
# Strong Physical Uncloneable Functions using Arrays of Resonant Tunnelling Diodes

**Benjamin James Astbury**

Department of Physics

Lancaster University

28th September 2017

This thesis is submitted in partial fulfilment of the requirements for the degree of

Masters of Physics

# Preface

This thesis is the result of work which I performed at Lancaster University, between October 2016 and September 2017. Except where otherwise stated the contents of this thesis is the result of my own work and is not the same as any others I have already submitted, or in the process of submitting, for any degree at any university or institution. The word count on this thesis does not exceed the maximum limit of 30,000 words.

B. J. Astbury

Thursday, 28th September 2017

“Motivation is fickle, you have to cultivate discipline in order to get stuff done…

and that’s so much more reliable than the emotional component of motivation”

Michael ‘Burnie’ Burns, 2017

# Acknowledgements

Throughout my years in education, leading up to this point, numerous people have been monumental in the reason I had the opportunity to work in this research group with such talented and inspiring people for which without them, this thesis would not be possible. First thanks go to my supervisor, Robert Young, who without him seeing my potential where others only saw a number, I wouldn't be in this position. Thank you for letting me join your research group and work on new and exciting concepts.

Next, I would like to thank the collaborators at the University of Manchester, School of Electrical and Electronic Engineering, who created the devices used within this work and performed initial measurements on. A further acknowledgement to Jonathan Roberts, who provided a stepping stone. Without them, the concept of this research would not be possible.

Recognition goes out to members of the research group, for which insightful conversations or help, even if I just talked at them so I could explain it to myself, keep me working in a steady direction. So many thanks to Hamzah Shokeir, Thomas McGrath, Christopher Woodhead but most of all Ramón Bernardo Gavito. Who without his guidance, I would still be completely lost.

I would further like to thank my housemates, for putting up with me, and friends for going for a drink or playing games together. To all those, I've played with and searched for the sleeper all I have to say is "Eyes up, Guardian".

I'd like to thank my family, for believing in me every step of the way even when I didn't. I love you and I'll never stop being thankful for everything you've done for me.

# List of Publications

B.J.Astbury, I.E. Bagci, J. Roberts, J. Sexton, M. Missous, U. Roedig, R. Bernardo Gavito, R.J. Young, “Strong Identities Using Resonant Tunnelling Diodes”, Poster-Presentation, 7ERSW, IMDEA Nanociencia (June 2017).

R. Bernardo Gavito, I.E. Bagci, J. Roberts, J. Sexton, B.J. Astbury, H. Shokeir, T. Mcgrath, Y. Noori, C. Woodhead, M. Missous, U. Roedig, R.J. Young, “Extracting random numbers from quantum tunnelling through a single diode”, arXiv:1706.07348 (Accepted with minor revisions, Scientific Reports).

# Abstract

In a modern world, where the malicious attacks of interconnected devices rises stemming from increased adoption of such systems. Security of these systems have repeatedly been bypassed, as such requiring secure validation through truly unique responses to an authentication request which cannot be impersonated. A resonant tunnelling diode has been shown useful by having a single unique and uncloneable response. The electrically driven device outputs a signature unique to the individual device which is uncloneable even by the manufacturer. The purpose of this work is to expand the range of responses of an individual authentication system using resonant tunnelling diodes.

The combination of resonant tunnelling diodes show a response unique to the base devices with multiple points of authentication. By creating an array structure where devices can be combined in different permutations, the set of responses can be increased. Varying the array's design can maximise the set of response to scale exponentially with the number of devices. The possibility of predicting a set of responses is explored through the initial measurement of base array devices. The risk is explored through the ability to deconvolute array responses into single device signatures and creation of subsequent array responses.

A designed and implemented 4x4, 16 device array with 256 responses is shown to have 99% uniqueness for each 4-peak permutation with a ~20% chance that any single peak will give a false negative response when compared with the expected output. The combination of devices is shown to be random in nature with how the device's signature shift when a second device is applied. The resultant system is given as a design for secure alternative to the current widely used authentication systems in small electronic devices. With such a system in place, security of information and devices can be significantly increased.

# Chapter 1 - Introduction into Authentication Based on Atomic Imperfections

## 1.1 Motivation

In a world where almost every electrical smart device, program and database is connected, with the ability to transfer large volumes of data through this network, the security of these systems is of utmost importance. In this modern day, devices like these have made it into our homes with the invention of intelligent personal assistants, such as Alexa, Cortana, Mycroft, etc [2-4], which are widely connected, controlling lights, heating and even financial interactions with only a spoken word. Other electronics include smart doorbells, bulbs, thermostats and more [5-7] all of which can connect to the Internet of Things (IoT) with the potential to be hijacked.

With the growing size of the Internet of Things (IoT), a network of connected devices, the need for secure communications has become more prevalent. These systems can, for example, be processing financial data, private communications and even contain huge amounts of personal information. Secure transmission between these interconnected devices is required such that nefarious parties are unable to pry into private data and identities. Due to the nature of the IoT, it is required that each individual node of the IoT be secured as once a single node is compromised, then the whole system of devices becomes compromised.

Some of the most worrying and devastating attacks on the IoT come in the form of widespread distributed denial-of-service attacks (DDoS) or manipulation of devices. DDoS attacks consist of sending networks huge amounts of data to slow or even halt systems. One of the largest being an attack on service provider Dyn [8]. Attacks to manipulate vehicles have been shown on a Jeep [9], the attack gave complete control over the movement of the vehicle while in transit containing passengers. Smaller devices include webcams, baby monitors and home security systems.

With a rise in interconnected devices, there has also been a proportional rise in reports of hacking, misappropriation and electronic theft. The range and scope of breaches to electronic systems can be wide, ranging from small and insignificant items to large.

To protect from breaches, each node of an IoT needs to be extremely difficult to bypass. A successful attack would be outweighed by the time and risk in doing so. Hereby the first requirement would be that each authentication code needs to be unique compared to its counterparts, such that brute-forcing a bypass by guessing at an authentication code would become difficult.

Further requirements allow for ease of implementation into any device within an IoT system. IoT devices tend to be small with low-power systems, as such solutions must abide by the restriction in being compact, low-cost with low-power requirements. As such solutions should be self-contained which would allow for the system to be as secure as possible, while also allowing it to be versatile in terms of its implementation into IoT devices.

## 1.2 Authentication and Identities in a Modern Society

The main way of bypassing such a code would be to decipher what the authentication code would be. If an authentication code is unique to a device for a user, then only users with authentication can access that device, program or database. The way in which authentication over a virtual space is achieved at a basic level is the sharing of a secret as registration, which becomes the authentication code. The user then provides the secret to the authentication system, which compares against the previously given secret. A match provides access whereas a mismatch does not.

Currently, the way in which this is carried out varies by level of security, data accessed and the maker of the system. Secrets shared for authentication purposes can range from personal information to physical objects which contain some amount of data. For example, most web accounts require a password, of a certain length composed of alphanumeric digits, which provides authentication. On the other end of the spectrum, a physical card and a randomly assigned code have a similar process for financial transactions. For ultra-high security, fingerprint or retina scans can be used as a form of unique personal information. Although the cost of such high-security tends to be too much for consumer grade security purposes. Thus, a unique outputting device, for low cost would be required to fill this gap, a device which, like a fingerprint, is defined as unique and unable to be cloned.

An authentication system with a unique output can be defined by a physically uncloneable function (PUF). While the output of such a system would be unique, it is also defined as being impossible or very difficult to reliably clone the physical system due to the unpredictable nature of the creation of the system providing the unique response.

## 1.3 Physical Uncloneable Functions

PUFs are generally defined as a system, physically unique in its creation, whereby when probed by a range of inputs give unique responses due to the internal structure of the system [10]. Each device has a distinctive output which is not replicated and is obtained from its paired input hence they can be used for identification and authentication. It is comparable to a set of one-way functions in that outputs are easily computable from the input, but the inverse is much harder to obtain. While the PUF is a one-way function, it has a requirement to be uncloneable so that an attacker, with access to the original system, is unable to replicate a copy of the system through any means. For PUFs to be as useful as possible there is a need for them to be low cost and simple to make, yet almost impossible to clone, even if the manufacturing is copied exactly.

Systems which are implemented as PUFs can be characterised by the security it provides and how it performs this. The merit of security in this instance falls to the response of a device when a challenge is applied to the system. The Challenge-Response Pairs (CRPs) from each representation can be a set of authentication tools, where a larger set often denotes a stronger system.

Additionally, PUFs can be characterised by the effect which creates the randomness such as classical interactions and quantum effects. However, some use introduced randomness as opposed to the intrinsic randomness of the devices to produce unique outputs of PUFs.

Introduced randomness is attractive due to its lower dependency on environmental variations and the ability to easily distinguish devices. Optical PUFs, made from transparent materials doped with particles which scatter the light [11], and coating PUFs, using a random scattering of dielectric particles to create random capacitance [11], are examples of introduced randomness.

However, intrinsic randomness is often preferred, as it can be included in a design without modification to the manufacturing process. Intrinsic classical randomness arises from a variety of methods. Static Random Access Memory (SRAM) can be used in a PUF, by reading each transistor cell upon start-up [12]. Due to manufacturing variations, each transistor cell is more likely to tend to a 0 or 1 when powered off. This can be read and used as a unique PUF [13].

A quantum PUF can utilise the quantum regime to produce truly random and unique outputs from quantum effects and atomic differences in the atoms comprising of the PUF. Once a quantum level is reached, minor differences in interfaces or material composition can vastly change the response garnished from a single system. With such a level of intricacy, the ability to reliably reproduce any PUF relying on quantum effects severely decrease. To reproduce a clone of any single device would require atomic-level engineering on a large-scale due to the number of atoms which make up an individual structure.

