
This is the author’s final accepted version.

There may be differences between this version and the published version. You are advised to consult the publisher’s version if you wish to cite from it.

http://eprints.gla.ac.uk/155951/

Deposited on: 23 January 2018

Enlighten – Research publications by members of the University of Glasgow
http://eprints.gla.ac.uk
Epitaxial Designs for Maximizing Efficiency in Resonant Tunneling Diode Based Terahertz Emitters

Răzvan Baba, Benjamin J. Stevens, Toshikazu Mukai, and Richard A. Hogg

Abstract—We discuss the modelling of high current density InGaAs/AlAs/InP resonant tunneling diodes to maximize their efficiency as THz emitters. A figure of merit which contributes to the wall plug efficiency, the intrinsic resonator efficiency, is used for the development of epitaxial designs. With the contribution of key parameters identified, we analyze the limitations of accumulated stress to assess the manufacturability of such designs. Optimal epitaxial designs are revealed, utilizing thin barriers, with a wide and shallow quantum well that satisfies the strained layer epitaxy constraint.

We then assess the advantages to epitaxial perfection and electrical characteristics provided by devices with a narrow InAs sub-well inside a lattice-matched InGaAs alloy. These new structures will assist in the realization of the next-generation submillimeter emitters.

Index Terms—resonant tunneling devices, semiconductor epitaxial layers, terahertz radiation.

I. INTRODUCTION

THE increasing connectivity of the world is driving the need for wireless technologies, with the THz spectrum [1], [2], having attracted significant recent attention due to the possibility of high data rates [3] with advantages in security and reduced network interference [4]. THz devices offer the possibility of large bandwidths per channel, with low susceptibility to atmospheric particulate matter. There is a well-known technology gap to generate waves in the near-millimeter region (0.3 GHz to 3 THz) at room temperature with high efficiency and commodity pricing. AlAs/InGaAs/InP resonant tunneling diodes (RTDs) are prime candidates to offer a mass manufacturable route to large bandwidths per channel, with a well-known technology gap to generate waves in the low susceptibility to atmospheric particulate matter. There is security and reduced network interference [4]. THz devices near-millimeter region (0.3 GHz to 3 THz) at room temperature with the possibility of high data rates [3] with advantages in security and reduced network interference [4].

RTDs have shown great versatility as a resonator across the spectrum, having been demonstrated with high output powers (420 µW) at a frequency of 548 GHz [9], 1 mW at 300 GHz and 5 mW at 160 GHz [10]. Maximizing RTD DC-to-RF conversion efficiency not only allows the realization of high power emitters, as area scaling is presently limited by catastrophic failure, but also allows higher frequency emitters to be conceived as conduction losses limit high frequency performance [8].

In this paper, we describe the DC-to-RF RTD THz emitter performance, in terms of coupling loss, electrical power loss in the extrinsic circuit, conduction loss of the RTD and the intrinsic resonant efficiency (IRE) specific to the semiconductor structure. For this purpose, we describe the present assessment criteria for RTDs and perform a quantitative study looking at the possible design trade-offs, by comparing with established trends. A Green’s function numerical software [11] is used for our modelling. We then describe a method to fit the simulation to the I-V characteristics measured from a device. An outline of a parametric design study is then made with the goal to assess the effect on the DC and potential small-signal AC power ratios. A quantitative analysis of the RTD epitaxial design for maximizing the emission efficiency for conventional InGaAs QW RTDs is explored to reveal whether there are possible

Fig. 1. A system diagram showing equivalent circuit model using complex impedances. The RTD THz emitter neglects the power source, and is compartmented in the semiconductor resonator device, and its radiating elements. The former includes the RTD epitaxial layers, modelled as a single element.

RTDs have been demonstrated as effective transceivers in short-distance communications up to 12 Gbps, allowing live streaming of uncompressed 4K video, as well as offering superior detection performance to Schottky barrier diodes [3], [6]. Tunable RTDs were achieved by integrating varactor epi-layers with the resonating element [7], opening the possibility to multi-channel, wide-band emitters. In addition to the record high frequency emission of up to 1.92 THz [8], RTDs have shown great versatility as a resonator across the spectrum, having been demonstrated with high output powers (420 µW) at a frequency of 548 GHz [9], 1 mW at 300GHz and 5 mW at 160 GHz [10]. Maximizing RTD DC-to-RF conversion efficiency not only allows the realization of high power emitters, as area scaling is presently limited by catastrophic failure, but also allows higher frequency emitters to be conceived as conduction losses limit high frequency performance [8].

In this paper, we describe the DC-to-RF RTD THz emitter performance, in terms of coupling loss, electrical power loss in the extrinsic circuit, conduction loss of the RTD and the intrinsic resonant efficiency (IRE) specific to the semiconductor structure. For this purpose, we describe the present assessment criteria for RTDs and perform a quantitative study looking at the possible design trade-offs, by comparing with established trends. A Green’s function numerical software [11] is used for our modelling. We then describe a method to fit the simulation to the I-V characteristics measured from a device. An outline of a parametric design study is then made with the goal to assess the effect on the DC and potential small-signal AC power ratios. A quantitative analysis of the RTD epitaxial design for maximizing the emission efficiency for conventional InGaAs QW RTDs is explored to reveal whether there are possible
new routes to maximize the IRE. Finally, we proceed to explore the use of a binary InAs sub-well inside a lattice-matched InGaAs QW, and repeat the procedure to discover optimal designs for maximal conversion efficiencies.

