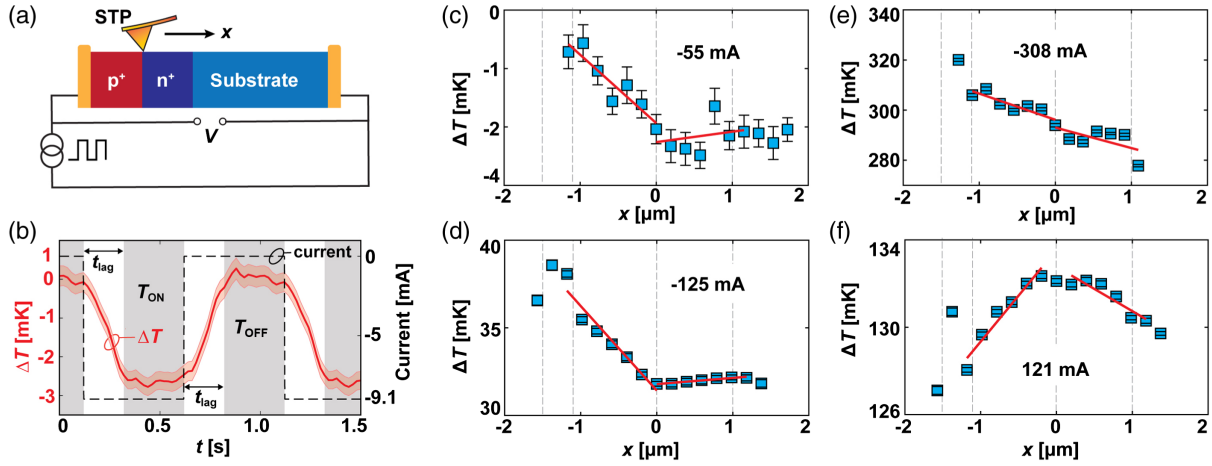


FIG. 3. Temperature profiles measured using scanning thermal microscopy. (a) Schematic describing how line scans were performed in a pixel mode by applying time-dependent two-level current pulses to the tunnel diode. Here, $x = 0$ corresponds to the location of the tunnel junction. (b) Data corresponding to a measurement close to the tunnel junction ($-25 \text{ nm} < x < 25 \text{ nm}$). The time-dependent change in the junction temperature (red solid line) was obtained by averaging data from 10 000 cycles for a current pulse of -9.1 mA (black dashed line). Shaded area indicates the 95% confidence interval of the mean. These data clearly show cooling at the junction by $\sim 2.7 \text{ mK}$ when a current pulse of -9.1 mA is applied. (c)–(f) Measured temperature change as a function of position (solid squares with corresponding 95% confidence interval) obtained in the active region of the tunnel diode for various diode currents. Linear fits to the temperature profiles in the p and n regions are shown by red lines.

and heating for biases of $\pm 20 \text{ mV}$ is found to be almost equal ($\sim 4 \text{ mK}$). Next, we performed measurements at higher biases ($V_B = 50 \text{ mV}$). Data from these experiments [Fig. 2(f)] reveal that heating ($\sim 11 \text{ mK}$) and cooling ($\sim 7 \text{ mK}$) become unequal at larger biases due to more pronounced Joule heating (both in the device as well as in the top electrode). The results for $V_B = 100 \text{ mV}$ are included in the Sec. S8 of [27] and indicate a heating of $\sim 21 \text{ mK}$ and cooling of $\sim 14 \text{ mK}$.

Next, we studied the temperature distribution in the device as a function of the device current (a feedback controlled current source was employed to ensure both a stable current and a current density that is invariant with time) by modulating the electric current (0.5 s on, 0.5 s off) to the tunnel diode. A scan of the sample temperature, along a line, was obtained by placing the STP at several predefined pixel locations [x in Fig. 3(a)] and measuring the surface temperature at each pixel over several thousand pulse cycles. Several scans with 16 pixels per line (pixel spacing of $\sim 180 \text{ nm}$) were obtained. Further, 64-pixel line scans (45 nm spacing) were also obtained at certain bias values to confirm the efficacy of 16-pixel scans in resolving the temperature profile (see Sec. S7.2 of [27]). At each pixel data from 5000 cycles were averaged, while the line scans were repeated multiple times for each current value (see [27]).