PUFs can be implemented into a secure system using a variety of algorithms, which ultimately describe how keys and information are passed between sender and recipient. Such algorithms can be via the use of CRPs so that the response from a PUF can be compared against the database response to allow authentication of a system. Emerging, untapped implementation of the Public-PUF (PPUF), uses a public and private encryption to facilitate secure communication and authentication. Communication via public-private keys, such that information is secured via the public key by the sender, un-encrypting of the information can only be done by a private key held by the recipient.

Authentication via a PPUF is facilitated by a public model of the hardware PUF. The hardware version would be able to compute the response from a challenge at a faster rate than that of the public model of the PUF. As such a timed-authentication technique can be employed, whereby the response from a client can be timed in conjunction with the response. The client, who hold the hardware PUF, computes the response at a much faster rate than any public model could, at which point the time of the response and the response is compared to that of the model. Therefore, only allowing the hardware model to be authenticated.

### 1.3.1 Different forms of PUF

Physically uncloneable functions come in a variety of different forms, each with different axioms which they require to determine the type, level of security and the ease of use. The security of a PUF increases normally with the number of unique responses output by an individual system along with its ability to be cloned through various means.

Unique Objects, (UNOs), are a physical display with some random characteristics, which when challenged by external measurements, can define a small, fixed output set which is unique from any other object. The physical system is impossible to recreate even if the exact fabrication and structure of the original UNO are known. However, due to its nature, it is possible to simulate the UNO this is guarded against by a visual verification of the device.

The visual device is often represented by a random function which is externally measured and outputs a unique signal of some capacity. Often properties like this are referred to as a ‘fingerprint’ of a unique object due to the visual nature of the device.

Unique objects are suited towards being used in a capacity that allowed them to be visually checked upon measurement, for signs of being tampered with. Hence, they are appropriate for use in confirmation of the authenticity of physical goods as opposed to the authentication of electronic devices in a virtual space.

Weak physical uncloneable functions are defined by the number of CRPs which the system can produce. The weakness of the system is derived from the small number of CRPs which it possesses and generally come from a linear increase of CRPs with respect to a property of the system. While it is described as ‘weak’, it is still useful as a form of security due to its unique outputs. However, to overcome its weakness, a requirement is that access to the CRPs is protected such that each response can be used multiple times. To aid with its protected access, an internal measurement system is required so that the responses from the system cannot be catalogued by a nefarious party.

Strong physical uncloneable functions are described similarly to their weak counterpart. While it’s unique outputs can be like that of a weak PUF the strength of the system is derived from the large set of CRPs. This large set tends to be a by-product of the likely exponential increase of CRPs with respect to some property of the system. As such, a system with such a large set of responses can be given to an attacker, with unlimited access and yet still be a secure system. This is such that, with the large set, all responses cannot be read in a reasonable time-frame nor can further responses be derived from previously categorised responses due to the unique and unpredictable nature of the responses. With the strength of the system, it is not necessary for the system to have an internal measurement system. The state of its measurement system can be left up to the nature of the representation of the strong PUF.

### 1.3.2 Challenges in PUF Technology

Since the inception of PUFs, the main obstacle to overcome is to prove that the physical representation of a PUF cannot be cloned. This being one of the main requirements of a PUF so that they should be uncloneable. Either their inherent random nature is unpredictable and difficult to reproduce, or outputting a unique and distinctive response such that a device trying to imitate the response would be impossible. Original requirements stated that the devices should be unable to be cloned, even by a manufacturer; difficult to fully characterise and difficult to access by the attacker. Therefore tamper resistances should be added to prevent this sort of unauthorized modification of the devices [14].

Some forms of PUF so far have been shown to lack these qualities through various means of manipulation by cloning of the physical device, mimicking a response, or tracking the outputs and using the database to authenticate a nefarious party. Using machine learning, an Arbiter PUF could be represented via a software model of the device predicting correct responses [15].

For higher security purposes, a software model becomes an impractical solution and a physical copy is required. An SRAM PUF, previously described in its operation, has been shown that it can be cloned creating a physical copy of the target response. This was achieved through the modification of a similar device via use of a Focused Ion Beam Circuit Edit (FIB CE) to thin the substrate and bias the nodes of the device. The two methods are the removal of transistors to achieve deterministic behaviours or a trimming of transistors to alter their performance. By allowing FIB CEs from the backside of the substrate, the modifications bypass the tamper-resistant protect on the PUF [13].

## 1.4 Quantum-Confinement PUF

In “Using quantum confinement to uniquely identify devices”, J. Roberts et al [1], uses Resonant Tunnelling Diodes (RTDs) as a weak physically uncloneable function with one CRP per device. The results given in this paper show that single RTDs, due to the uncontrollable atomic differences appearing during the fabrication process, produce unique signatures in their I/V spectrum.

RTDs, being the physical representation of a quantum well made from semiconductor materials, are shown to be easy to fabricate. The response of an RTD is difficult to reproduce due to the unpredictable nature of its atomic structure. The reproducibility of a single system is hence difficult to reproduce or copy due to the atomic differences in each RTD. However, these devices are shown to be easy to measure and robust in their measurement. RTDs can be challenged in an electronic circuit for which its current-voltage (I/V) spectrum is seen to be both stable and unique. Such that repeated measurements show very little deviation in peak position, found to be upwards of 95% confidence limits with a standard deviation of around 1.96.

Furthermore, the position of an RTD’s peak is split into 256 bins across both axes. The probability of a peak switching bins when re-measured is given at 11.4% and 0.54% for the x and y-axis respectively. Improvements to the devices are discussed by refining interfaces during the MBE growth of RTDs.

### 1.4.1 Improving the QC-PUF

The QC-PUF from the size of its CRPs such that each device only contains 1 response per device, thus constituting a weak PUF. However, due to the RTD being used as an electronic circuit component it becomes being viable to place multiple RTDs in a single circuit and therefore increasing the output of the devices. By varying the readable devices in a single circuit, an increased number of CRPs can be measured from a single system. Furthermore, depending on the number of devices which are read in a single circuit, it would be possible to increase the bit output of the devices from one to the number of devices in series. This can be derived from multiple peaks shown for any single response of an array due to the series nature of these electronic devices.

By increasing the RTDs in a single system and placing them in an array structure, an exponential increase in CRPs with respect to the devices can be achieved. The set of CRPs could be defined as a strong PUF due to the number of CRPs which a single system creates.

Being that RTDs are made from semiconductor materials and are very small (on the order of μm), this representation of a strong PUF would be ideal for IoT applications as a small, low power, cost-effective solution for security measures in each node of an IoT.

## 1.5 Contributions

This research expands on previously explored devices which have a single unique and robust response by combining multiple devices in various arrays. The merits of each array are explored and an optimal design is produced. By varying the dimensions of the array of resonant tunnelling diodes, it is possible to explore the effect on the set of responses. The array is shown to allow its outputs to be maximised and tailored to the security level required. The significance of this research is the increase from a single response PUF system to an exponentially increasing response system under the definition for a strong PUF.

The resonant tunnelling diodes are explored through the resulting effect when devices are combined in series. Combinations are shown to be unique to the constituent devices but convoluted enough that it becomes very difficult to find constituent devices. While some types of PUF design have been shown to be clone-able, this research verifies the difficulty and complexity of successfully cloning the PUF system described and designed herein.

The importance of a system with a large set of unique responses come from the need for secure systems as modern dependence on mobile systems increases. Conventional security falls short due to user negligence or weak security parameters such that nefarious parties can access systems with relative ease. The system described herein would increase the security of IoT systems and make unauthorised access difficult or impossible due to the inherent random nature of the devices used.

## 1.6 Outline of Thesis

In this research, we postulate the use of a known weak PUF and its unique combinations to give a representation of a strong PUF. By employing the classical interactions between quantum devices, we aim to prove the uniqueness of an exponentially increasing depiction of a strong PUF for each permutation and its inability to predict the output of each permutation even if given the output of the single devices or the previously attained responses.

In chapter 1, the background to the security of systems and their current state has been explored with the motivation for why a more robust security representation is required. A study of PUFs and different types available is given and the possibility of cloning on devices is explored. An introduction to a PUF using quantum effects and the postulation for an improvement from weak PUF to Strong PUF is discussed.

The next chapter, background and theory, covers a basic introduction to semiconductor materials for the purposes of use within this research and the cited research papers which provide some background. An Introduction to the theory behind confinement of carriers at a quantum level in multiple dimensions is given with the density of states for each of the confinement levels. A further explanation of the conduction mechanisms of resonant tunnelling diodes and the interesting negative differential region. Where the uniqueness of the device and various electronic implementations are explored including memory storage, oscillators and random number generators. Finally, a more detailed exploration into a quantum confinement PUF including an explanation of the uniqueness and robustness of single devices.

The third chapter detailing experimental methods includes fabrication of devices, the design of arrays of RTDs to produce a large set of challenge-response pairs, electronic characterization of the array system and deconvolution of the spectra. Fabrication details the MBE growth of devices along with the preparation and integration into an electronic circuit. Design of the array details a systematic improvement of the array design to increase challenge-response pairs while reducing exposure to potential side-channel attacks. An experimental setup to characterize the devices unique output is shown along with a deconvolution of the devices unique output. Lastly, a method to show robustness and uniqueness quantitatively is shown.

A results chapter follows which first explores the interaction between two devices to better understand how devices interact when in series. Robustness measurements and uniqueness overlap of devices are found which will show how defined each permutation is compared to subsequent permutations alongside measurement of the properties, the shift in voltage caused by devices in conjunction is explored along with how devices cluster with respect to the derivative devices and the device causing the shift. Further, 2x2 to 4x4 arrays are categorised similarly to the combination of in-series devices. The spread of data along with how well defined unique points are without any correlation in the spread. Lastly, the possibility of finding responses from the deconvolution of previous permutations is explored via a software model of the combination of devices.