In Fig. 1, we show an equivalent circuit for the RTD THz emitter and note several device parameters that impact conversion efficiency. To achieve this, we note several components: the DC power supply and the electrical leads or traces, seen as extrinsic impedance elements, the fabricated RTD device, consisting of the intrinsic I-V characteristic of the RTD, and its associated waveguide and antenna. The summary diagram represents the various elements as complex impedances, all which need to be matched to satisfy the maximal power transfer conditions. \( Z_e \) is the input impedance of the power supply unit. \( Z_{on1,2} \) represent the impedance due to the traces, module and packaging found in a typical system. \( Z_{em} \) and \( Z_L \) represent the emitter and collector contact impedance due to the metallization. \( Z_{sheet} \) represents the impedance that can appear due to various mesa fabrication configurations, including the contribution of sidewall leakage current. The intrinsic characteristics of the RTD are included as a single element (a non-linear voltage controlled oscillator). \( Z_{col} \) is given by the \( \lambda/4 \) coupler and \( Z_L \) represents the load, or antenna, impedance. In order to produce a tunable system, several elements require controllability [7].

In this work we do not consider the elements external to the RTD resonator element. For the RTD resonator we later describe a figure of merit we refer to as the intrinsic resonator efficiency (IRE). The label intrinsic is carefully chosen to illustrate that device fabrication (with arising issues such as minimizing contact resistance, preventing sidewall leakage) benefit from the extensive canon [12]–[14] that exists from manufacturing devices with similar requirements. Whilst the resonant tunneling process is generally acknowledged as a femtosecond-scale event, the RTD device behaves as a high frequency electrical resonator, analogous to a Fabry-Pérot cavity. In order to exploit this characteristic for free-space THz wave emission, the RTD requires coupling into a waveguide and suitable antenna. The small-signal THz behavior can be predicted from the I-V characteristics using methodology for optimizing the RF topologies as discussed by Asada and Suzuki [15].

Experimentally, attempts to maximize the output power of RTD-based THz emitters through band-gap engineering of the epitaxial structure include: collector delta-doping or grading alloy content of the emitter [15], a sub-well [16], lowering conduction losses [8], in addition to modifications to the antenna element [17]. Such attempts often come with the risk of sacrificing epitaxial growth quality [18], requiring careful consideration of stress accumulation due to lattice misfit [19].

There have been extensive design analysis and simulation of RTD devices for a range of applications [20], [21], however, no such studies have targeted the feasibility and epitaxial design of RTDs specifically as THz emitters. This paper aims to address this gap in the literature, through evaluation of established design trends and ensuring the initial parameter sensitivity analysis is comparable to the established canon. We then evaluate these trends, by computing the IRE for a number of parameter changes, and then impose strain limitations in order to seek epitaxial perfection. Finally, the paper assesses whether there are alternative growth strategies that provide comparable or increased performance with better process tolerance.

![Fig. 2. Sample modelled RTD I-V characteristic shown with electrical chip power (P\textsubscript{chip}, mid-point of negative differential resistance) and extractable power (P\textsubscript{ext}). Inset] Approximation of the active-region conduction band potential under bias in the near-resonance condition, shown with the 1\textsuperscript{st} quasi-bound state \( E_1 \). Highlighting parameters: well width (WW), well depth (WD), and barrier width (BW), used later in this work.

II. INTRINSIC RESONANT EFFICIENCY

Qualitatively, enhancing PVCR (or \( J_{\text{peak}} \)) without a reduction in \( J_{\text{peak}} \) (or PVCR) (or increasing both simultaneously) has been highlighted as advantageous in increasing THz emission power [16]. We previously suggested how improvement of both leads to a marked improvement in THz emission power [22]. Recently, a conversion efficiency figure of 0.33% was reported for RTDs [17], therefore considerable room for improvement exists in all elements highlighted in the discussions around Fig. 1.

Fig. 2 presents a model [23] “N-shape” DC I-V characteristic for the RTD semiconductor layers, neglecting the contribution of external circuit elements. The turning points on this curve, create a peak and a valley feature, between which the device exhibits negative differential resistance (NDR).

The inset shows the conduction band edge potential (at the \( \Gamma \) point) of the RTD active region. The ‘on resonance’ biasing condition is such that the injector energy level \( E_0 \) aligns with the QW quasi-bound \( E_1 \) level. This allows free flow of carriers. We also highlight three compositional parameters which are used in the sensitivity analysis in section IV.

