Strain Compensated InGaAs/AlAs Triple Barrier Resonant Tunneling Structures for THz Applications

Craig P. Allford and Philip D. Buckle

Abstract—We report a theoretical study of InGaAs/AlAs triple barrier resonant tunneling heterostructures, which are optimized for operation in the terahertz frequency range, and compare these to current state-of-the-art double barrier structures reported in the literature. We consider the effect of strain introduced due to the large lattice mismatch between the substrate, quantum well, and potential barrier materials and describe designs with strain compensated active regions. Constraints have been imposed on the designs to minimize charge accumulation in the emitter quantum well, which is often associated with more complex triple barrier structures. The use of a triple barrier structure suppresses the off-resonance leakage current, thus increasing the maximum output power density, with \( \approx 3 \) mW/\( \mu \)m\(^2\) predicted at 1 THz. The use of thinner potential barriers also reduces the carrier transit time through the structure, which increases the maximum output frequency, predicted to be \( \geq 4 \) THz for optimized structures.

Index Terms—Resonant tunneling diodes (RTDs).

I. INTRODUCTION

The frequency range from 300 GHz to 10 THz, typically known as the terahertz region of the electromagnetic spectrum, is of great interest due to its potential applications. Enhanced security imaging [1], which exploits the unique “terahertz fingerprint” of many nonconductive materials to identify hidden objects, ultrafast wireless communications for short range high-capacity line of sight communication [2], and noninvasive highly sensitive medical imaging due to the nonionizing nature of the terahertz radiation [3] are a few of the potential applications offered by radiation in this frequency band.

Despite the development of several optical and electrical devices, which operate in the THz frequency range, the applications and commercial opportunities are still limited. Optical terahertz sources that have been developed are challenged by difficulties in obtaining suitably low energy band-to-band transitions and the need for cryogenic cooling to operate at THz frequencies [4], and as such are incompatible with compact modern electronic circuitry. Solid state sources, however, despite being compact, are limited in operating frequency by the carrier transit time, which is often too long for devices to operate in the THz frequency band [5]. Therefore, the lack of practical and coherent THz radiation sources has led to the term “terahertz gap” [6] being used to describe this frequency range.

Resonant tunneling diodes (RTDs) exploit the phenomenon of quantum mechanical tunneling, which is inherently a fast process, and unlike conventional electronic devices the operational speed of RTDs is mainly governed by the carrier tunneling time through the structure rather than a conventional transit time. Thus, RTDs are widely recognized as the fastest solid-state electronic devices.

Conventional double barrier RTDs, which are well studied, are used in current state-of-the-art devices [7]; however the most recent improvements in the measured emission frequency of these devices have been as a result of improved device fabrication and design, rather than improved structure design. Triple barrier RTDs are less studied and exhibit more complex resonant tunneling mechanisms [8]. These have been considered for their potential to operate at higher frequencies and with more output power than conventional double barrier structures [9]–[12].

The addition of a third potential barrier is essential to reduce the background leakage current, which dominates the off-resonance valley current in double barrier structures, whilst maintaining the high transmission coefficient through the device, which contributes to a large on-resonance current. This in turn improves the difference in the peak to valley current \( \Delta I \), which is an important device parameter for both maximum output power and frequency of oscillation.

To achieve oscillation frequencies in the THz window with sufficient output power to be utilized in real-world applications, an optimized double barrier resonant tunneling structure has been developed in the In\(_{0.53}\)Ga\(_{0.47}\)As/AlAs (indium gallium arsenide/aluminum arsenide) material system grown on lattice matched InP (indium phosphide) and reported by Kanaya et al. [13], and similar structures are currently being utilized in state-of-the-art RTDs, which have recently been reported operating with a frequency of 1.92 THz [14] at room temperature.

Due to the increased degree of complexity of triple barrier RTDs a thorough understanding of the behavior of these structures is needed, along with careful consideration and optimization of the triple barrier design. A recent study optimizing a triple barrier structure in the GaAs/AlGaAs material system has been...
reported [15]; however high-frequency operation of devices in this material system are less practical for general exploitation due to the required low temperatures.