A conclusion discusses the ability of the proposed strong PUF designed to create a useful physical authentication system. A final section discusses further work to improve the system and explore its possible weaknesses. The movement to a single complete system with the ability to be used in an IoT system is described with a more in-depth description found in Chapter 3.

# Chapter 2 – Background and Theory

## 2.1 Semiconductors

This section highlights the key concepts of semiconductors in solid state physics, which is for the purposes of using semiconductors as a form of PUF. Using the background, it can be understood where the uniqueness of physical structures can be found and what causes them.

### 2.1.1 Introduction to Semiconducting Materials

A semiconductor is a material which, in terms of its electron transmission properties, exist between a metal and an insulator (Figure 1), due to its narrow energy bandgap. The bandgap is the region which occurs between the lowest point of the conduction band and the highest point of the valence band. In a metal, the valence and conduction band overlap causing the bandgap to be non-existent hence the charge carriers can

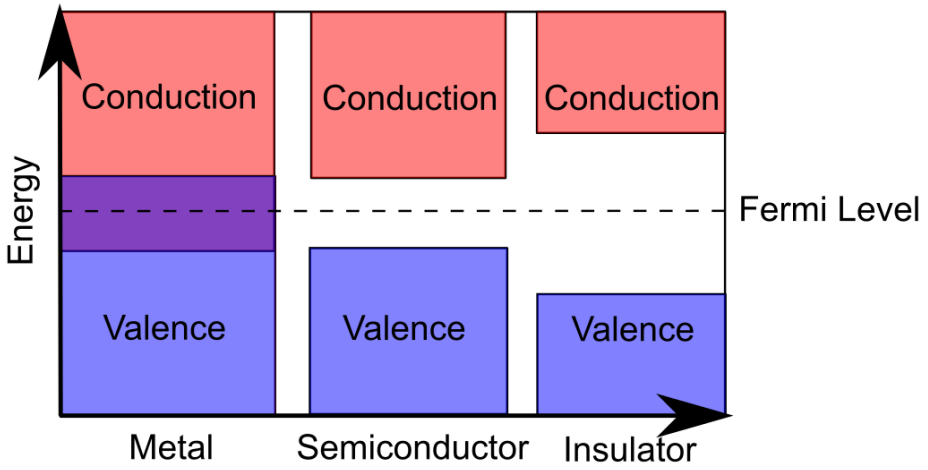


*Figure 1 : Simplified band structure of a metal, semiconductor and an insulator. In a metal where the valence band and the conduction band overlap, carrier transport is easy, whereas an insulator has a large bandgap and hence carrier need a large energy to pass the bandgap to the conduction band. A Fermi level where the 50% population probability is given at thermal equilibrium.*

pass through the material with ease. On the other hand, insulators have a large differenc in conduction and valence bands such that the transmission of charge carriers encounter a large difference in energy when traversing the material.

For charge carriers to populate the conduction band from the valence band, they require energy larger than or equal to the bandgap. The large energy required leads to the carrier population of the conduction band being negligible even when the material reaches a normal operating temperature (e.g. Room temperature). In contrast to both metals and insulators, semiconductor have a narrow band gap, such that carriers needn't have much energy to make the transition from valence band to conduction band but cannot freely make the transition as in a metal.

### 2.1.2 Fermi Level

The Fermi level (Figure 1), is defined as the hypothetical energy level where a 50% chance occupancy of an energy level occurs within the material at thermal equilibrium. With the use of doping (intentionally creating impurities in the crystal), the Fermi level can be shifted closer to the conduction band or the valence band. The shift occurs due to the doping material having a different number of electrons to the surrounding material. The shift of the Fermi level changes the electrical properties of the material, such that a Fermi level shifted towards the conduction band, an n-type semiconductor, allows the conduction band to be populated easily. However, a shift towards the valence band, a p-type semiconductor, makes it more likely for the upper states of the valence band to be empty. A p-n junction is created when these two materials are placed next to each other and is used in electronics for a variety of purposes such as to excite structures with a flow of electrons.

The Fermi level is also defined as the maximum energy that an electron can have at absolute zero. This is described by the Fermi-Dirac distribution (1) whereby at absolute zero the probability of an electron in a state above the Fermi level ($\mu$) is zero, and below the Fermi level is one.

$$f(\varepsilon) = \frac{1}{e^{(\varepsilon-\mu)/kT} + 1} \quad (1)$$

Where $\varepsilon$ is energy, T is absolute energy and k is the Boltzmann constant.

As the temperature increases, as does the likelihood that an electron can be excited into the conduction band, hence the occupation probability above the Fermi Level increases. As the system can be described as closed, an increase of conduction occupancy leads to the reduction the occupancy below the Fermi level.

### 2.1.3 Semiconductor PUFs

Integrated PUFs on a silicon wafer are most often the result of semiconductor process variations. Hence, even when an exact fabrication method is used, unpredictable outputs of systems can be produced.

A VIA-PUF [16] is a perfect example of a consistent fabrication method creating unpredictable results which can form the basis of a PUF utilising a semiconductor process. A via is a way to connect vertically adjacent layers of a device together. These are often used in printed circuit board fabrication, notably for the connection of grounding planes for the system. Here, the author of the VIA-PUF [16] proposes using the probability formation in circuit layouts. As the size of the via is reduced, the probability for the via to be successfully formed decreases accordingly. Thereby, a successful formation is assigned a digital value of 1, otherwise, is assigned a value of 0.

## 2.2 Quantum Confinement

Quantum confinement can be described in multiple dimensions, by confining it in one-dimension (creating a quantum well) up to three-dimensions of a quantum dot where the charge carrier is confined in all spacial dimensions.

### 2.2.1 Confinement in One-Dimension

A one-dimensional confinement system is described by a quantum well, which in an ideal model is described as infinitely deep. From the previous assumption, it removes the possibility of the carrier to escape when its energy reaches such a level that the state it occupies is more than that of the barrier energy. Quantum wells are described by a region which is occupiable by carriers and the surrounding region is classically forbidden for carriers to occupy. This creates a region where the carrier is confined between two classically forbidden regions. The quantum confinement effect occurs as the size of the occupiable region is reduced to a comparable distance to that of the de Broglie wavelength of electrons and holes.

To create a device in which these properties exist, layered heterostructures of semiconductors are epitaxially deposited on a substrate (3.2.1). Here, a material with a depth comparable to the de Broglie wavelength is sandwiched between two materials with a much larger bandgap. The fill layer results in confinement for electrons and holes, creating a quantum well. For a single electron in motion through a semiconductor crystal, the de Broglie wavelength is described by:

$$\lambda_{\mathrm{deB}} = \frac{\mathrm{h}}{\sqrt{\mathrm{m} * \mathrm{k_B T}}} \qquad (2)$$

Where m* is the effective mass of the carrier, h is the Planck constant and $k_b$ is the Boltzmann Constant. For an electron in a GaAs crystal (where m* = 0.067) at room temperature (300K), the de Broglie wavelength is ~42nm. Thus, quantum confinement effects govern the properties of the structure with a size on the order of tens of nm.

The properties of the devices used with the system have an effect which is governed by the confinement occurring within a quantum well. These effects can best be approximated using the ‘particle in a box’ approximation.

## 2.2.2 'Particle in a Box' Approximation

The 'particle in a box' approximation describes a particle's free movement in a small space surround by impenetrable barriers

The confinement causes the energy levels to become discrete states, described by the Schrödinger equation when applied to an infinite well:

$$-\frac{\hbar^2}{2m^*}\frac{d^2}{dx^2}\Psi_n(x) + V(x)\Psi_n(x) = E_n\Psi_n(x) \quad (3)$$

Where V(x) is the potential energy of the model, $\Psi_n(x)$ is the wavefunction which describes the fundamental behaviour of the particle such as position, momentum and energy. $E_n$ describes the eigenenergy of the system for each value of n, the principal quantum number. Further, $\hbar$ is the reduced Planck constant and m* is the effective mass of the carrier. Considering the infinite well height and the condition that the

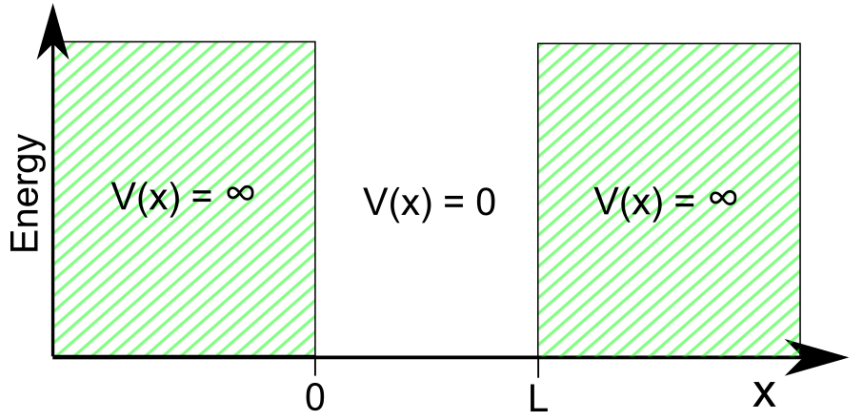


*Figure 2 : Ideal quantum well representation where the well is infinitely deep and has a width of L where the boundaries of the well are 0 and L along the x-axis. It is assumed that the potential outside of the well is infinite as this a classically forbidden region*

particle cannot exist outside of the well under the condition:

$$V(x) = \begin{cases} 0 \\ \infty \end{cases} \quad (4)$$

The solutions become relatively simple such that:

$$E_n = \frac{\hbar^2 k^2}{2m} = \frac{\hbar^2}{2m}\left(\frac{n\pi}{L}\right)^2 \quad (5) \qquad \Psi_n(x) = A\sin\left(\frac{n\pi x}{L}\right) \quad (6)$$

For these solutions, the first three energy levels are shown in Figure 3. An important characteristic of these solutions is that $E_n$ is inversely proportional to the square of L. This, for the purposes of unique identification, shows that even atomic differences in the width of the well will change the discrete energy at which the energy states occur at.