The time-dependent surface temperature obtained by averaging 10 000 cycles (data from two line scans) at a pixel closest to the junction is shown in Fig. 3(b) for a current pulse of -9.1 mA (black dashed line). The obtained data (red solid line) show surface cooling of $\sim 2.7 \text{ mK}$ when the pulse is applied. The surface temperature change

in response to the applied current pulse at each pixel is simply calculated from $\Delta \bar{T}(x) = \bar{T}_{\text{on}}(x) - \bar{T}_{\text{off}}(x)$, where $\bar{T}$ represents the time average of the measured temperature. The temperature profiles, $\Delta T(x)$, are shown for several values of current pulse amplitude in Figs. 3(c)–3(f) (data for other current pulse magnitudes shown in Fig. S16 of [27]). For forward-biased measurements the temperature profile is convex at the junction, indicating the existence of a heat source at the junction. Conversely, in reverse-biased measurements with current in the range of -55 mA to -308 mA , the temperature profile is concave at the junction, indicating local cooling. Next, the junction heat flux (Q_j) was estimated by performing a linear fit to the temperatures in the p- and n-doped regions for estimating the temperature gradients $(\partial T / \partial x)_p$, $(\partial T / \partial x)_n$ in the p and n regions, respectively [see Figs. 3(c)–3(f)] and by employing Fourier's law [38] to a small control volume enclosing the junction,

$$Q_j = k \left[\left(\frac{\partial T}{\partial x} \right)_p - \left(\frac{\partial T}{\partial x} \right)_n \right], \quad (1)$$

where $k = 5 \text{ W m}^{-1} \text{ K}^{-1}$ for $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ at room temperature [36]. The junction heat flux vs device current is presented in Fig. 4(a) and clearly shows cooling fluxes for a range of reverse biases.

Next, we modeled carrier transport in tunnel diodes using an extension of Kane's model for electron transmission as described in [7] and briefly summarized in Sec. S12 of [27]. In this approach the junction heat flux (Q_j) is given by

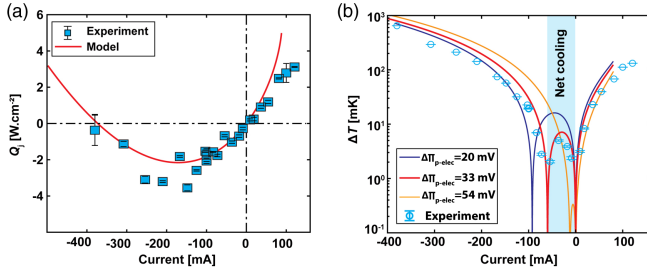


FIG. 4. Comparisons of the experimental and modeled junction heat flux and temperature. (a) Junction heat flux due to interband tunneling as a function of device current, obtained from both *in situ* scanning thermal microscopy imaging and modeling. (b) Junction temperature change as a function of device current obtained from *in situ* scanning thermal microscopy experiments as well as a 3D finite-element heat conduction model for different values of $\Delta\Pi_{p\text{-elec}}$. Good agreement with the experimental data is obtained for $\Delta\Pi_{p\text{-elec}} = 33$ mV.