In this paper, we report a theoretical optimization study of triple barrier RTDs, which build on the current state-of-the-art double barrier designs, whilst still considering the practical growth requirements for strain compensation of the InGaAs/AlAs material system.

II. REQUIREMENTS FOR THz EMISSION

The negative differential conductance regions exhibited in the current–voltage (I–V) characteristics of resonant tunneling devices, combined with an appropriate resonant circuit allow for emission of high-frequency radiation, which can extend into the THz frequency regime. The theoretical maximum oscillation frequency of an RTD ($f_{\text{MAX}}$) has been known for many years and is given by (1) [16] and based on the small-signal equivalent circuit of an RTD given by Brown et al. [17]

$$f_{\text{MAX}} = \frac{1}{2\pi} \left( \frac{1}{2L_{QW}^2 C_D} \right)^{1/2} \left\{ 2L_{QW} - \frac{C_D}{G^2} + \left[ \frac{C_D}{G^2} - 2L_{QW} \right]^2 - \frac{4L_{QW}^2 (1 + R_S G)}{R_S G} \right\}^{1/2}$$

where $C_D$ is the space charge capacitance resulting from the charging and discharging effect of charge carriers within the device depletion regions, $G$ is the negative differential conductance, and $R_S$ is the device series resistance, which includes contact resistance, spreading resistance, and the resistance of the emitter and collector regions. $L_{QW}$ is known as the quantum inductance and is given by

$$L_{QW} = \frac{\tau_{\text{nd}}}{G}$$

where $\tau_{\text{nd}}$ is the tunneling time through the RTD structure. In suitably designed structures, $L_{QW}$ can be shown to not have a significant impact on the high-frequency operation of RTDs [18] and (1), to a good approximation, reduces to [19]

$$f_{\text{MAX}} = \frac{1}{2\pi C_D} \sqrt{\frac{G}{R_S} - G^2}.$$  \hspace{1cm} (3)

Thus, to maximize the device oscillation frequency, minimizing the passive device components such as the series resistance and parasitic capacitances, whilst maximizing the negative differential conductance is necessary, where the average negative differential conductance of a resonant peak in the device I–V characteristic can be calculated from [20]

$$G = \frac{3}{2} \frac{\Delta I}{\Delta V}$$

where $\Delta I$ is the peak to valley current difference and $\Delta V$ is the peak to valley voltage difference.

Whilst maximizing $G$ is an important consideration in optimization of the maximum frequency of oscillation, for practical devices and for commercial applications of such devices, the maximum output power is important.

The maximum output power for resonant tunneling devices with a static I–V characteristic (in the steady state), which is represented by a cubic polynomial can be calculated [21], where

$$P_{\text{MAX}} = \frac{3}{16} \Delta I \Delta V.$$  \hspace{1cm} (5)

However, for practical applications it is more appropriate to consider the frequency-dependent output power, $P_{\text{MAX}}(f)$ [22]

$$P_{\text{MAX}}(f) = \frac{3}{16} \cos \left[ \omega(f) \tau_T \right] \Delta I \Delta V.$$  \hspace{1cm} (6)

The total intrinsic delay of the charge carriers traveling through the resonant tunneling structure $\tau_T$ serves to decrease the maximum output power with increasing frequency and is given by [23]

$$\tau_T = \tau_{\text{nd}} + \frac{\tau_{\text{dep}}}{2}$$

where $\tau_{\text{nd}}$ is the carrier tunneling time, and $\tau_{\text{dep}}$ the carrier transit time in the depletion region. This finite transit carrier time reduces the negative conductance with increasing frequency and thus for high power output at high frequency, a large negative differential conductance region ($\Delta I$ and $\Delta V$) as well as a short device transit time is required.

With the ever increasing need for low power consumption devices it is also extremely important to consider the power efficiency of the RTDs. Since the output power of the device is related to the difference between the peak and valley voltages $\Delta V$ and not the magnitude of the applied voltage itself, to minimize the wasted power it is desirable for the current resonance peak to occur at a low voltage whilst maintaining a large current.

To analyze and compare the input to output power efficiency of these structures a figure of merit suggested by Baba et al. [24], which is the ratio of the time-averaged electrical chip power $P_{\text{chip}}$ to the steady-state extractable power $P_{\text{MAX}}$ is used.