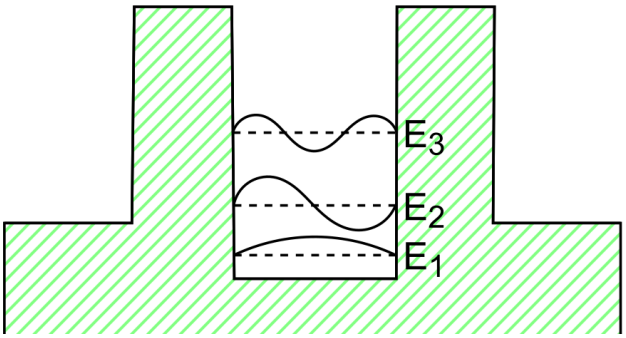


*Figure 3 : Energy levels of a quantum well structure. The first three energy levels are occupied by the first three harmonic wavelengths.*

In reality (due to the finite barrier height) the energy levels are confined to those which occur under the barrier height and also leads to quantum tunnelling of the carrier wavefunctions into the barriers. Interestingly, the well width can be tuned such that the eigenenergy allows for the emission of a specific wavelength. With these solutions and as can be seen in Figure 3, the first energy level (zero-point energy) is not zero, this is a product of the Heisenberg uncertainty principle, given by:

$$\sigma_x \sigma_P \geq \frac{\hbar}{2} \tag{7}$$

where $\sigma_{x,}\sigma_P$ are the standard deviations of position and momentum respectively.

Due to the particle being confined in a region of space, the variation on its position is confined also. To avoid violating the uncertainty principle, the particles moment cannot be zero, such that hence the particle must have some finite energy. As the well-width increases, its position becomes more defined, and hence the momentum and zero-point energy must increase to compensate.

## 2.2.3 Confinement in Multiple Dimensions

Confinement in one-dimension can be easily expanded to approximate confinement in an increasing number of spatial dimensions. As such, two and three-dimensional confinement can be represented by a quantum wire and quantum dot respectively.

The density of states (often used to characterise a given quantum structure) elaborates on the number of states per energy level. Throughout all the representations of quantum confinement, it is shown that as more dimensions of confinement are introduced the density of states becomes more discrete. This is most evident when 3D confinement in a quantum dots shown fully discrete energy levels.

$$\mathrm{DOS}_{\mathrm{bulk}} = \frac{\sqrt{2}}{\pi^2\hbar^3}\mathrm{m}^{3/2}\mathrm{E}^{1/2} \qquad \mathrm{DOS}_{2\mathrm{D}} = \frac{\mathrm{m}}{\pi\hbar^2}$$

$$\mathrm{DOS}_{1\mathrm{D}} = \pi\sqrt{\frac{2\mathrm{m}}{\mathrm{E}}} \qquad \mathrm{DOS}_{0\mathrm{D}} = 2\delta(\mathrm{E}-\mathrm{E}_{\mathrm{C}}) \tag{8}$$

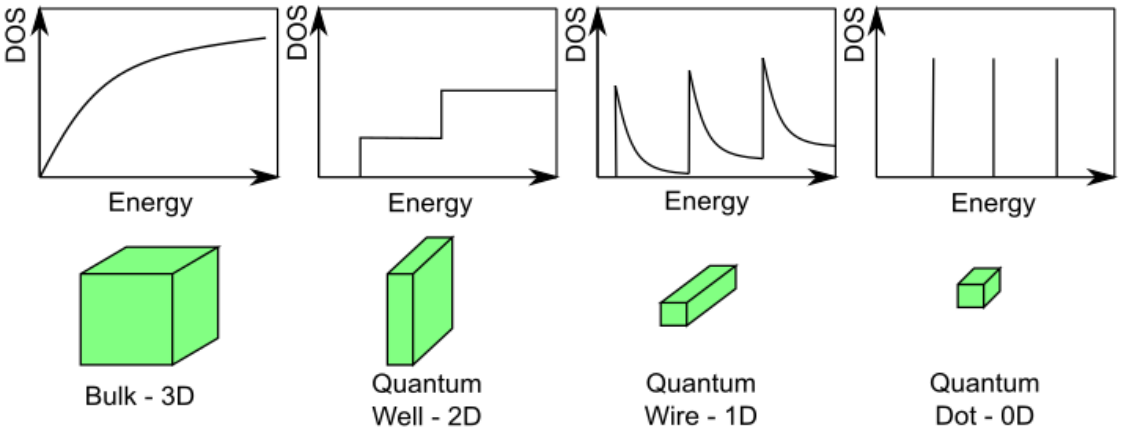


*Figure 4 : Graphs which show the dependence of the density of states on the energy of the system for a bulk material with three degrees of freedom, a quantum well with two degrees of freedom, and quantum wire with one degree of freedom and a quantum don't with zero degrees of freedom.*

It is interesting to note here that a quantum well with 2 degrees of freedom, that the density of states does not depend on the energy of the system. A quantum dot acts much like a single atom in its density of states such that it has fully discrete energy levels much like the energy levels of a single atom for its excitation states.

## 2.3 Resonant Tunnelling Diodes

Resonant Tunnelling Diodes (RTD), as the physical realisation of a quantum well, employ quantum confinement and tunnelling as the main mechanisms for the transfer of electrons through the system. The resonant tunnelling diode is characterised by its N shaped current/voltage (I/V) curve, which is a product of its negative differential resistance (NDR). The NDR is a region of the curve for which as the voltage continues to increase, the current decreases. Therefore, RTDs find use in electronics which lead to interesting effects which can be utilized by electronic technologies.

Resonant Tunnelling Diodes are made up of a thin, narrow band-gap material (e.g. InGaAs) sandwiched between wide bandgap material (e.g. AlAs) which forms the double barrier and well region of the quantum well. Highly-doped, narrow band-gap materials are placed either side of the well structure forming an electron source/sink, commonly referred to as the emitter/collector regions. Due to the finite height of the barrier, a finite number of energy levels can be observed. In the case where the well height is small enough we only see the first energy level.

### 2.3.1 Electron Transport Mechanisms

The RTDs I/V curve can be simplified down into 3 sections: Tunnelling region, NDR and the Thermionic region. Each of these sections is due to an electron transport characteristic or the switch between electron transport mechanisms.

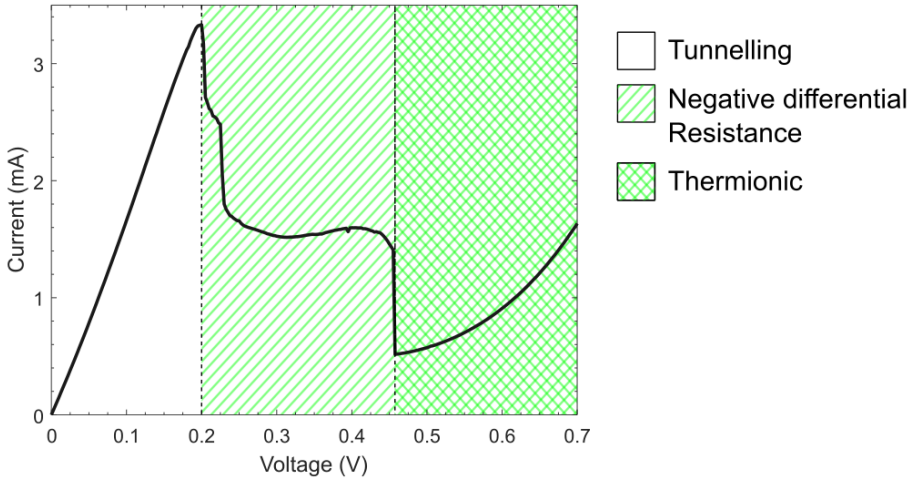


*Figure 5: Single Current/Voltage spectra of a resonant tunnelling diode divided up into the conduction mechanisms which is the dominant mechanism given as Tunnelling and Thermionic. The negative differential is shown as the drop in current from a increase in voltage due to a change in conduction mechanism.*

The first section of the curve, the tunnelling region, is primarily due to the quantum tunnelling effect. This mechanism is what gives the device its name. Here, the curve increases almost linearly with the amount of voltage applied. This is due to the electrons, with varying energy, coming into resonance with the first energy level of the quantum well structure. As the Fermi level of the emitter comes into resonance with the first energy level current flows. With an increasing voltage bias and the electrons becoming more densely packed as the energy level shifts to the lower edge of the conduction band in the emitter region, we see the current increase as more electrons come into resonance with the first energy level.

A sudden drop in current is observed when the energy level passes beyond the lowest energy of the conduction band where no electrons exist. This shift sees the main transport mechanism switching to a thermionic emission of the electrons which have enough energy to pass over the top of the barrier. As more of the electrons have the energy to bypass the barrier region, they will continue to do so, hence causes an exponential increase of current. During the switch of mechanisms, the current is not reduced to zero, as some electrons will already have the required energy to bypass the barrier region.

### 2.3.2 Negative Differential Region

The negative differential resistance (Figure 5) will occur at a range of currents and voltages due to the ability to shift the energy level by small amounts as shown previously with the energy levels inverse proportionality to $L^2$. However, due to how this region is created, it is naturally unstable as such the system will show signs of varying current in this region as it switches between the tunnelling and thermionic mechanisms. Due to the uncontrollable creation of the energy level in the well, the peak position will occur at a unique point per device, as no device can be reliably recreated in the knowledge that it will clone another device. The slight tuning of the well width affects the energy level and hence affects where the peak will be found. It can be seen in Figure 5 that the NDR region has multiple plateauing regions where the current seems to stagnate, this is thought to be due to trapping of a charge in the confinement region which is shifting the energy level up.

