Thermal Characterization of Gallium Arsenide THz Schottky Diodes Heterogeneously Integrated on Silicon Using Thermo-Reflectance Measurements

Souheil Nadri, Student Member, Christopher M. Moore, Student Member, Noah D. Sauber, Student Member, Linli Xie, Student Member, Naser Alijabbari, Member, Michael E. Cyberey, Member, Arthur W. Lichtenberger, N. Scott Barker, Fellow, Mona Zebarjadi, and Robert M. Weikle, II, Fellow

Abstract—This paper presents the first thermal characterization of terahertz Schottky diodes using thermo-reflectance imaging. The devices under test are quasi-vertical gallium arsenide Schottky diodes, heterogeneously integrated on a high resistivity silicon substrate. The measurement technique is non-contact and non-invasive. It relies on the change in light reflectivity with changing material surface temperature. The diode is excited with a voltage pulse synchronized in time with the image acquisition. Heating and cooling temperature profiles and 2-D temperature maps are obtained for 5.5 and 3.5 µm diameter diodes. From the measurements, we extracted the device thermal resistances, junction temperatures, and thermal time-constants. The characterization results are verified against an electrical transient current-temperature measurement based on I-V pulsing. Based on the characterization, RC circuit and finite-element thermal models are developed for this device geometry. This study identified thermal bottlenecks and process/design optimizations, which could improve the power handling of the diodes. The results also indicate that the temperature of the quasi-vertical diodes decays to ambient much faster than previously reported submillimeter-wave Schottky diodes.

Index Terms—Heterogeneous integration, imaging, junction temperature, Schottky diode, thermo-reflectance, thermal characterization, thermal parameters.

I. INTRODUCTION

HIGH-speed semiconductor devices are now subject to high power densities [1] due to advances in fabrication and integrated circuit technologies. Therefore, thermal management has become crucial to ensure that circuits don’t suffer from aging effects [2], performance degradation [3], or catastrophic failures [4]. An example of such a device is the terahertz Schottky diode. It is the technology of choice for producing solid-state power in the terahertz region of the frequency spectrum [5]. To reach higher output powers at frequencies greater than 1 THz, it is necessary to drive the input stage of the frequency multiplier chain with large power levels. Recent progress in power amplifiers has made it possible to achieve watt levels of power at W band frequencies (75-110 GHz). Consequently, power management considerations have become crucial in the design of diode-based frequency multipliers [6].

The implementation of submillimeter-wave Schottky diode devices has evolved significantly over the decades, from a single element whisker-contacted geometry [7], to a planar chip with integrated finger contact [8], and finally to a fully monolithic component that is commonly integrated into thin membrane structures [9], or heterogeneously integrated onto a host substrate such as quartz [10] or silicon [11]. Silicon is favored over quartz for its mechanical robustness and superior thermal properties [12]. This work presents the first thermal characterization study of terahertz quasi-vertical gallium arsenide Schottky diodes heterogeneously integrated on silicon, using visible light thermo-reflectance measurements.

Previous efforts to study the thermal behavior of submillimeter-wave Schottky diodes include thermal modeling [12] [13] [14], electrical methods [15] [16] [17], and imaging techniques [13]. Finite-element thermal simulations are useful to understand the heat flow through a device and to offer a comparison with measurements. However, they often fail to capture all fabricated device behavior, due to unknown material parameters (thermal conductivities, heat capacities, thermal boundary resistances, etc), fabrication non-idealities (metal quality, rough surfaces, etc), and practical measurement conditions. The electrical method based on pulsed I-Vs and S-parameter measurements uses the diode itself as a thermometer. It is a non-contact method that allows the characterization of devices, even if they are packaged (unlike imaging methods) [18]. However, it suffers from many drawbacks such as a poor spatial resolution, pulsed heating, and using an electrical parameter to perform both excitation and sensing functions. Some of these issues have been addressed in [19]. However, electrical methods cannot inherently provide temperature maps, which limits them to only measuring average temperatures [18]. Additionally, they can only provide cooling temperature profiles, since they use low currents for
sensing, and high currents for heating.