III. SIMULATION AND STRUCTURE DESIGN DETAILS

The simulations were performed using the WinGreen simulation package [25], which is based on a nonequilibrium Green’s
A function approach to quantum transport in laterally extended layered heterostructures.

Material layer properties such as electron effective mass, dielectric constant, bandgap energy, and valence band offsets are defined in a material database, extracted from [26]–[29], which were modified to also take into account the effects of strain, where appropriate, and simulation temperature, which was maintained at 300 K.

To qualify the simulations and ensure that the theoretical results were comparable to experimental devices, double barrier structures reported by Kanaya et al. [13] and a triple barrier structure reported by Sekiguchi et al. [30] were simulated. The scattering parameter, which describes elastic scattering mechanisms such as phonon, impurity, and interface roughness, is implemented in a single optical potential function as an imaginary self-energy for the Green’s functions. This value was fine tuned for both double and triple barrier structures by comparison of the theoretical and experimental I–V characteristics, which show good agreement, and are shown in Fig. 1. The substructure observed in the negative differential conductance region of the experimental traces in Fig. 1 is associated with the time-averaged measurement of high-frequency oscillations caused by instabilities in the measurement circuit. These features are not present in the simulated data, which considers only the steady-state solutions. The input parameters for the simulated Kanaya et al. double barrier structure with a 12 nm spacer are given in Table I and these parameters form the basis for the simulations presently reported, with only the layer thickness, or layer material varied.

Previous studies of triple barrier resonant tunneling structures have shown that significant charge accumulation can occur in the quantum wells of the structure [31], which is a problem for high-frequency applications due to the increased device capacitance. As a result, careful consideration and design of the triple barrier structure is required to minimize any potential charge accumulation.

To a first approximation, charge accumulation in these structures can be minimized by ensuring that

$$B_E \geq B_M + B_C \quad \text{(8)}$$

where $B_E$, $B_M$, and $B_C$ are the layer widths for the emitter, middle, and collector barriers, respectively. As the electron transmission probability crudely depends on the barrier widths, this imposes limits on the structure such that the transmission probability into the emitter quantum well is less than the transmission probability out of the emitter well, thus minimizing charge accumulation.

However, in reality, at energies equal to carriers in the three-dimensional (3-D) emitter and with bias across the structure

$$T_E > T_M > T_C \quad \text{(9)}$$

where $T_E$, $T_M$, and $T_C$ are the transmission probability for the emitter, middle, and collector barriers, respectively. Thus, provided that (9) is still true at resonance, barrier width combinations such as 3 monolayers (ML), 2 ML, and 2 ML for the emitter, middle, and collector barriers, respectively, will allow for minimized charge accumulation.

### IV. Simulation Results and Discussion

The structure designs for those simulated are given in Table II, where the requirements for minimizing charge accumulation and thin barriers for high current density with fast tunneling times have been imposed. The potential barrier material has been chosen as aluminum arsenide (AlAs) to maximize the height of the confining potential barriers compared to the indium gallium arsenide composition ($\text{In}_x\text{Ga}_{1-x}\text{As}$) of the emitter quantum well ($W_E$). The width of the collector quantum well

<table>
<thead>
<tr>
<th>Layer Widths for the Simulated Triple Barrier RTDs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design</td>
</tr>
<tr>
<td>---------</td>
</tr>
<tr>
<td>In53A</td>
</tr>
<tr>
<td>In53B</td>
</tr>
<tr>
<td>In53C</td>
</tr>
<tr>
<td>In53X</td>
</tr>
</tbody>
</table>

All structures utilize AlAs barriers and an $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ collector quantum well, with an emitter quantum well alloy composition, $\text{In}_{0.90}\text{Ga}_{0.10}\text{As}$ for designs In53A, In53B, and In53C and $\text{In}_{0.80}\text{Ga}_{0.20}\text{As}$ for In53X.
Fig. 2. (a) Zero bias (0.0 V) conduction and valence band potential profiles for the In53B design (black solid lines). The Fermi level at 0.00 eV is also shown (dashed blue line) along with the electron charge density (red dashed–dotted line). (b) The conduction and valence band potential profiles and electron charge density at 0.90 V for the In53B design.