The NDR is useful in many physical systems of various electronic technologies. Often it is exploited for its fast switching of mechanisms, the stability of the carrier mechanism and the difference in its signature depending on its driving source.

### 2.3.3 Implementing a Resonant Tunnelling Diode

Resonant tunnelling diodes have a range of uses within electronic components because of their fast switching capabilities. The most prominent use of the RTD is to use the speed at which it can switch from peak-to-valley. This is useful in high-frequency oscillators for use within signal generation technologies. The RTD has been shown to be able to produce signals into a THz range [17-20].

Further switching capabilities have been demonstrated by Sung-Yong Chung et al, by using RTDs in a three-terminal bipolar transistor [21]. As such, the designed transistor shows an ability for an adjustable peak-to-valley current ratio. The design implementation allows amplification and switching of high frequencies commonly used in modern circuits and radio-frequency systems.

Due to the nature of an RTD and its two-state electron transport mechanism, it has been shown [22] that by connecting RTDs as circuit elements, that memory cells can be created. By connecting multiple RTDs together, multiple states can be achieved. As such, a normal binary system achieves two-states described by 0 or 1. A three and four state system of memory storage is presented by RTDs.

Electronic systems, which until recently have used software code to simulate random numbers, have shown advancement using an RTD. R. Bernardo Gavito has shown the ability to create a true random number generator using a current driven RTD which outputs random switching behaviour between states [23]. As opposed to a peusdo-random number generator which uses a complex algorithm based on many factors to generate a seemingly random number, the design specified outputs a random string of 0 and 1 bits. In the same way, in which a memory implementation allows for multiple states for storage of bits, the same method can be applied. This would allow each device to output multiple bits per input.

## 2.4 Resonant tunnelling Diodes in Authentication

In secure communications, each node needs to have secure access such that only valid users can access the information or services therein provided. This authentication requires each user to have a unique and impossible to replicate code such that access cannot be gained from users not authorized to have access to the node. An RTD outputs a unique signature derived from the energy level which changes per device due to the sub-monolayer differences in the width of the quantum well. The Quantum Confinement PUF, QC-PUF, describe by J. Roberts et al [1] takes the unique output of an RTD and uses it as an identifying secret for which authentication can be achieved. Each RTD shows a single Challenge-Response pair per device and is a unique physical object such that it constitutes a weak PUF.

Measurements showing unique, reproducibility of single RTD are performed on 4$\mu m^2$ devices which show a peak range of 70mV and 4mA. Devices, as explained in subsequent sections, are shown to be unique through lack of overlap of peak positions and a robust signal output over repeated measurements.

### 2.4.1 Peak of Tunnelling Current

The peak at which the device switches mechanism (the point where the energy level is in resonance with the conduction band), is shown to be unique. This is shown by each device having a peak which occupies a different area within a current-voltage plane. Figure 6(a) shows the average peak position for 26 devices over 100 spectra. The peak position is calculated using a Gaussian fit as it can be used to approximate the tunnelling current. There is shown no overlap between the peaks, however, Figure 6(b) shows the red boxed region in which the average peaks seem clustered, but it can be seen that there is no overlap between devices. By using 1.96,3.09 and 3.99 standard errors for the 95%, 99.95 and 99.997% confidence is plotted showing the reproducibility of subsequent measurements.

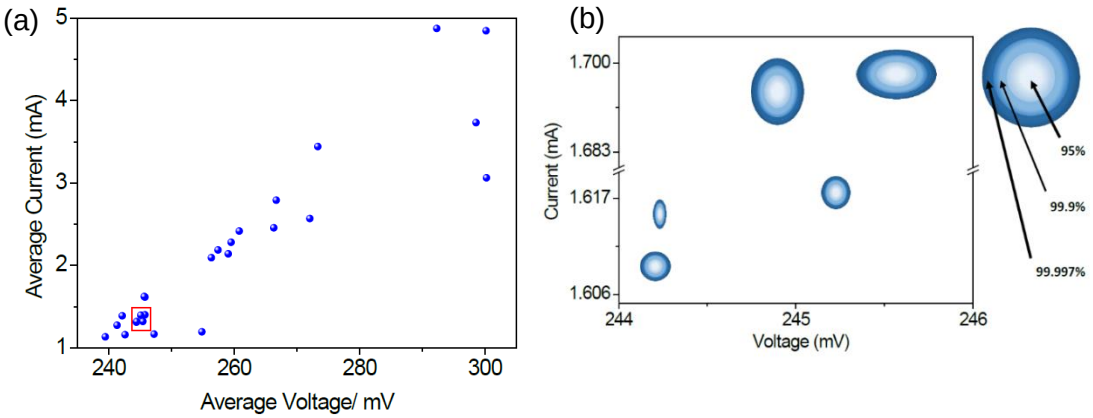


*Figure 6 : (a) 26 device peak position of resonant tunnelling diodes extracted using a Gaussian fit. (b) Red Boxed region in (a) plotted with standard deviations of 1.96, 3.06 and 3.99 for the 95%, 99.95 and 99.997% confidence regions. (Figure courtesy of J. Roberts from: Using Quantum Confinement to Uniquely Identify Devices [1].)*

## 2.4.2 Robust Current/Voltage Characteristics

For RTDs to be effective as a unique form of authentication and identification need to show robustness in their output such that the output is reliable and consistent. It is shown that for a single device, the I-V spectra is consistent with 100 measurements of a single device. Hence, showing the reproducibility of a single peak position and the differential current-voltage. Repeated measurement consistently lies within 2 standard errors of the average value. This repeat measurement is shown in Figure 7 where the spectra are offset for clarity. It is noted that an average measurement would be used for the implementation of the device to reduce the possibility of a false reading.

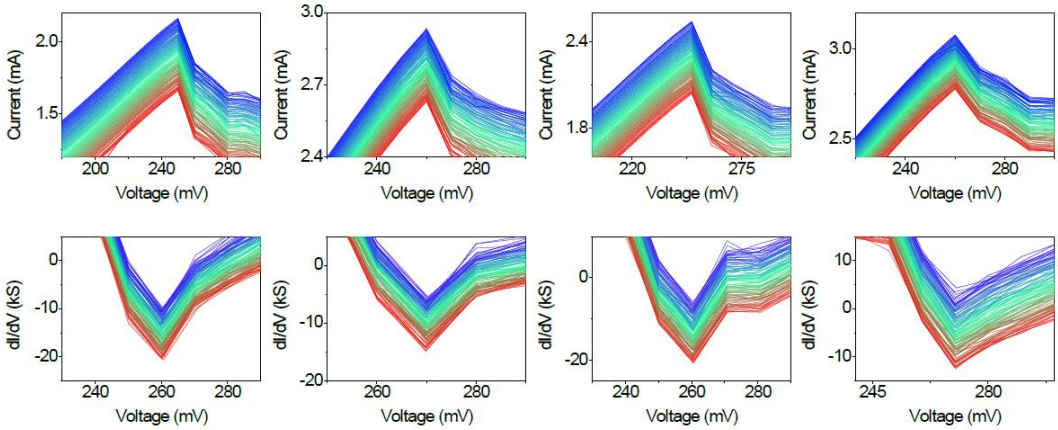


*Figure 7: Current-Voltage spectra of 4 single devices where the measurement is repeated 100 times. Spectra are offset for clarity of robust peak position. The current-voltage differential is also given for robustness of the spectrums properties. (Figure courtesy of J. Roberts from Using Quantum Confinement to Uniquely Identify Devices [1].)*

# Chapter 3 – Experimental Methods

## 3.1 Arrays of RTDs

Arrays of RTDs can be created in many ways, which can affect the speed at which the CRPs increase, the number of CRPs per system and the security of each CRP. Yet a balance must be struck so that an increase is achieved, yet information on the implementation of the physical system is not freely given away to an attacking party in the event of a side channel attack. The number of devices, d, used in these implementations can be described by their position in an M by N array corresponding to the rows and columns, respectively.

### 3.1.1 Addressable Array

Each RTD contains a single CRP, while each CRP from each RTD is unique due to the properties described previously, but it does not provide a large enough set of CRPs to make it useful in a ‘strong’ authenticable system.

Combining single RTDs with a single CRP, each RTD is addressable as an individual would increase the set of CRPs for a single system. An array of devices which can be uniquely addressed will have a set of CRPs equal to the number of single devices in the system.

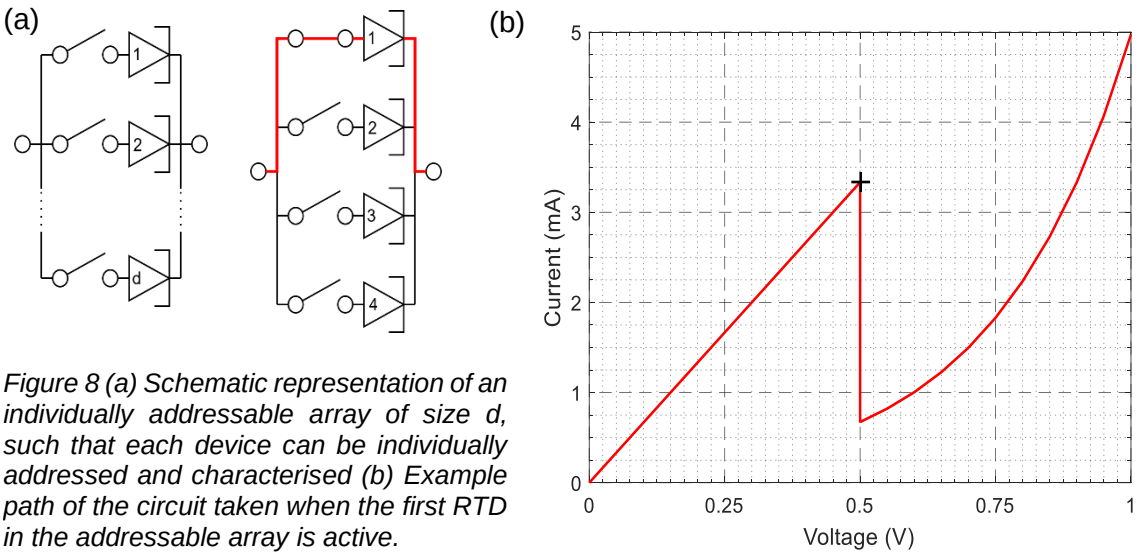


*Figure 8 (a) Schematic representation of an individually addressable array of size d, such that each device can be individually addressed and characterised (b) Example path of the circuit taken when the first RTD in the addressable array is active.*

An array with each device in parallel with each other (e.g. 1 column) will cause the devices to have less in-series resistance as each switch will be dependent on every other switch in the array and hence have no more than 1 switch in-series. This means only 1 device can be switched on for any 1 sweep, and each device can be characterised individually.