$$Q_j = \frac{1}{q} \int dE [q|\Pi_n| - q|\Pi_p| + (E_{F,n} - E_{F,p})] \frac{\partial J}{\partial E}, \quad (2)$$

where J is the current density. Results from these calculations are shown in Fig. 4(a) (solid line) and are in good agreement with the experimental data. We also developed a 3D finite element heat conduction model, where the junction heat flux (i.e., computed values of Q_j) was utilized as a localized source term at the junction to obtain the internal temperature distribution of the diode (see Sec. S11 of [27]). Results from both our experiments and the model indicate that the maximum cooling power occurs at a device current (I) of ~ -170 mA, corresponding to a junction current density (J) of ~ -84 A/cm² (see Sec. S10 of [27] for a description of how J is estimated). The maximum cooling power observed from the experiment is ~ 3.5 W/cm², which is in good agreement with the model's prediction of 2.17 W/cm². In addition, in Fig. 4(b) we show the measured junction temperature change vs device current along with modeling results for three different values of $\Delta\Pi_{p\text{-elec}}$ (20 mV, 33 mV, and 53.7 mV). We note that the exact value of $\Delta\Pi_{p\text{-elec}}$ depends on the precise alloy composition and deposition or annealing parameters of the top ohmic contact, which determine its transport parameters [20,21]. A good match between the model and experiment is observed for $\Delta\Pi_{p\text{-elec}} = 33$ mV, indicating that net cooling at the junction occurs at small reverse-bias currents of $-70 \text{ mA} < I < 0 \text{ mA}$, corresponding to current densities of $-34.3 \text{ A/cm}^2 < J < 0 \text{ A/cm}^2$. The model overestimates the temperature change by $\sim 40\%$ in both the cooling and heating regions, likely due to simplification of heat transfer paths in the model (see Sec. S11 of [27]) that were made to eliminate multiple fitting parameters and increasing the interpretability of the model.

Our measurements of temperature fields within functional electronic devices with nanometer resolution will

significantly advance understanding of nanoscale energy conversion and dissipation processes. The demonstration of cooling in reverse-biased tunnel diodes paves the way for exploring and integrating internal cooling approaches into semiconductor devices for enhanced performance in various applications.

Acknowledgments—P. R. and E. M. acknowledge support from DOE-BES through a grant from the Scanning Probe Microscopy Division under Award No. DESC0004871 (Scanning Thermal Microscopy). P. R. and E. M. acknowledge support from the Office of Naval Research under Grant No. N00014-24-1-2132 (Instrumentation) and the National Science Foundation under Grant No. CBET 2232201 (Study of Tunnel Junctions). The authors acknowledge Shaomin Wu in Professor John Bowers' research group at UCSB for providing the epitaxial material.

- [1] A. Majumdar, *Nat. Nanotechnol.* **4**, 214 (2009).
- [2] E. Pop, *Nano Res.* **3**, 147 (2010).
- [3] E. Pop, S. Sinha, and K. E. Goodson, *Proc. IEEE* **94**, 1587 (2006).
- [4] J. Schlee, J. Mateos, I. Íñiguez-de-la-Torre, N. Wadefalk, P.-A. Nilsson, J. Grahn, and A. Minnich, *Nat. Mater.* **14**, 187 (2015).
- [5] X. F. Fan, G. H. Zeng, C. LaBounty, J. E. Bowers, E. Croke, C. C. Ahn, S. Huxtable, A. Majumdar, and A. Shakouri, *Appl. Phys. Lett.* **78**, 1580 (2001).
- [6] I. Chowdhury, R. Prasher, K. Lofgreen, G. Chrysler, S. Narasimhan, R. Mahajan, D. Koester, R. Alley, and R. Venkatasubramanian, *Nat. Nanotechnol.* **4**, 235 (2009).
- [7] Z. Li, J. Xue, and R. J. Ram, *J. Appl. Phys.* **126**, 165107 (2019).
- [8] K. P. Pipe, R. J. Ram, and A. Shakouri, *Phys. Rev. B* **66**, 125316 (2002).
- [9] L. Esaki, *Phys. Rev.* **109**, 603 (1958).
- [10] S. M. Sze, *Semiconductor Devices, Physics and Technology*, 2nd ed. (Wiley, New York, 2002).
- [11] L. Britnell *et al.*, *Science* **335**, 947 (2012).
- [12] A. M. Ionescu and H. Riel, *Nature (London)* **479**, 329 (2011).
- [13] D. Sarkar, X. J. Xie, W. Liu, W. Cao, J. H. Kang, Y. J. Gong, S. Kraemer, P. M. Ajayan, and K. Banerjee, *Nature (London)* **526**, 91 (2015).
- [14] T. N. Theis and P. M. Solomon, *Science* **327**, 1600 (2010).
- [15] J. F. Geisz, R. M. France, K. L. Schulte, M. A. Steiner, A. G. Norman, H. L. Guthrey, M. R. Young, T. Song, and T. Moriarty, *Nat. Energy* **5**, 326 (2020).
- [16] S. Essig *et al.*, *Nat. Energy* **2**, 17144 (2017).
- [17] R. Cariou *et al.*, *Nat. Energy* **3**, 606 (2018).
- [18] J. L. Young, M. A. Steiner, H. Doscher, R. M. France, J. A. Turner, and T. G. Deutsch, *Nat. Energy* **2**, 17028 (2017).
- [19] O. Khaselev and J. A. Turner, *Science* **280**, 425 (1998).