Fig. 3. Simulated current density–voltage characteristics for a selection of structures based on the (a) In53B and (b) In53C designs (Table II) with varying collector quantum well width. The characteristics for different monolayers collector well widths are shown in comparison to the double barrier structure by Kanaya et al. [13].

has been chosen to result in an active region, which is as close to being strain compensated as possible, where the strain introduced by the AlAs (tensile) and In$_x$Ga$_{1-x}$As (compressive for $x > 0.53$) layers has been approximately balanced when compared to the In$_{0.53}$Ga$_{0.47}$As collector and emitter regions of the device, which are lattice matched to InP. The width of the collector quantum well ($W_C$) is optimized whilst the well alloy composition remains fixed at In$_{0.47}$Ga$_{0.53}$As.

The conduction and valence band potential profiles and electron charge density for the In53B design active region are shown in Fig. 2 at 0.0 V [see Fig. 2(a)] and at 0.90 V [see Fig. 2(b)], where the largest current resonance in this structure is observed.

The electron charge distribution, shown in Fig. 2(b), indicates that there is minimal charge accumulation in the emitter quantum well in this structure, and is comparable to the simulated electron charge distribution for the Kanaya et al. double barrier structure (not shown).

Examples of the forward bias simulated current density–voltage characteristics are shown in Fig. 3(a) and (b). Due to the thin middle barrier in the triple barrier heterostructure design there is a large energy splitting between the quasi-confined quantum well 2-D electron states, and as such the transmission probability of electrons at these energies is very high. This results in an $I$–$V$ characteristic in which two resonant peaks are present, as can be seen in Fig. 3(a) and (b).

The magnitude of the first and second observable resonant current peaks varies with collector quantum well width, as well as between different structure designs (In53A, In53B, In53C, and In53X). For high-frequency applications the largest resonant feature of the current density–voltage characteristics for the structure designs previously described were analyzed to extract important measures such as an average negative differential conductance ($G$), steady-state output power density ($P_{\text{MAX}}$), output to input power ratio, and resonance peak voltage, where plots of these, for the variations of the structures described, are shown in Fig. 4(a) and (b) and Fig. 5(a) and (b).

With a low output to input power density ratio, increased negative differential conductance, and output power density, the
In53A structures perform poorly in comparison to the optimized double barrier structure by Kanaya \textit{et al.}, as well as the In53B, In53C, and In53X triple barrier structure designs. This poor performance can be attributed to the thicker, 4 ML emitter barrier width and further emphasizes the need for thin barriers in these structures.

The In53B set of designs, however, are much more promising, with nearly double the negative differential conductance and output power density than that of the optimized double barrier structure. The resonances in these structures occur at higher voltages and, as such, despite In53B 11 ML and In53B 12 ML designs having the highest output power to input power density ratio, these values are still of lower efficiency than the optimized double barrier structure.

The series of In53C designs are very interesting structures, with resonances that occur at high voltages, but with extremely large $\Delta I$ and $\Delta V$ thus resulting in large output power density and negative differential conductance. These designs also appear to be relatively efficient with output to input power density ratios between $\approx 4\%$ and $6\%$, due to the large peak resonance current in comparison to the off-resonant current. However, these structures pose a higher degree of uncertainty in the calculated values due to the high voltages at which the resonances occur. At such high voltages there are many more leakage mechanisms, which can contribute to the off-resonant current, which these simulations do not consider. As such the series of In53C designs must be treated with some caution.

The final set of designs, In53X that utilize an In$_{0.80}$Ga$_{0.20}$As emitter quantum well rather than In$_{0.90}$Ga$_{0.10}$As, offer the lowest resonance peak voltages of all of the triple barrier structure designs. However, despite good efficiency with the wider collector quantum wells (15 and 16 ML) and negative differential conductances similar to the double barrier structure, the output power density for these structures is low and as such these structures overall perform quite poorly.
From this analysis it is trivial to see that optimizing a structure, which improves upon all of the parameters considered here is not possible, as each design has some improvements, but sacrifices other aspects in order to achieve these benefits. Thus, the design of the structures should be tailored to the application, rather than using a simple superior design. For example, the series of In53C designs are likely to have superior output power density and negative conductance but become less efficient as a result.