However, as each device still only creates 1 CRP, this does not constitute a large enough set or an exponential increase in CRPs. By restricting the access to the system, this would constitute a weak PUF as each device has been shown to be random and unique.

### 3.1.2 Square Array

A Square array is represented by having an equal number of rows and columns. The rows and columns are defined by the number of total devices, d, and the square root gives the number of both rows and columns, M. For example, 16 devices give a 4 by 4 array. Each device would be paired with a switch, meaning that each switch can be in either an on or off state, independent of every other switch state.

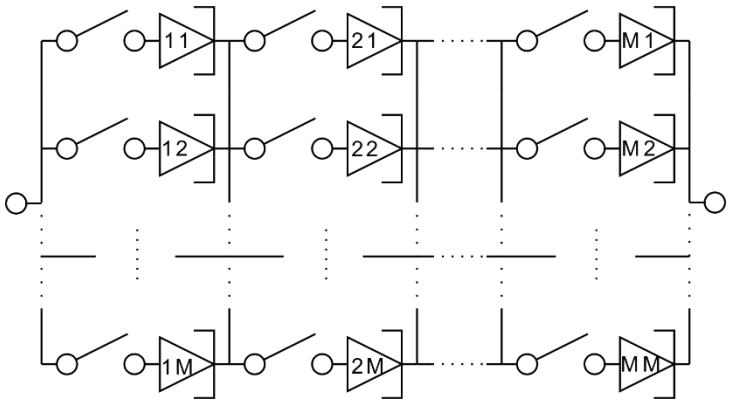


*Figure 9: (a) Schematic representation of a square array of size M, where $M^2$ equals the number of devices. Each RTD is coupled with a switch for turning each one on and off.*

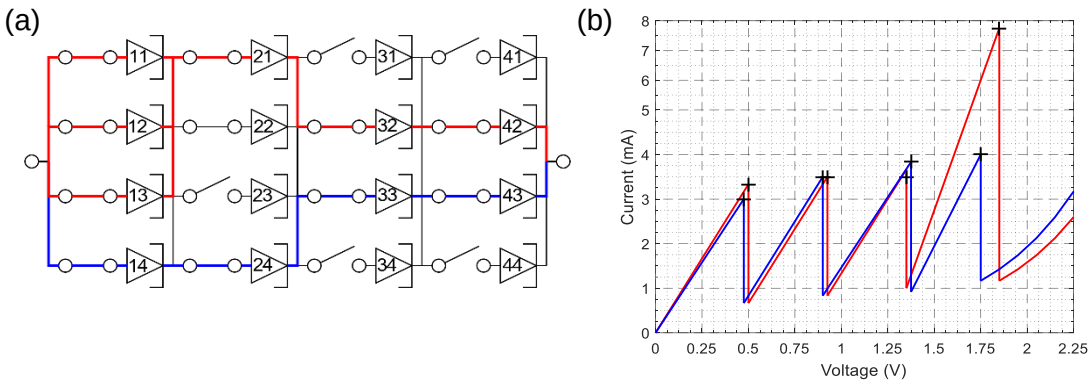


*Figure 10 : (a) Example permutation paths through a 4x4 16-device array* $(2^4 - 1)^4 = 50{,}625$ *permutations excluding open permutations. (b)Current/Voltage trace of the example paths (a) such that 1 permutation contains only series devices. A second path with a set of parallel devices which hence gives a larger peak in the 4th peak position.*

As each device has 2 states, there are $2^{M^2}$ permutations in this array representation. However, for each column, if no device is selected then the circuit for the array will be open. Therefore, each column will have 1 less usable permutation, so it will have $(2^4 - 1)^4$ permutations. Hence, a 16 device, 4 by 4 array would $(2^4 - 1)^4 = 50{,}625$ permutations which are not open circuits. With an exponential increase, this would constitute a strong PUF, which increases very quickly, thus the number of permutation gets very large, even dwarfing previous permutations numbers as the devices increase.

While this array has the largest growth rate of any of the representations stated here, the issue lies in how the system is setup and its outputs. Parallel devices cause a superposition of their peak position elevating to a higher current hence requiring more power to run. Furthermore, as the number of devices in a permutation can vary from a minimum of M to d, this would allow an attacker to gain valuable system information in a side-channel attack.

It should be noted that while an extra parallel switch can be added to the bottom of each column to allow a column where a device is not selected to be bypassed. The number of peaks which can be seen would decrease for the permutations which were previously open. For 16 permutations, they would have individual devices and hence only 1 peak, and no combinations. So, the number of peaks will then vary from 1 to M, and devices in each permutation can vary from 1 to d.

### 3.1.3 Linear Array

A linear array is represented by all devices existing on a single row, with the same number of columns as devices, where a second row is populated by only switches (Figure 11). Each device is accompanied by a switch and a further switch in parallel where one is in an ‘on’ position (Figure 12). This representation works by allowing only a select, consistent number of devices to be in the circuit in any permutation, e.g. 2, 3, or 4 devices. Keeping the number of devices in each permutation constant decreases the available physical information which is helpful to an attacking party. This representation allows each device to be paired with every other device, without repeating a permutation of devices in the system.

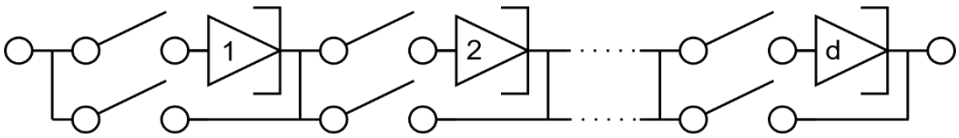


*Figure 11: (a) Schematic representation of a linear array, comprised of d devices in series. Each device is coupled with two switches which are in parallel with each other.*

This representation’s permutations can be denoted by

$$\sum_{i=1}^{d-n+1} i = \frac{(d-n+1)(d-n+2)}{2} \tag{9}$$

Where d is the number of devices and n is the number of devices per permutation. This formula gives a polynomial increase, which by the definition of a strong PUF is not a large enough increase to allow adequate security. However, with a large enough set of devices, the system may constitute a large enough set of permutation that it could be described as a strong PUF.

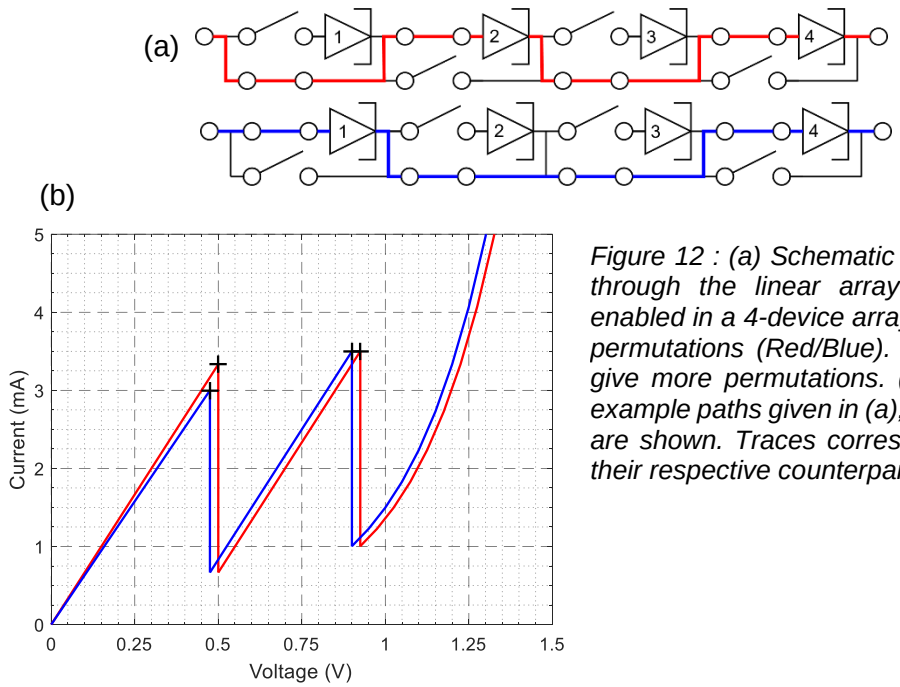


*Figure 12 : (a) Schematic of potential paths through the linear array with 2 devices enabled in a 4-device array with (3)(4)/2 = 6 permutations (Red/Blue). Longer array will give more permutations. (b) Traces of the example paths given in (a), where two peaks are shown. Traces correspond in colour to their respective counterparts in (a).*

## 3.1.4 Dependent Switch Array

The following design is based on the square array; however, the design differs in one aspect, each switch becomes dependant on the status of the switches in the same column as it (Figure 13). In this representation, each column can only have one switch active, hence no devices will be in parallel with any other devices (Figure 14). The permutations for this representation of d devices will have M different choices of the active switch for each column, and N columns. Hence $M^N$ permutations of different devices establishing an exponential growth for the design. The exponential growth would constitute a large set of CRPs for this representation to be classed as a strong PUF.