We had previously introduced the IRE as a method to optimize the DC-to-RF conversion efficiency [23]. The IRE is the ratio of two parameters: the time-averaged electrical chip power, \( P_{\text{chip}} \), and the small signal AC power, \( P_{\text{NDR}} \). The chip power is given by

\[
P_{\text{chip}} = I \cdot V \quad (1)
\]

Where the bias point is chosen as the mid-point of the NDR (\( dI^2/dV^2 \rightarrow 0 \)), as measurements have shown this to correspond of maximum THz emission [24], [25]. We note that practical devices, particularly those with a small cross-section, may oscillate in this region, displaying an average of the current under DC measurements [20]. The extractable NDR power, a predictor of the THz power, derived from the 3rd order polynomial approximation of the I-V plot [26], is defined as
\[ P_{NDR} \approx \frac{3}{16} \Delta I \Delta V \quad (2) \]

Where \( \Delta I \) and \( \Delta V \) are defined by the peak and valley currents and voltages, respectively. These values are highlighted schematically in Fig. 2. We therefore define our intrinsic resonator efficiency (IRE) as

\[ IRE = \frac{P_{NDR}}{P_{chip}} \quad (3) \]

Maximizing the IRE not only allows optimum conversion efficiency, but also allows high power emitters to be realized if we consider scaling the RTD area up to the maximum possible current density limited by catastrophic failure of the device. Furthermore, not only does the IRE include the contribution of the valley current, but the bias point takes into account the slope given by the NDR. Other criteria for RTD performance derive from static characteristics in the literature include the linewidth of the 1\textsuperscript{st} quasi-bound state divided by the peak current [27] as well as a ratio of carrier transit times to optimize maximum unity gain bandwidth [28].

In Fig. 3, reproduced with updates after Sugiyama et al. [16] (with permission), Horizontal: measured peak to valley current ratio. An observed trend is marked with a dotted line.

An inverse linear trend over 3 decades of peak current densities (\( J_{peak} \)) is observed, with deviation being attributed to design variations and different layer/interface quality[16]. Previous assessment criteria [16] compare the PVCR against the \( J_{peak} \), with an increase in either PVCR and/or \( J \) (with no deterrent to either) being highlighted as advantageous.

This figure plots high current density MOVPE-grown RTD devices [29], [30] with 1.1 nm AlAs barriers. In contrast, the current THz emission record holder [8] contains a highly-strained QW with 0.9 nm AlAs barriers. Presented here are a variety of designs including MBE, MOVPE-grown devices accompanied by an intermediary QW containing an InAs region (sub-well) bounded by lattice-matched InGaAs to avoid the effects of alloy scattering [18] as much as possible. The accumulated stress in sub-well structures can be engineered to resemble the case of a purely ternary, high indium composition alloy. We note that our devices [24] were tested in reverse bias.

III. METHOD

Our model utilizes a computationally efficient equivalent of a Hartree-corrected Schrödinger solver [11], employing the non-equilibrium Green’s functions (NEGF) method[31], often used for near-ballistic electron transport cases [32]–[34]. A 1-D model is commonly used for such devices as ballistic transport dominates [35]. Good agreement with experimental data has been reported using this method [36]–[38]. The simulation successfully accounts for charge build-up in the QW which may introduce unwanted intrinsic bistability. We previously predicted these effects to have a minimal contribution in these high-J structures [29]. In all instances however, a Poisson solver takes into account charge distribution throughout the structure, and the electron-electron scattering potential is also included. This leads to a better estimation of the quasi-bound energy levels compared to the transmission matrix theory alone.

In order to present unambiguous structural designs, all active area dimensions are expressed in monolayers (ML), with 1 ML being half the (001) InP substrate lattice constant. This does not necessarily correspond to the atomic radii, though it provides a convenient basic unit for the simulation. A 1-D mesh size 1ML = 0.293nm. The MOVPE growth of the InGaAs/AlAs RTD on a semi-insulating InP substrate is described elsewhere[39]. In this reference we also comment about the doping uniformity. The modelled structure consists of i) 20 nm In\(_{0.53}\)Ga\(_{0.47}\)As, 3·10\(^{18}\) ii) 7 ML In\(_{0.53}\)Ga\(_{0.47}\)As injector iii) 4 ML AlAs iv) 15 ML In\(_{0.8}\)Ga\(_{0.2}\)As (QW) v) 4 ML AlAs vi) 20 nm In\(_{0.53}\)Ga\(_{0.47}\)As collector vii) 25 nm In\(_{0.53}\)Ga\(_{0.47}\)As, 3·10\(^{18}\). To account for barrier penetration, the Hartree calculation region includes a subset of layers ii to vi.

We will refer to this as our reference structure, yielding a \( J_{peak} \) of ~7 mA·μm\(^{-2}\) (0.7 MA·cm\(^{-2}\)) in reverse bias, previously used to generate a primary oscillation of 350 GHz when coupled to a slot antenna [24]. An RTD with a similar epitaxial structure was used to achieve an output power of ~400μW at 550GHz [9].

The modelling package computes the density of states from the NEGF. The transmission probability is then calculated for every applied bias step to determine the quasi-bound well levels. A general carrier scattering parameter (\( \Gamma \)) within the NEGF serves to account for wave-packet perturbations (and acts as a broadening function). This \( \Gamma \) is fit to ensure quantitative agreement between simulation and experiment. In practice, carrier scattering can be attributed to material interface roughness and compositional modulation [18], leading to an increase in the magnitude and change in shape of the valley current. With this fitting parameter alone, it is possible to obtain a current density estimation within 5dB for a range of devices.