To compare the frequency-dependent power of these simulated structures against the optimized double barrier structure, the tunneling time through the RTD ($\tau_{\text{rtd}}$) has been calculated from the full width half-maximum of the transmission coefficient for the In53B 11 ML, In53B 12 ML, and In53C 9 ML structures and found to be 16 fs, 13 fs, and 13 fs, respectively. The transit time through the collector depletion region ($\tau_{\text{dep}}$) is assumed to be equal to that calculated by Kanaya et al. [13], which is 60 fs. Therefore, from (7), the total delays of the charge carriers traveling through the structure ($\tau_T$) are found to be 46 fs, 43 fs, and 43 fs for designs In53B 11 ML, In53B 12 ML, and In53C 9 ML, respectively. The intrinsic response frequency limit ($f_c$) due to the tunneling and transit times through the structure can be calculated from [20]

$$f_c = \frac{1}{2(2\tau_{\text{rtd}} + \tau_{\text{dep}})}$$  \hspace{1cm} (10)

and are found to be $\approx$5.43 THz, $\approx$5.81 THz, and $\approx$5.81 THz for designs In53B 11 ML, In53B 12 ML, and In53C 9 ML, respectively. The calculated frequency-dependent power density for these designs, from (6), along with the optimized double barrier structure by Kanaya et al. [13] are shown in Fig. 6.

There is a clear improvement in the output power density in the THz region with the maximum frequency of oscillation also increased due to the reduction in the total transit time through the structure. This is a very important result as these increases significantly improve the output power density at frequencies between approximately 2 and 5 THz, and as such increase the practicality for applications of room temperature RTD devices in this frequency range.

To assess the potential suitability of real-world structures realizable with current basic fabrication technologies, the maximum oscillation frequencies of typical $1 \mu m^2$ mesa devices are calculated using (3). Literature values for $R_S$ and $C_{\text{p}}$ have been used. $R_S$ that includes contact resistance, spreading resistance, and bulk resistance is assumed to be 0.247 $\Omega \mu \text{m}^{-2}$ from [32] and $C_D(=C_{\text{dep}} + C_{\text{rtd}} + C_{\text{c}})$ has been calculated. The capacitance associated with the device depletion region ($C_{\text{dep}}$) is assumed to be 6 fF$\mu \text{m}^{-2}$ for a 12 nm spacer layer (taken from [13]), with the contact capacitance ($C_{\text{c}}$) assumed to be 14 fF$\mu \text{m}^{-2}$ from similarly reported structures [32]. Additional capacitance due to the tunneling and transit delay of carriers for the small-signal case and in the low frequency regime ($\tau_{\text{dep}} \omega < 1$) can be given by [20]

$$C_{\text{rtd}} \approx \left(\tau_{\text{rtd}} + \frac{\tau_{\text{dep}}}{2}\right) G$$  \hspace{1cm} (11)

which, for these structures, can be assumed to be $\approx 5$ fF$\mu \text{m}^{-2}$. Estimates of the maximum frequencies for a $1 \mu m^2$ device for designs In53B 11 ML, In53B 12 ML, and In53C 9 ML, with negative differential conductance values $G$ of 72.9 mS$\mu \text{m}^{-2}$, 72.8 mS$\mu \text{m}^{-2}$, and 104.4 mS$\mu \text{m}^{-2}$ are found to be $\approx$3.4 THz, $\approx$3.4 THz, and $\approx$4.0 THz, respectively.