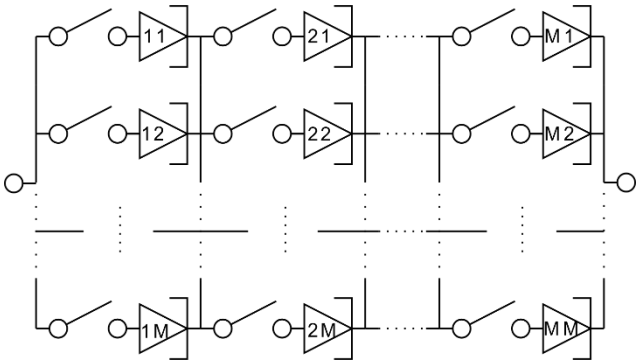


*Figure 13: Schematic representation of a square array, where each RTD is set up the same as before, yet each switches state is based on the state of each switch in parallel with it.*

While this representation does decrease the permutations per device from the previous square array, it does avoid the use of excessive power due to high currents from the superposition of parallel devices. Hence this representation while being a lower power which is a requirement for that of integration with electronics with limited power. Additionally, with a constant number of devices per permutations, it would reduce the exposure of the system to information leakage.

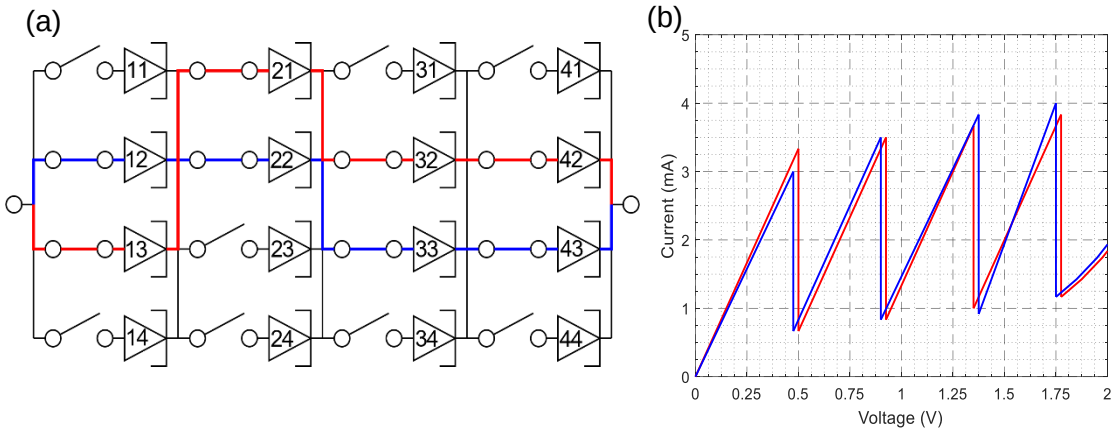


*Figure 14: (a) Schematic representation of two example paths through a 4x4, 16-device array with $4^4 = 256$ permutations (Red/Blue) (b) Current/Voltage example traces of the two example paths given in (a). Traces correspond by colour to the paths respectively.*

### 3.1.5 Rectangular Array

While the design of this strong PUF array has the same concept as a dependable square array, it has a few layout differences to maximise the permutations that can be output from this array (Figure 15). By changing the ratio of M: N, rows: columns respectively, the array is then rectangular in nature. The rate at which the exponential growth occurs can be tuned to achieve maximum CRPs.

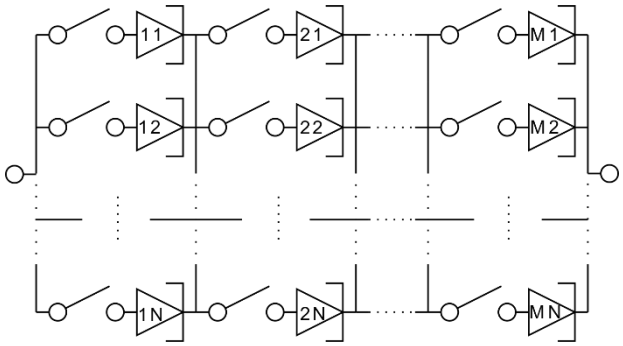


*Figure 15: Schematic representation of a rectangular array where each RTD has the same coupled switch setup as the 4.3.1 and 4.3.3 square array. The columns (M) and rows (N) can vary here to increase the number of permutations. However, this means that the number of peaks seen will be equal to the number of columns.*

As previously discussed, the general expression which gives the number of permutations of the system is given by $\mathrm{M}^{\mathrm{N}}$. Generally, the permutations growth is proportional to the number of permutations, as an exponential growth rate. The expression for the growth of the system is given by

$$\mathrm{NM}^{\mathrm{N}-1} \tag{10}$$

However, this expression is constrained by $\mathrm{M} \times \mathrm{N} = \mathrm{d}$, therefore if M or N is increased, the alternate variable will be reduced. M and N will also be constrained to only have values equal to that of the factors of d thus the constraints are satisfied, and the array is of a quadrangle nature (Figure 16).

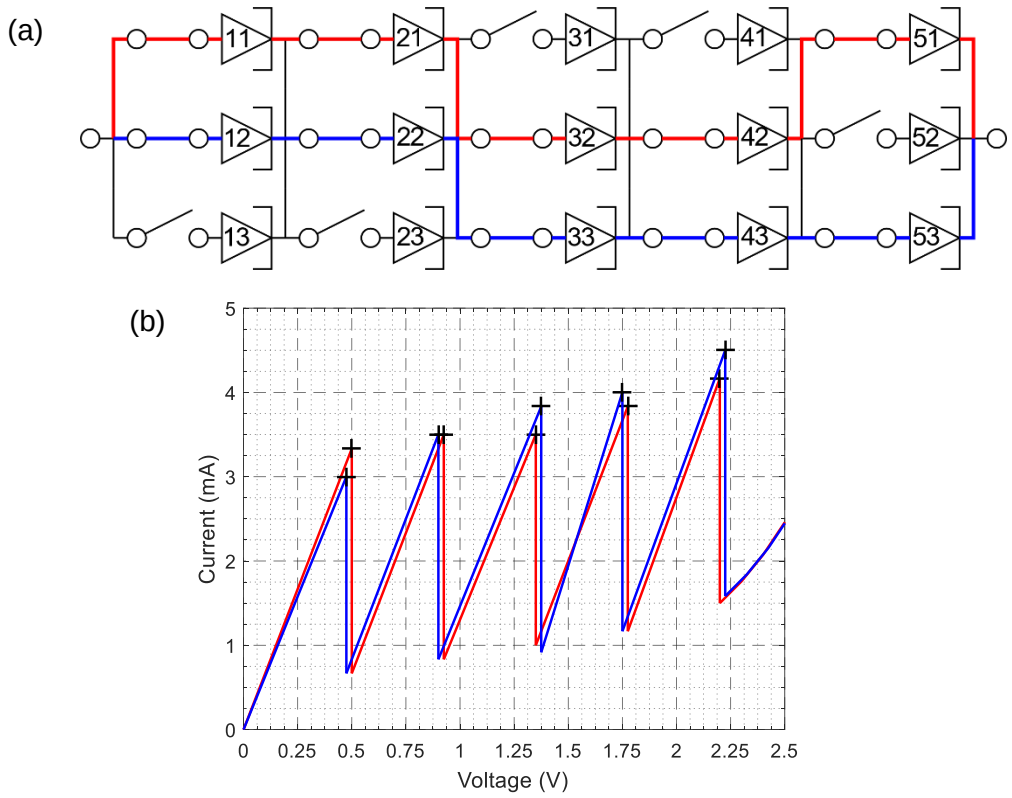


*Figure 16: (a) Schematic representation of two example paths through a 5x3, 15-device array with $3^5$ = 243 permutations (Red/Blue). (b)Current/Voltages example traces of the two example paths given in (a) for the 5x3 array. Traces correspond by colour to paths respectively*.

By keeping the number of devices constant, it can be noted that increasing the value of M (columns) will increase the number of permutations which the array will output. Whereas, increasing the value of N, rows, will decrease the number of available permutations. This is shown with examples where N and M take a value ranging from 1 to 20 (Figure 17). It can be seen in Figure 17, a vertical array (N>M) has the lowest output of permutations per device. Opposing that, a horizontal array (M>N) gives the most permutations per device. The largest increase in permutations per device is a point where the number of rows is between 2 and 4, centred on *e*. This is due to the increase being exponential that as N decreases, (and M increases) the rate at which the permutations increase decreases drastically.