![Fig. 3. Assessment criteria of fabricated RTDs, comparing epitaxial growth methods, updated after Sugiyama et al. [16] (with permission). Horizontal: measured peak to valley current ratio. An observed trend is marked with a dotted line.](image)

![Fig. 4. Modelled and measured (Ref. [24], with permission) V-I characteristics of the RTD. Fit achieved using mesa resistance determined from consecutive wet etch measurements.](image)
discussed here [8], [16], [24]. In addition to epitaxial growth non-uniformity, the departure from reality may arise from peripheral current leakage, deep trap levels, contact resistance, all which hint towards a large dependence on the fabrication process and material quality on the final experimentally obtained I-V characteristic [24].

Fig. 4 plots the simulated I-V characteristic of the RTD active region. Since the simulator output is unidimensional, raw values are expressed in terms of current density, which we extrapolate to our mesa size (3.3 μm²). A fit is produced that takes into account the ∆V, added by the extrinsic circuit, which is compared with the fabricated device by Jacobs et al. [24]. A good fit to the NDR could be achieved only by modifying Γ to a suitable value. The accuracy of the E₁ level prediction, and thus Vₚₑak, was verified using low temperature photoluminescence spectroscopy [40]. In order to transform the simulated curve to include the extrinsic circuit, a series (emitter & collector contacts) and parallel (contribution of sheet resistance, side leakage) resistor are added, resulting in an adequate fit. These values were deduced from successive wet etch and electrical measurement steps [24].

Caution needs to be exercised when employing the IRE for devices that have not been stabilized for DC operation; the valley current is obfuscated by small signal oscillation, with the appearance of a plateau region. Partial stabilization methods were proposed by E.R. Brown et al. [41]. A source of criticism of modelling devices with Green’s functions using the Born self-consistency approximation lies with the insufficient accuracy in the calculation of current density as a result of iterative solver approach, which may lead to physically insignificant results in certain cases [42]. However, for the purpose of this analysis, we have successfully modelled several devices within acceptable quantitative error. The 1.92 THz device from [8] resulted in ~26 mA·μm² (2.6 MA·cm²) without further adjustment of fit parameters, whereas the authors quoted 50 mA·μm². The value of Γ between doped/undoped layers dictates the shape of the IV-characteristic [37]; a low value, perhaps counter-intuitively, lowers the current density, but sharpens the NDR drop (triangle or Z-shaped I-V curve), whilst a high value increases the current density and lowers the slope of the NDR (smooth N-shape I-V curve). To further improve the fit quality, the device active region (area where the Fermi-equivalent potential is tilted) needs a careful selection such that the charge stored in the QW corresponds to those estimated by Shimizu et al. [43]. Whilst the modelled voltage-dependent capacitance is a chief parameter required in predicting the gain bandwidth of the device, this remains a subject to be covered in more detail in future work. In addition, the maximum operating frequency is dependent on the relaxation times given by an electron tunneling from the collector and to the emitter [28], and specifically the unintentionally doped spacer layers ii) and vi). The thickness choice is close to the optimum for the reference structure, but in order to reduce the explored parameter space, these are assumed static throughout this study.

Finally, a fine-grain fit may also take into account epitaxial characterization results, to alter the local layer widths and compositional variations accordingly.

One of the main issues in estimating the oscillator output power points back to the reported hysteretic region of measured I-V curves [44], [45]. The nature of an observed hysteresis can be either due to the intrinsic bistability of the device [46][47], i.e. quantum well charge accumulation and depletion due to quasi-bound level misalignment, or extrinsic, due to the external circuit and parasitic components such as contact resistance, inductance and line-loading effects [48]. An extensive modelling study has also been made to ascertain the effect of various design parameters on the hysteresis and plateau region [20] for memory applications.

We have previously reported on the simulation of a device known to exhibit bistability [29], [46]. We confirmed that the variation between the forward/reverse sweeps are attributable to the asymmetric spacer structure (and barriers, should imperfections exist). However, in our high-J structures, conduction loss is dominant over bistability effects.

IV. SENSITIVITY ANALYSIS

We first proceed with a design sensitivity analysis. Our investigation proceeds by varying 3 key parameters of the RTD: the barrier width, QW width, and QW depth. Firstly, we present the data without any regard to lattice misfit strain.

In turn, Fig. 5 plots the IRE for our standard structure varying a single parameter, and its associated transmission probability for: (a, d) the AlAs barrier width, (b, e) the QW width in ML, and (c, f) the QW indium composition [23], [49]. The E₁ quasi-bound energy level (the first peak of the lower graphs) is overlaid to show the relative change with a variation in these structural parameters. The same vertical scales are used for ease of comparison.

Reducing the barrier width (Fig. 5(a) ) from the initial value of 4ML is observed to result in an exponential increase in J_peak. This also comes at the expense of PVCR, attributed to the reduced dwell time [5], in line with previous work [16]. The increase in IRE may also in part be due to a decrease in NDR which results in a reduced Pₙₚₑak with essentially identical P₁ₜₜ. However, we note that a maximum IRE figure is obtained with 2ML barriers. A strong increase in non-resonant tunneling current occurs for 1ML barriers which reduces IRE and renders the NDR region non-observable for well widths below 13ML. We also note the broadening of the resonance linewidth (Fig. 5(d)), accentuated by thinner barriers.

Finally, a fine-grain fit may also take into account epitaxial characterization results, to alter the local layer widths and compositional variations accordingly.