The final technique for thermal characterization is thermography, or thermal imaging. A review of the different thermography techniques used for IC characterization can be found in [18]. Among them, infrared imaging is the most widely used. Nevertheless, it suffers from frame rate limitations and a diffraction limited spatial resolution on the order of ten micrometers. Charge-coupled device (CCD) based thermo-reflectance microscopy alleviates these issues by using visible light, which results in sub-micron spatial resolution. This characterization method is based on measuring the relative change in reflectivity of the device surface as a function of change in temperature. This paper presents the thermal characterization of two diameter size (3.5 and 5.5 micrometers) terahertz Schottky diodes, using thermo-reflectance thermometry. Section 2 describes the quasi-vertical diode geometry. Section 3 covers the fundamentals of thermal modeling. Section 4 describes the measurement setup. Sections 5 and 6 are the results from the thermo-reflectance and I-V measurements respectively. Finally, section 7 discusses the results and develops circuit and finite element models.

II. QUASI-VERTICAL SCHOTTKY DIODE GEOMETRY

The quasi-vertical diode process was first developed at the University of Virginia to address series resistance concerns [20]. Furthermore, the diode’s thermal management improves due to the heterogeneous integration of gallium arsenide on silicon, since Si offers a larger thermal conductivity compared to GaAs. The processing steps for fabricating the quasi-vertical diodes under test are thoroughly discussed in [11]. They incorporate an indium gallium arsenide (InGaAs) cap layer to the epitaxy to allow the formation of a low resistance Ti/Pd/Au/Ti metal stack-up ohmic contact that does not require annealing. The metals are evaporated over the entire GaAs surface prior to bonding. After that, SU-8 is used to bond GaAs onto Si using a low temperature (140 °C) curing process. Once bonded to the silicon substrate, the GaAs handle is removed in a nitric acid solution (HNO₃; H₂O₂; DI at 50 °C), followed by a slow citric acid etch (C₆H₅O₇·H₂O₂ at 50 °C). The exposed AlGaAs etch stop layer is removed in hydrofluoric acid, revealing the GaAs device layer field. After removal of the handle GaAs, the diode mesa areas are defined photo-lithographically and formed by a sequence of selective etches that stop on the silicon surface [11]. The remaining steps utilize standard lithographic patterning to form the anode, airbridge finger contact, ohmic contact overlay metallization, and other circuit features on the silicon surface [11]. A scanning electron microscope (SEM) image of a completed quasi-vertical diode is shown in fig. 1(a). Fig. 1(b) is a close-up view of the anode contact for a 5.5 µm diameter diode. The quasi-vertical heterogeneously integrated Schottky diode geometry is summarized in Fig. 2.

III. THERMAL MODELING

A. Finite-Element Simulation

The 3D finite element solution is acquired via the Ansys Mechanical [21] simulation tool. The simulated geometry corresponds to Fig. 1 and Fig. 2. The mesh size is adjusted to 3 µm around the critical features (diode stack and air-bridge). A heat flux is applied directly beneath the diode contact. Its value equals the total power dissipated divided by the anode area. The bottom and side facets of the substrate are set to ambient temperature (293 K). The remaining surfaces are assumed to be adiabatic, i.e. assuming the effect of convection and radiation cooling mechanisms are negligible. The relevant material thermal properties used in this analysis are listed in Table I. The gold thermal conductivity is 214 W m⁻¹K⁻¹ based on Wiedemann–Franz law, which relates the electrical and thermal conductivities, and a measured value of 3 × 10⁷ S m⁻¹ for the electrical conductivity of the film. Upon setting up the proper simulation boundaries and material thermal properties, the steady state and transient solutions are obtained by solving the heat equation.

It is difficult to predict the thermal conductivity of the diode material stack. It contains many sub-micrometer layers in thickness, which means that their thermal conductivities will be less than the bulk values. It is also difficult to estimate the effect of doping and interfaces (such as Ti/GaAs and InGaAs/GaAs) on the overall thermal resistance. For example, the Ti/GaAs thermal interface can add a significant thermal boundary resistance, which depends on the fabrication conditions [23]. Therefore, the layers are lumped into a single
element with thickness 1.37 μm. Its thermal conductivity will be used as a fitting parameter in the results section. The other fitting parameter is the plated gold mass density, which should be less than the bulk property due to the plated gold quality.