- [20] I. Bat'ko and M. Bat'kova, *Eur. Phys. J.* **31**, 191 (2005).
- [21] L. Cui, R. Miao, K. Wang, D. Thompson, L. A. Zotti, J. C. Cuevas, E. Meyhofer, and P. Reddy, *Nat. Nanotechnol.* **13**, 122 (2018).
- [22] T. S. Fisher and D. G. Walker, *J. Heat Transfer* **124**, 954 (2002).
- [23] P. J. Koppinen and I. J. Maasilta, *Phys. Rev. Lett.* **102**, 165502 (2009).
- [24] M. Walter *et al.*, *Nat. Mater.* **10**, 742 (2011).
- [25] A. Reihani, S. Yan, Y. X. Luan, R. Mittapally, E. Meyhofer, and P. Reddy, *Appl. Phys. Lett.* **118**, 163102 (2021).
- [26] A. Reihani, Y. X. Luan, S. Yan, J. W. Lim, E. Meyhofer, and P. Reddy, *ACS Nano* **16**, 939 (2022).
- [27] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.133.266301> for additional details regarding device and scanning probe fabrication, characterization of the device, scanning thermal microscopy, and heat transfer and exchange modeling in the device, which includes Refs. [28–35].
- [28] N. Shibata, S. D. Findlay, H. Sasaki, T. Matsumoto, H. Sawada, Y. Kohno, S. Otomo, R. Minato, and Y. Ikuhara, *Sci. Rep.* **5**, 10040 (2015).
- [29] K. Kim, W. Jeong, W. Lee, and P. Reddy, *ACS Nano* **6**, 4248 (2012).
- [30] A. Reihani, J. W. Lim, D. K. Fork, E. Meyhofer, and P. Reddy, *ACS Sens.* **6**, 387 (2020).
- [31] A. Pan and C. O. Chui, *J. Appl. Phys.* **116**, 054508 (2014).
- [32] A. Schenk and S. Sant, *J. Appl. Phys.* **128**, 014502 (2020).
- [33] J. B. Krieger, *Ann. Phys. (Leipzig)* **36**, 1 (1966).
- [34] E. O. Kane, *J. Appl. Phys.* **32**, 83 (1961).
- [35] K. H. Goetz, D. Bimberg, H. Jürgensen, J. Selders, A. Solomonov, G. Glinskii, and M. Razeghi, *J. Appl. Phys.* **54**, 4543 (1983).
- [36] S. Adachi, *J. Appl. Phys.* **54**, 1844 (1983).
- [37] J. Jaramillo-Fernandez, E. Chavez-Angel, R. Sanatinia, H. Kataria, S. Anand, S. Lourdudoss, and C. M. Sotomayor-Torres, *CrystEngComm* **19**, 1879 (2017).
- [38] L. M. Jiji, *Heat Conduction*, 3rd ed. (Springer, Berlin, Heidelberg, 2009).