One of the main issues in estimating the oscillator output power points back to the reported hysteretic region of measured I-V curves [44], [45]. The nature of an observed hysteresis can be either due to the intrinsic bistability of the device [46][47], i.e. quantum well charge accumulation and depletion due to quasi-bound level misalignment, or extrinsic, due to the external circuit and parasitic components such as contact resistance, inductance and line-loading effects [48]. An extensive modelling study has also been made to ascertain the effect of various design parameters on the hysteresis and plateau region [20] for memory applications.

We have previously reported on the simulation of a device known to exhibit bistability [29], [46]. We confirmed that the variation between the forward/reverse sweeps are attributable to the asymmetric spacer structure (and barriers, should imperfections exist). However, in our high-J structures, conduction loss is dominant over bistability effects.

IV. SENSITIVITY ANALYSIS

We first proceed with a design sensitivity analysis. Our investigation proceeds by varying 3 key parameters of the RTD: the barrier width, QW width, and QW depth. Firstly, we present the data without any regard to lattice misfit strain.

In turn, Fig. 5 plots the IRE for our standard structure varying a single parameter, and its associated transmission probability for: (a, d) the AlAs barrier width, (b, e) the QW width in ML, and (c, f) the QW indium composition [23], [49]. The E₁ quasi-bound energy level (the first peak of the lower graphs) is overlaid to show the relative change with a variation in these structural parameters. The same vertical scales are used for ease of comparison.

Reducing the barrier width (Fig. 5(a) ) from the initial value of 4ML is observed to result in an exponential increase in J_peak. This also comes at the expense of PVCR, attributed to the reduced dwell time [5], in line with previous work [16]. The increase in IRE may also in part be due to a decrease in NDR which results in a reduced Pₙₚₑak with essentially identical P₁ₜₜ. However, we note that a maximum IRE figure is obtained with 2ML barriers. A strong increase in non-resonant tunneling current occurs for 1ML barriers which reduces IRE and renders the NDR region non-observable for well widths below 13ML. We also note the broadening of the resonance linewidth (Fig. 5(d)), accentuated by thinner barriers.

Finally, a fine-grain fit may also take into account epitaxial characterization results, to alter the local layer widths and compositional variations accordingly.

One of the main issues in estimating the oscillator output power points back to the reported hysteretic region of measured I-V curves [44], [45]. The nature of an observed hysteresis can be either due to the intrinsic bistability of the device [46][47], i.e. quantum well charge accumulation and depletion due to quasi-bound level misalignment, or extrinsic, due to the external circuit and parasitic components such as contact resistance, inductance and line-loading effects [48]. An extensive modelling study has also been made to ascertain the effect of various design parameters on the hysteresis and plateau region [20] for memory applications.

We have previously reported on the simulation of a device known to exhibit bistability [29], [46]. We confirmed that the variation between the forward/reverse sweeps are attributable to the asymmetric spacer structure (and barriers, should imperfections exist). However, in our high-J structures, conduction loss is dominant over bistability effects.

IV. SENSITIVITY ANALYSIS

We first proceed with a design sensitivity analysis. Our investigation proceeds by varying 3 key parameters of the RTD: the barrier width, QW width, and QW depth. Firstly, we present the data without any regard to lattice misfit strain.

In turn, Fig. 5 plots the IRE for our standard structure varying a single parameter, and its associated transmission probability for: (a, d) the AlAs barrier width, (b, e) the QW width in ML, and (c, f) the QW indium composition [23], [49]. The E₁ quasi-bound energy level (the first peak of the lower graphs) is overlaid to show the relative change with a variation in these structural parameters. The same vertical scales are used for ease of comparison.

Reducing the barrier width (Fig. 5(a) ) from the initial value of 4ML is observed to result in an exponential increase in J_peak. This also comes at the expense of PVCR, attributed to the reduced dwell time [5], in line with previous work [16]. The increase in IRE may also in part be due to a decrease in NDR which results in a reduced Pₙₚₑak with essentially identical P₁ₜₜ. However, we note that a maximum IRE figure is obtained with 2ML barriers. A strong increase in non-resonant tunneling current occurs for 1ML barriers which reduces IRE and renders the NDR region non-observable for well widths below 13ML. We also note the broadening of the resonance linewidth (Fig. 5(d)), accentuated by thinner barriers.
Initially, increasing the QW width (WW = 9 .. 13 ML) (Fig. 5(b)) results in a monotonic increase in IRE, attributed to bringing the 1st quasi-bound resonant level closer to the conduction band potential, at the expense of an increased carrier dwell time[5]. The reduction in IRE by further increasing the QW width (>13ML) occurs as the broadened resonance linewidth begins to fall below the Fermi energy level where the optimal bias point would lie, therefore the resonance energy becomes misaligned with respect to the emitter states. Our modified reference structure with 4MLs barrier and InAsGaAs QW has a maximal IRE with a 13ML QW. A cause of concern with wider QW is shown in Fig. 5(e), is brought by the drop in the 2nd resonance level. It is conceivable this may align with indirect bandgap states to result in an increase in phonon-activated non-resonant tunneling.