### B. Circuit Electrical Analog

The unidirectional flow of heat in solids is analogous to the flow of charge in an electrical network. Therefore, models consisting of thermal resistors \( R_{\text{th}} \) and capacitors \( C_{\text{th}} \) have been developed \[24\] to study heat dissipation in semiconductor devices. Each material through which heat flows can be represented as a shunt combination of \( R_{\text{th}} \) and \( C_{\text{th}} \) given by \[25\]:

\[
R_{\text{th}} = \frac{t}{A\kappa} \\
C_{\text{th}} = \rho Vc_p
\]

In (1) and (2), \( t \) is the material’s thickness, \( A \) is the cross-sectional area, \( \kappa \) is the thermal conductivity, \( \rho \) is the density, \( V \) is the volume, and \( c_p \) is the specific heat capacity. For a linear system, the steady state junction temperature of the diode is given by:

\[
T_j = T_{\text{amb}} + P_T R_{\text{th,tot}}
\]

Where \( T_j \) is the diode junction temperature, \( T_{\text{amb}} \) is the ambient temperature, \( P_T \) is the total power dissipated within the diode, and \( R_{\text{th,tot}} \) is the total diode thermal resistance. For the remainder of this paper, the “th” subscript indicating thermal will be dropped, since all the impedances are assumed to be related to thermal flow.

The transient heating or cooling behavior of the diode can be represented by \( N \) exponentially rising (heating) or decaying (cooling) terms, each having a time constant equal to: \( \tau_n = R_n C_n \), where \( n \) is an integer from 1 to \( N \). When heat flows through a device, one can show \[19\] that the transient junction temperature of the diode after the heating voltage pulse is turned off can be written as:

\[
T_j(t) = T_{\text{amb}} + T_1 * exp\left(-\frac{t}{\tau_1}\right) + ... + T_N * exp\left(-\frac{t}{\tau_N}\right)
\]

where \( T_n (n > 0) \) is the temperature raise associated with an internal thermal resistance inside the device.

---

**Table I: Material Thermal Properties**

<table>
<thead>
<tr>
<th>Material</th>
<th>Thermal conductivity (W/m K)</th>
<th>Thermal capacity (J/Kg K)</th>
<th>Mass density (Kg/m³)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diode Stack</td>
<td>&lt;44</td>
<td>350</td>
<td>5320</td>
</tr>
<tr>
<td>Plated gold</td>
<td>214</td>
<td>129</td>
<td>&lt;19300</td>
</tr>
<tr>
<td>SU-8</td>
<td>0.25</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Silicon</td>
<td>150*(300/T)¹/³</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

From this, an RC thermal model (Fig. 3) is constructed for the fabricated quasi-vertical diodes under test (based on the geometry shown in Fig. 2). The ground symbol in Fig. 3 symbolizes the silicon heat sink. The circuit model can be simplified based on the geometry of the device, and the material properties in Table 1. The SU-8 resistence is an open circuit, since its thermal conductivity is very small (thermal insulator). The two-dimensional gold spreading and gold overlay have a large thermal conductivity. Therefore, they can be seen as a thermal short. Finally, the diode stack is replaced with a single RC combination, accounting for size and boundary effects.

**Fig. 3.** (a) SEM of the quasi-vertical diode profile showing the device mesa, underlying ohmic contact, and anode contact airbridge (b) Close-up SEM view of the anode contact for a 5.5 micrometers diameter diode.

### IV. Thermo-Reflectance Measurement Technique and Setup

Thermal characterization consists of studying the temperature response of a circuit component due its internal self-heating. The key figures of merit include the junction temperature, internal heat resistances, and thermal time-constants.
[19]. To characterize the thermal behavior of a semiconductor device, one must measure a temperature dependent physical phenomena associated with it. The thermo-reflectance measurement technique relies on the change in refractive index, and therefore surface reflectivity ($\Delta R / R$), with changes in temperature ($\Delta T$). The relation between these two quantities can be approximated to first order as:

$$\frac{\Delta R}{R} = \left( \frac{1}{R} \frac{\partial R}{\partial T} \right) \Delta T = \chi \Delta T$$

(5)

where $\chi$, which is typically of the order of $10^{-2} - 10^{-5} \text{ K}^{-1}$, is the thermoreflectance calibration coefficient that depends on the sample material, the wavelength of the illuminating light, the angle of incidence (and thus, by extension, the surface roughness), and the composition of the sample in the case of multi-layer structures. The values of this calibration coefficient reported in the literature for plated gold at 530 nm [26][27][28][29][30] vary from $-3 \times 10^{-4}$ to $-2.3 \times 10^{-4} \text{ K}^{-1}$. In this work, we choose the value of $-2.65 \times 10^{-4} \text{ K}^{-1}$. The error resulting from this choice will be evaluated later in the paper.