However, at these predicted high frequencies, the low frequency regime assumption made in (11) is not valid, and thus the full frequency dependence of both $C_{\text{rtd}}$ and the negative differential conductance must be considered [20]

$$C_{\text{rtd}}(\omega) = \frac{2\sin \left(\omega \left(\tau_{\text{rtd}} + \frac{\tau_{\text{dep}}}{2}\right)\right) \sin \left(\frac{\omega \tau_{\text{dep}}}{2}\right) G(\omega)}{\omega^2 \tau_{\text{dep}}}$$  \hspace{1cm} (12)

$$G(\omega) = \frac{2 \cos \left(\omega \left(\tau_{\text{rtd}} + \frac{\tau_{\text{dep}}}{2}\right)\right) \sin \left(\frac{\omega \tau_{\text{dep}}}{2}\right)}{\omega \tau_{\text{dep}}}$$  \hspace{1cm} (13)

Solving (3), (12), and (13) iteratively to a converged solution results in revised $f_{\text{MAX}}$ values of $\approx 3.1$ THz, $\approx 3.1$ THz, and $\approx 3.4$ THz for $1 \mu m^2$ mesa devices of structure designs In53B 11 ML, In53B 12 ML, and In53C 9 ML, respectively.

By further decreasing the mesa size, the reduction in device capacitance results in an increased $f_{\text{MAX}}$, but a decrease in the absolute output power of the device. Thus, the tradeoff between frequency and output power still applies here, and the device design must be tailored to the application. However, these structures show considerable promise for practical applications in the THz frequency range.

It is also likely that the triple barrier structures can be further optimized to increase device efficiency by altering the 3-D emitter region alloy composition such that the resonances occur at lower voltages. Further performance optimization can also be carried out by altering the barrier and well compositions, as well as a reoptimization of the collector spacer considering the tradeoff between the device capacitance and collector depletion region transit time, similar to that carried out by Kanaya et al. [13].
In reality, however, the large-scale, low-cost, high-volume manufacturing of these structures still remains extremely challenging due to the complex engineering of the associated circuit required for such RTD devices and the precise epitaxial growth of such thin layers. The variability of the tunneling current for proposed electronic devices has been previously explored [33] and is still applicable today. Although research devices with a low yield are likely possible, a single monolayer variation in the emitter barrier can result in a significant change in the I–V characteristic and so result in a dramatically different device response. Therefore, any attempt to scale such devices to low cost manufacturable yields will still require significant work on the control of epitaxy and fabrication processes.

V. CONCLUSION

We have designed, simulated, and optimized a series of strain compensated InGaAs/AlAs triple barrier resonant tunneling heterostructures. The use of a triple barrier structure active region improves both output power density ≃ 3 mWμm⁻² at 1 THz and maximum output frequency ≥ 4 THz in comparison to state-of-the-art double barrier structures. However, growth and fabrication of devices based on these structures still remains extremely challenging due to the complexity of the circuitry required and precise epitaxial growth of such thin layers.

REFERENCES

Craig P. Allford received the B.Sc. degree in theoretical and computational physics and Ph.D. degree from Cardiff University, Cardiff, U.K., in 2009 and 2016, respectively. His thesis was entitled “Resonant tunnelling in GaAs/AlGaAs triple barrier heterostructures.”

In 2016, he was a Postdoctoral Research Associate with Cardiff University, where his research interests included resonant tunneling structures for THz frequency applications and the fabrication and characterization of quantum devices in indium antimonide and closely related materials for novel electron spin and ballistic transport based devices. He is currently a Postdoctoral Research Fellow with the University of Warwick, Coventry, U.K. He also holds an Honorary Postdoctoral Research Associate position with Cardiff University.

Dr. Allford is a member of the Institute of Physics, U.K.

Philip D. Buckle received the B.Sc. degree in physics and Ph.D. degree from Sheffield University, Sheffield, U.K., in 1991 and 1994, respectively. His thesis concerned resonant tunneling structures.

He was a Postdoctoral Researcher at the then University of Manchester Institute of Science and Technology. In 1998, he then joined QinetiQ Malvern (formerly DERA), Malvern, U.K. to lead the Narrow Bandgap Quantum Device Technologies Group, with extensive experience in semiconductor nanofabrication and novel device physics. He is currently a Reader of condensed matter and photonics with the School of Physics and Astronomy, Cardiff University, Cardiff, U.K. He has co-authored more than 70 peer-reviewed journal papers and was one of the founders of the Institute for Compound Semiconductors, Cardiff University, a major innovation investment that will bring European leading large-scale translational semiconductor fabrication facilities to the U.K.

Dr. Buckle was a QinetiQ Fellow. He is a member of the Institute of Physics, U.K.