Increasing the well depth is achieved by altering the indium composition in the InGaAs alloy (Fig. 5(c)). This has a two-fold effect: a higher indium composition decreases the band gap, but the lowered effective mass translates into a decrease in the available density of states within the QW, resulting in a lower $J_{peak}$. It may appear as a trade-off between $J_{peak}$ and PVCR, however, our simulations predict that a higher indium fraction will reduce the linewidth of the resonant state, resulting in a higher Q-factor of the resonance. Additionally, the resulting reduction of the operational bias, a probable source of phase noise in voltage controlled-oscillators, sets the recipe for a desirable emitter. Fig. 5(f) shows an opposing trend of the 2nd resonance compared to the QW shift; at first sight, increasing the mole fraction towards binary InAs appears to have no downsides. To summarize thus far, the ideal RTD would have a combination of 2 ML barrier pairs between an InAs QW, with possible local optima for the WW. However, before we proceed into the full parametric analysis we need to consider misfit strain.

V. ACCUMULATED STRESS CONSIDERATIONS

The coherent growth of strained layers upon a substrate without the formation of dislocations is a key challenge for epitaxy. There are several models available to calculate the so-called critical thickness for a layer of given strain. Matthews & Blakeslee (M&B) [50] developed a model based purely on mechanical equilibrium theory, yielding a lower theoretical limit of the critical thickness. People & Bean (P&B) considered the format of misfit dislocations to be determined solely by energy balance rather than mechanical equilibrium [51], [52]. In general energy balance (P&B) considerations indicate that thicker strained layers can be grown than predicted by mechanical equilibrium (M&B). Experimental studies of InGaAs/InP epitaxy indicate that poor surface morphology occurs when M&B limits are exceeded, and poor PL emission and poor surface morphology occur as the P&B limit is exceeded [53].

We now proceed with the analysis of the accumulated stress due to the growth of partially strain-balanced structures. The growth of strained layers, strain-balanced QWs and our partially strain-balanced structures is represented schematically in Fig. 6 (not to scale). Fig. 6(a) shows the case for the growth of a strained layer upon unstrained layers. In this case stress accumulates in the structure until a force (M&B) or energy (P&B) limit is reached (shown by a dotted line) and misfit dislocations are formed. Fig. 6(b) shows the case for the strain-balanced structures where the average lattice constant of the structure is designed to be equal to that of the substrate in order to overcome limitations in growing multiple QW layers [54]. This situation balances accumulated stress within the structure. Our situation is shown in Fig. 6(c) where partial strain-balancing is achieved, with a In$_{0.85}$Ga$_{0.15}$As QW (+1.72% lattice constant) and binary AlAs barriers (-5.32% lattice constant).

We solve the recurrent equation (5) of M&B using standard material parameters [55], [56], assuming the InP substrate the unstrained lattice constant, and finally plot the critical thickness depending on the mole fraction x. According to this M&B model, 5.04 ML of InAs can be uniformly grown on an InP substrate before strain relaxation through the formation of misfit dislocations become a problem. This area is shown in green in Fig. 7. To obtain the yellow region, we add the critical thickness of the tensile AlAs barrier to this value, as InAs is compressive. We assume the worst-case scenario where only the first barrier matters in the relaxation scheme [53]. The red zone will thus exceed this limitation, where growth may be possible, with the risk of introducing increasing numbers of defects the further the distance from the strain-balanced point. Furthermore, the black zone also exceeds the P&B limit. We are not aware of any high-quality epitaxial growth attempts that breach this limit.

![Fig. 6. Illustrative schematic (not to scale) of (a) strained layer (central) grown within an unstrained matrix. (b) strain-balancing in a MQW stack incorporating alternating strained (thinner) and lattice-matched (thicker) layers, and (c) our structure incorporates partial strain-balancing. Sample semaphore colour-coding to match the stress calculations shown in Fig. 7. (001) growth direction represented by z. Dotted lines delimit different layers.](image_url)

![Fig. 7. Modelled schematic showing which combinations of QW width and indium composition may be successfully realised. (See text)](image_url)
VI. RESULTS FOR THE TERNARY QW DEVICE

Fig. 8 plots the IRE as a function of well width and indium composition for an RTD design (a) 1ML AlAs barriers, (b) 2ML AlAs barriers, (c) 3ML AlAs barriers, and (d) 4ML AlAs barriers. The color coding of Fig. 7 is preserved for structures which satisfy the requirements for lattice misfit dislocation formation. Once more, the variation of barrier thickness is shown to have the main effect upon IRE. From the analysis of this graph, we suggest an optimal device structure (point β, IRE = 0.158) with 2MLs barriers, 17ML QW and a QW alloy of In$_{0.75}$Ga$_{0.25}$As, a perhaps counter-intuitive result based on current literature where barrier thickness tends to be kept constant, and the push is for deeper and narrower QWs [9], [16], [38]. The reduction in IRE seen with certain combinations of well width & depth is attributed to the drop of the resonant state below the injection energy level. Further improvement of this figure may be achieved by reducing the carrier scattering within the structure, either by improving epitaxial processes or designs (discussed in the following section), or by reducing the operating device temperature [57]. An alternative material system with greater conduction band energy offsets may also enhance this figure.

The simulated I-V of the optimized structure is plotted in Fig. 8(e), along with the reference structure[24]. Not only is the PVCR and $J_{\text{peak}}$ higher, but it also exhibits lower biasing requirements compared to the reference design (identified by point α). We note again that this structure is not strain balanced, but designed to be compressively strained to the maximum acceptable limit based on the adjusted projection of the M&B model, in order to avoid the onset of the generation of misfit dislocations. We also note that our findings indicate the challenge now made to the epitaxial process in realizing high performance THz emitters. Not only are highly uniform thin layers required (e.g. 2ML AlAs barriers), but the limits of strain relaxation must be pushed in order to maximize efficiency of these THz light sources.