The setup shown in Fig. 4 is used to measure the change in diode surface reflectivity under a voltage-pulse device excitation. It consists of a Microsanj NT210A unit which encompasses a signal generator, a pulse generator, and a control box. The setup also includes a green (530 nm) LED light source and a CCD camera to sense the intensity of the reflected signal. 530 nm light is chosen because the reflectivity of a gold surface is sensitive at this wavelength and smaller reflected signal. The setup also includes a green (530 nm) LED encompassing a signal generator, a pulse generator, and a control box. The setup also includes a green (530 nm) LED light source and a CCD camera to sense the intensity of the reflected signal. The setup also includes a green (530 nm) LED light source and a CCD camera to sense the intensity of the reflected signal. The setup also includes a green (530 nm) LED light source and a CCD camera to sense the intensity of the reflected signal. The setup also includes a green (530 nm) LED light source and a CCD camera to sense the intensity of the reflected signal.

Because the change in surface reflectivity has a small magnitude, the signal measured by the CCD camera is averaged for each CCD exposure, and over many continuous device thermal excitation cycles at a selected frequency. The CCD records reflectance information only for the duration of the LED illumination pulse. Consequently, the measurement requires synchronizing the device excitation, the illumination, and the image acquisition. Fig. 5 shows the different timing signals involved. A square-wave voltage pulse train excites the diode. The ON voltage level is set based on the desired power to be delivered to the diode. The diode voltage and current are monitored using an oscilloscope connected in shunt with the device, and an ammeter in series with the setup. For each period, the voltage is ON for 50 $\mu$s with a 30% duty cycle. These parameters allow the diode to cool down completely to ambient temperature before the next excitation is applied. From these parameters, the Microsanj software calculates the minimum width of the LED pulse, and therefore the minimum time resolution of the measurement, to be 500 ns. The delay of the illumination pulse relative to the device excitation pulse is controllable. The full diode thermal transient can then be assembled by acquiring a sequence of images as the illumination pulse delay is adjusted in increments relative to the rising edge of the diode excitation pulse. In this work, for each increment in pulse delay time, the reflected light intensity is measured by the CCD camera is averaged for four minutes. System timing and CCD acquisition were controlled by custom Microsanj hardware triggers and SanjView program [31].

V. THERMO-REFLECTANCE MEASUREMENT RESULTS

The thermo-reflectance characterization is performed for three devices: two 5.5 $\mu$m (D1 and D2) and one 3.5 $\mu$m (D3) diameter diode. The temperature maps of the devices are obtained using the procedure described in the previous section. As mentioned earlier, the heating voltage is switched to zero at time = 50 $\mu$s. Figs. 6 and 7 show the two-dimensional (2-D) temperature maps of D1 and D3 at four different times during cooling. The input powers to these diodes are respectively 6.86 mW and 6.88 mW. We use a post-processing shadown algorithm to remove data from curved surfaces, which result in erroneous intensity readings. Consequently, these data points appear as black pixels in the 2-D image. The surface of the gallium arsenide mesa is also shaddowed, since this material is not sensitive to light at 530 nm wavelength. All the 2-D images are calibrated for gold using a coefficient of thermo-reflectance of $-2.65 \times 10^{-4} \text{ K}^{-1}$.

The heating and cooling curves are obtained from the 2-D temperatures maps. For the 5.5 $\mu$m diodes, the temperature
is averaged over a rectangular region of interest (ROI) shown in Fig. 6-d. This region contains 196 measured pixels, and is located at the top of the anode pillar. It is the closest visible region to the anode-mesa contact, which is the hottest spot of the device. Similarly, for the 3.5 $\mu m$ diode, the temperature corresponds to the average temperature for the ROI region shown in Fig. 7-d, which contains 16 measured pixels.

The heating and cooling profiles for D1 (5.5 $\mu m$ diode) at different power levels is presented in Fig. 8. The temperatures are temperature rises relative to ambient (293 K), since thermo-reflectance imaging measures changes in surface reflectivity. In this figure as well all the subsequent ones, symbols indicate measured values while lines represent fitted or modelled values. In Fig. 8, the straight lines are fitted from the cooling curves based on eq. 4. For clarity, we focus on the cooling curves at three power levels and their corresponding fitting curves (Fig. 9).

Fig. 6. Transient cooling 2-D temperature maps for diode 1 (5.5 $\mu m$ diameter) calibrated for gold surface at (a) $t = 49.7 \mu s$ (b) $t = 51.9 \mu s$ (c) $t = 54.6 \mu s$ (d) $t = 57.3 \mu s$.