In Fig. 9 we plot $J_{\text{peak}}$ as a function of PVCR for the 3ML and 2 ML barrier devices from Fig. 8 (c) and (d), respectively. These are the previous metrics established by Sugiyama et al.[16]. The colored lines denote identical mole fractions, whereas the markers denote identical WW. It is noted that the scatter from the line of best-fit in Fig 3 is ~factor 10 in I and ~factor 4 in PVCR, and the axis limits are significantly smaller in Fig. 9 than in Fig. 3. Furthermore, the data in Fig. 3 contains a sub-set of designs, with varying epitaxial processes, of which a subset will be reported due to self-selection by the authors and the selectivity of the peer review process. We therefore consider that the data in Fig. 9 is not in disagreement with the canon.

The dotted line is the trend line as seen in Fig. 3. The difference in the distribution of 4ML and 2 ML devices is worth noting. In our sensitivity analysis we predicted that a higher indium mole fraction would improve the device performance, whereas this statement is not true for deep QW with long WW. The point of inflection seen in the 2 ML case illustrates the combinations of QW widths and depths that result in $E_1$ levels which fall under the $E_F$ when the required bias is applied, resulting in sub-optimal performance. This does not occur for the parameters discussed in the 4ML case. Unfortunately, this visualization (Fig. 9b) does not necessarily highlight the difficulties associated with uniform growth of thin barriers.

For instance, the placement of the starting and optimized devices ($\alpha$ and $\beta$, respectively), suggest that there may be significant room for improvement in both cases, even if both are approaching the limits of pseudomorphic growth. Furthermore, these criteria fail to account for changes in the...
slope of the NDR, an otherwise critical parameter for optimum power transfer of the resonator to the rest of the system. This unique combination of parameters, as well as the uncertainty in $\Gamma$, leads to the difference in our recommendation of having the widest, deepest QW possible within limit of the accumulated stress and desired resonator frequency, given the current parameter set.

In$_{x}$Ga$_{1-x}$As layer in the ternary QW case. One such example is given in our simulation of the device described in Table II of Sugiyama et al. [16]. For the equivalent structure of 5ML/5ML/5ML of ternary/InAs/ternary alloy, the equivalent indium mole fraction if the QW were a constant composition is $x=0.686$.

VII. RESULTS FOR THE SUBWELL DEVICE

The devices with a constant mole fraction of InGaAs QW shown previously are one of the conventional methods of growing RTDs. However, as evidenced by Fig. 3, there are many more examples of growing a deeper InAs sub-well sandwiched between lattice-matched InGaAs QW layers. There are two main reasons for this approach: i) minimize local barrier thickness extension due to the pseudomorphic AlAs lattice [18] ii) results in a reduction of calibration steps in the initial wafer epitaxial process.

Fig. 10 shows the schematic representation of the conduction band potential of the sub-well device. The InAs layer creates an additional Type-I QW, the sub-well. We initially investigate the merits of this structure using the previously established methodology. All parameters are related to the width of the epitaxial layers: $A$ is the distance in ML from the entry barrier, whereas $B$ is the distance from the exit. $W$ is the width of the InAs sub-well. We note that the ratio between $W$ and overall QW width ($A+W+B$) is analogous to changing the mole fraction of the equivalent In$_{x}$Ga$_{1-x}$As layer in the ternary QW case. One such example is given in our simulation of the device described in Table II of Sugiyama et al. [16]. For the equivalent structure of 5ML/5ML/5ML of ternary/InAs/ternary alloy, the equivalent indium mole fraction if the QW were a constant composition is $x=0.686$.

![Fig. 10. Schematic modelled conduction band diagram of the Ternary/InAs/Ternary epitaxial alloy device.](image)

![Fig. 11. Comparison of simulated uncalibrated I-V characteristics against 2 ternary alloy devices with mole fraction $x=0.7$ and 0.8, respectively, and a lattice-matched ternary alloy InGaAs/InAs/InGaAs sub-well device. For comparison purposes, same modelling fit parameters are used as the devices in Fig.8-9.](image)
with the different set of geometric parameters previously mentioned. The modified reference structure has W, A, B are all 4 ML, but varying a single parameter (a) varying the sub-well width (W), whilst A, B are constant (b) keeping the overall well width fixed whilst W increases (c) the relative offset position in ML from the emitters.

Fig. 12. Modelled intrinsic resonator efficiency figure of merit for individual changes to a structure of 4 ML InGaAs/4ML InAs/4ML InGaAs (a) adjusting QW width with a constant sub-well width(b) keeping QW width constant (c) varying the position of the sub-well inside the QW (d), (e), (f) are their corresponding energy transmission results. The overlaid line in (a), (b), (c) shows the shift in the 1\textsuperscript{st} quasi-bound state. The dotted line is a guide for the eye.

Increasing the sub-well width (Fig. 12(a)) from the initial value of 4ML results in moderate gains with a clear optimum. Shorter sub-wells, therefore reducing the relative concentration of indium, increases the 1\textsuperscript{st} quasi-bound energy level with respect to the Fermi level, resulting in higher device bias requirements. With several combinations, the 2\textsuperscript{nd} resonance merges with continuum states above the AlAs $\Gamma$-point of 1.16eV. However, greatly increasing the well width may have a deleterious effect upon the cut-off frequency of the device, due to an increase in carrier transit times.

If the overall well width is to be kept constant (Fig. 12(b)), there is a monotonic improvement with the sub-well width. In line with increasing the indium mole fraction of the ternary device, this results in apparently direct improvements without penalty, as the 2\textsuperscript{nd} quasi-bound state is pushed into a higher level.

Altering the position of the sub-well whilst keeping its width constant (Fig. 12(c)) is another important factor to investigate, as the precise position inside the well cannot be guaranteed. A surprising result was the increase in the IRE as the sub-well is placed closer to the well. We suggest this effect is due to the average of well potential appearing to be flatter under the bias compared to the formation of a triangular well in the ternary device, resulting in less linewidth degeneracy. An analogous case has been suggested with a ternary device, by ramping up the indium mole fraction [58]. Parabolic conduction band potential QWs have also been created in the GaAs/AlGaAs system[59], however, this is not a direct comparison due to the closer band offsets of the material pair, likely favoring tunnelling through indirect band gap states.

In Fig. 13, we expand upon the previous parameter sensitivity analysis and present the result for (a, b) 4 ML barriers and (c, d) 3 ML barriers, with the sub-well placed in the center, and immediately attached to the emitter-side barrier, respectively. The simulation predicts a best-case IRE of ~5% (+30% improvement vs. ternary device) for the 4ML devices. At a +29% improvement, from 11% to 15.1% IRE, the trend is preserved for a 3ML device, rivalling the previously suggested 2 ML barrier which may decrease the yield of highly-similar diodes. We remind the reader that ~3 ML devices have been successfully grown to obtain 1.92THz [8] primary oscillation.

The accumulated stress limitations in these devices may manifest themselves as Stranski-Krastanov (3D) growth modes, particularly with wider sub-wells. In Fig. 7, we predicted that pending suitable flow and temperature conditions, 5 ML of InAs may be successfully grown on lattice matched InGaAs. However, the question that arises is
whether a high-quality interface between the AlAs barrier and InAs sub-well placed towards the emitter side is possible, taking into account the lattice constant mismatch of -6.06%. This provides a new challenge to the epitaxial process.

VIII. CONCLUSION

We have described a figure of merit based on measurements from our own fabricated RTDs that can be used as a design aid to maximize THz radiation output. We highlight the importance of the chip power and explain how the variation of certain structural parameters affects the efficiency of the device. An optimal device is proposed by varying 3 essential geometrical parameters. The requirements for strain balancing indicate that a wider QW with lower In mole fractions, but much narrower AlAs barriers provides a route to higher efficiency THz sources. We have shown the advantage with respect to the accumulated stress and intrinsic resonant efficiency of moving to a sub-well structure, and have shown such structures benefit from careful design of QW width and position. Additional advantages in reduced alloy scattering and barrier perfection may lead to even greater intrinsic resonant efficiency enhancements for sub-well structures.

ACKNOWLEDGMENT

R.B. would like to acknowledge the fruitful discussions with Dr. Yaoseoung Kim of ROHM, Co., Ltd.

REFERENCES

Rázvan Baba received the M.Sc.(Eng) degree in electronic engineering from The University of Sheffield, Sheffield, U.K., in 2014. He is currently in The University of Glasgow, Glasgow, U.K., working towards his Ph.D. degree in resonant tunneling diodes for terahertz applications.

Benjamin J. Stevens received the Ph.D. degree in advanced GaAs-based lasers from The University of Sheffield, U.K., in 2010. He received a summer fellowship from the Japanese Society for the Promotion of Science joining the Asakawa Group, Tsukuba, Japan, where he learned selective area molecular beam epitaxy. In July 2009, he worked in the National Centre for III-V Technologies, The University of Sheffield, where he is involved in metal–organic vapor phase epitaxy growth. Since 2015 he is working at IQE plc. His research interests include lasers and regrowth technologies.

Toshikazu Mukai received the M.E. degree in engineering science from Osaka University, Osaka, Japan, in 2005. In 2005, he joined ROHM Co., Ltd., Kyoto, Japan. He was involved in developing SiC power devices for three years. He has studied the engineering of resonant tunneling diodes as terahertz oscillators and detectors for terahertz applications.

Richard A. Hogg received the Ph.D. degree in physics in 1995 from the University of Sheffield, U.K. He was a Research Fellow in NTT Basic Research Laboratories, Atsugi, Japan. Received a 2-year EU-Japan fellowship as a Visiting Researcher in Professor Arakawa’s Laboratory, University of Tokyo. For 3 years, he was in Toshiba Research Europe’s Cambridge Laboratory. In 2000, he joined Agilent Technologies Fibre-Optic Component Operation, Ipswich, U.K. In 2003, he joined The University of Sheffield, later chaired as Professor of Semiconductor Devices. Since 2015, he is Professor of Photonic Devices and Systems at the University of Glasgow. Research interests include device physics and engineering, fabrication, and applications of laser, amplifier, and superluminescent diode devices.