Fig. 7. Transient cooling 2-D temperature maps for diode 4 (3.5 $\mu m$ diameter) calibrated for gold surface at (a) $t = -49.8 \mu s$ (b) $t = 51.7 \mu s$ (c) $t = 54.2 \mu s$ (d) $t = 56.7 \mu s$.

Fig. 8. Heating and cooling temperature profiles for D1 device (5.5 $\mu m$ in diameter) at different power levels. The fitted curves extracted from the measurement are shown as straight lines.

Fig. 9. Cooling temperature profiles for D1 device (5.5 $\mu m$ in diameter) at different power levels. The fitted curves extracted from the measurement are shown as straight lines.
The thermal parameters extracted from the fitting curves shown in Fig. 9 are summarized in Table II. As expected, the peak temperature rise scales linearly with power according to eq. (3). The fitted value for the temperature rise agrees well with the measurement. Furthermore, this quasi-vertical device exhibits a single time constant of $\sim 2 \mu s$. This is unlike other diodes reported in the literature [19], which have slower thermal processes that result from their geometry and fabrication. The total measured thermal resistance is calculated using eq. (3). It has a value that is comparable to other diodes reported [19]. A small dependence on the input power is observed. This could be due to the diode series resistance, which becomes more significant as the diode bias increases. In other words, at high power levels, some of the heat that was generated at the diode junction is now dissipated across the series resistance. The error in the peak temperature rise is comprised of two components: the error due to the choice in coefficient of thermal reflectance ($\varepsilon_{\chi}$) and measurement error ($\varepsilon_m$). $\varepsilon_{\chi}$ is calculated by obtaining the temperatures corresponding to the ROI region given the minimum and maximum values for $\chi$ found in the literature ($-3 \times 10^{-4}$ to $-2.3 \times 10^{-4}$ K$^{-1}$). $\varepsilon_m$ is the standard error of the mean from seven thermo-reflectance measurements of D1 based on a 95% confidence level. $\varepsilon_m$ has a value of ±0.54. Finally, the fitting error in $T_1$ and $\tau_1$ correspond to a 95% confidence bound with an R-square value of 0.999.

To fit to this measurement, the values of the missing material parameters are adjusted in the circuit and finite-element models. The circuit model shown in Fig. 3 is used to obtain initial values for the diode stack and plated gold thermal conductivities, and the mass density of plated gold. These values are then used as inputs in the Ansys model. They are tuned further until a match between the measured and modelled curves is achieved. We extracted a value of 4.4 ($\pm 1$) W mK$^{-1}$ for the diode stack thermal conductivity, and 12981 Kg m$^{-3}$ for the gold mass density. Fig. 10 compares the measured cooling curves for D1 with the ANSYS model. Additionally, Fig. 11 compares the measured cooling curves for D3 (3.5 $\mu$m in diameter) with the Ansys model.

Table III summary

Rationale for diode stack

VI. ELECTRICAL VERIFICATION

Finally, a brief discussion of the diode’s internal thermal dependence is necessary. In this work, the diode junction temperature is measured mainly using a thermo-reflectance technique. However, a verification is also performed using the I-V pulsed method [19], which relies on the temperature dependence of the current-voltage (I-V) relationship of a Schottky diode, which can be expressed as [32]:

$$I(V) = I_{sat} \exp\left(\frac{q(V - IR_s)}{\eta kT_j}\right)$$  \hspace{1cm} (6)

where the saturation current $I_{sat}$ is:

$$I_{sat} = S A^* T_j^2 \exp\left[\frac{-q\Phi_B}{\eta kT_j}\right]$$  \hspace{1cm} (7)

In (6) and (7), $\eta$ is the ideality factor, $R_s$ is the series resistance of the diode, $q$ is the elementary charge, $V$ is the
applied voltage, $k$ is Boltzmann’s constant, $A^{**}$ is the modified Richardson constant, and $\Phi_B$ is the barrier height.

The procedure for the electrical verification is very similar to the one described in

$$k = \frac{\text{current density}}{\text{electric field}}$$

**Fig. 12. Calibration curves at different bias points.**

**Fig. 13. Electrical vs thermo-reflectance cooling measurement comparison.**

### VII. CONCLUSION

REFERENCES


[21] ANSYS Mechanical, ANSYS Inc, Southpointe, 275 Technology Drive, Canonsburg, PA 15317, USA.