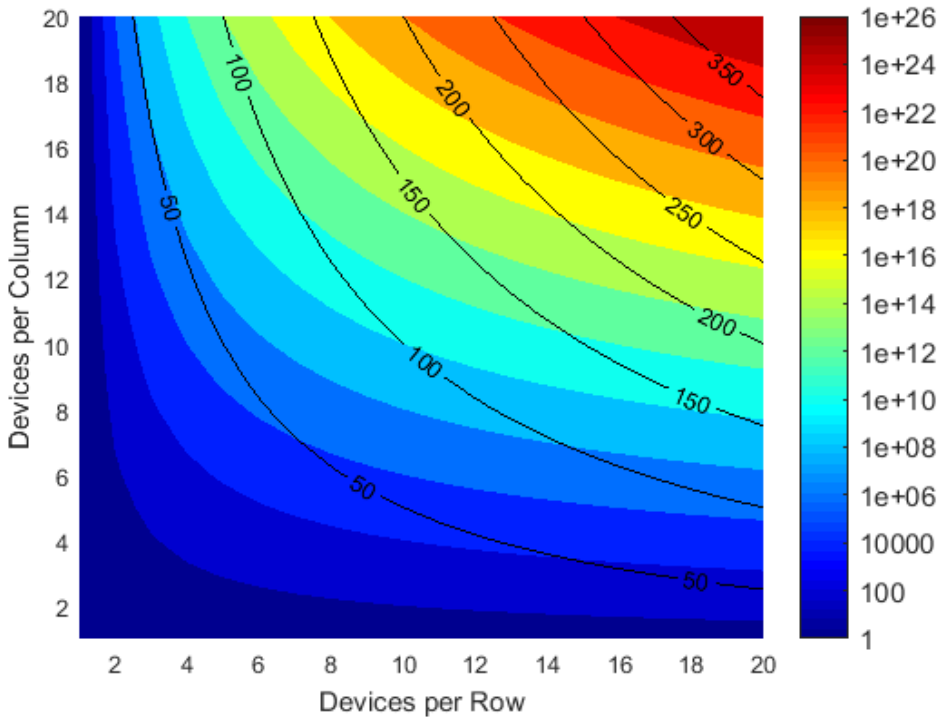


*Figure 17: Representation of challenge-response pairs from different number of rows and columns in an array. The colour bar gives the number of responses for an array. Black contour gives the number of devices total in a given array.*

## 3.2 Fabrication

### 3.2.1 Molecular Beam Epitaxy

The diodes used are made by M Zawawi et al, at Manchester University and are further described for readers in their paper, 'Fabrication of Submicrometer InGaAs/AlAs Resonant Tunneling Diode Using a Trilayer Soft Reflow Technique with Excellent Scalability'. The process for their production is described in a dedicated paper.

Structures are fabricated in a Molecular Beam Epitaxy to generate an InGaAs/AlAs double barrier on an InP substrate. The double barrier structure is surrounded by a 20nm InGaAs spacer layer, to prevent dopants from diffusing into the un-doped region of the barrier. Beyond the spacer layer are emitter and collector layers made from 25nm doped InGaAs. The emitter contact is created from highly doped 45nm InGaAs, while the collector contact is 400nm highly doped InGaAs.

Initially, the width of the emitter is set to be 1µm by i-line photolithography. A soft re-flow technique is used to reduce width down to 350nm such that the emitter metal comprised of titanium (50nm) and gold (200nm) onto the surface of the highly doped InGaAs emitter. Using the top metal contact as a hard etch mask, a Reactive Ion Etch (RIE) with $CH_4$: $H_2$ (1:8) for 20 mins results in a 210nm anisotropic sidewall down to the collector layer.

A further wet etch to define areas by using photoresist and UV-photolithography. With an etch rate of 90nm/minute using $H_2O$: $H_2PO_4$: $H_2O_2$ at 50:3:1 to etch 300nm down to the InP to isolate the devices. This final etch also provides the lateral undercut on the air bridge. Finally, the collector contact is deposited by thermal evaporation of titanium/Gold (50nm/500nm).

### 3.2.2 Device Preparation

Fabricated chips contain an array of mesa sizes of 4 $\mu m^2$, 9 $\mu m^2$, 16 $\mu m^2$, 25 $\mu m^2$, and 36 $\mu m^2$, where the device characteristic is proportional to the size of the mesa region. Measurements in this work were all performed on a single size of the device to simulate the most probable overlap scenario. However, there is no qualitative difference between the different sizes of the mesa region so that any size of mesa would be integrated into the designed system. In order to show the most overlap possible, all the devices used have a mesa size of 36 $\mu m^2$, 16 of which are used in the largest array proposed therein.

RTDs are bonded from the emitter and collector regions to a ceramic chip carrier using a TPT HB05 Ultrasonic wire bonder. Connections are made using 25 μm-diameter gold wire. The chip carrier is placed into a 28-pin IC socket on a PCB board, whereby connections to the array circuit are made using standard copper wires and connectors for the possibility of variation of devices. Measurements are all taken at room temperature using methods provided in subsequent sections. During the measurement process, devices are covered with a 3D-printed cover used to protect devices from unwanted dust or damage.

### 3.2.3 Device Integration

For each RTD to be addressable, it is coupled to a switch (7066N Quad Bilateral) in series. This is achieved by connecting each device to the array independently of others to allow for damaged devices to be removed and devices to be switched to vary the device signatures. Furthermore, for the array to vary from 1 by 1 to the maximum number of rows and columns, each column is paired with 1 further switch. Hence the number of switches used will be the maximum number of devices and columns to give an array which can vary in size. While this is not necessary for an array with a fixed number, the variation of the array allows increased practicality of the physical system. Variation in the size of the array allows the full categorisation of all array sizes for experimental tracking of the interaction between RTDs.

A (7066N) Quad-bilateral switch requires a 3.3V or 5V source of power, depending on the required resistance. A higher voltage often provides a lower resistance, but with that, the activating voltage is proportional to an increased powering voltage. This would hence require balancing with respect to being a low-power system. With the use of an Arduino which outputs 3.3V to the I/O pins, a 3.3V source for the switch allows some variation in the I/O voltage such that activation pin on switches will never be underpowered.

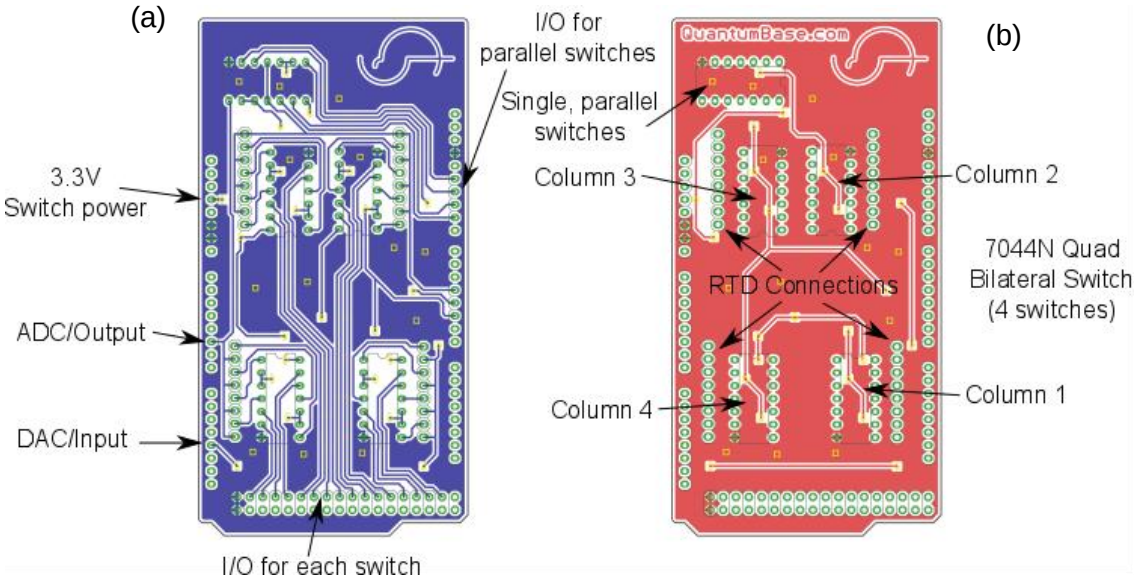


*Figure 18: (a) Bottom board layout for printed circuit board of variable 4x4 array. I/O pins used for activating switches can be seen at the bottom on the schematic. Input of RTD array structure labelled (DAC/Input), Output labelled (ADC/Output). DAC/ADC lines such that the array structure can be easily inserted into a setup which does not require a Source-Measure Unit (b) Top board layout for printed circuit board. In anti-clockwise manner starting from the bottom right corner, 4 switches which control the RTDs in each column, denoted respectively in the image. Connections for where each RTD can be connected are denoted. RTDs connect sequentially, input-output.*

## 3.3 Electronic Characterization

The use of the designed array which will be described in following sections requires a few modules in its design. In a complete PUF system, some modules in the current design become redundant as other systems can replace them to reduce size, cost and weight of the system while providing a similar or equal purpose.

### 3.3.1 Source-Measure Unit

The voltage sweeps, and measurement of current are provided by a Keithley 2602B Source-Measure Unit (SMU). This can be carried out by other devices in the design, but for more accurate measurements and a stable source, the SMU is advantageous. The downside of this module is that it is not low power and is bulky, thus, is not practical to use in an IoT system which may also be a mobile system. The design of a more complete and smaller representation will be detailed later in this section.

### 3.3.2 System Control Modules

The control of the switches in the array is given to an Arduino Due. Connected by stackable headers to the PCB (Fig) containing the array of switches and headers for which the devices can be connected into, the Arduino can address each switch individually. This means that using a complete variable array, alongside having an M by N array, we can also measure each RTD individually with the representation given in **3.1.5.** The Arduino receives the challenge through the serial port which, in turn, is decoded to give the code for which of its I/O ports shall be switched to an active position of ~3.3V. The switches are first all switched to an 'off' position before each required switch is switched to its 'on' position completing the circuit through the chosen RTDs.

The SMU is controlled by a Raspberry Pi 3, which contains the main program used to run the array and acts as a central hub to the representation. It uses the number of RTDs along with the array dimensions to create a permutation number which it cycles through to create the challenge. The challenge is subsequently passed to the Arduino for each permutation for the control of the required switches. It then passes a set of instructions to the SMU for the sweep and records the data, saving it to a file and displaying it on the screen.

### 3.3.3 Complete Strong PUF system

A full system includes a controlling device, e.g. Raspberry Pi, a system to control the input of the array system, e.g. Arduino Due, and a voltage input/current measurement device, Keithley 2602B SMU. As the Raspberry Pi is the controlling module, all inputs and outputs are channelled through this, hence data management and variability of inputs are given here where they are passed to the Arduino to fulfil the request for a specific pattern which outputs a specific signal. The Keithley is controlled directly by the Raspberry PI as a slave module retrieves data and allows the Raspberry Pi to read the dataset.

A full database containing controlling programs and PCB schematics is given for interested readers at https://github.com/Benasbo12/StrongArray.

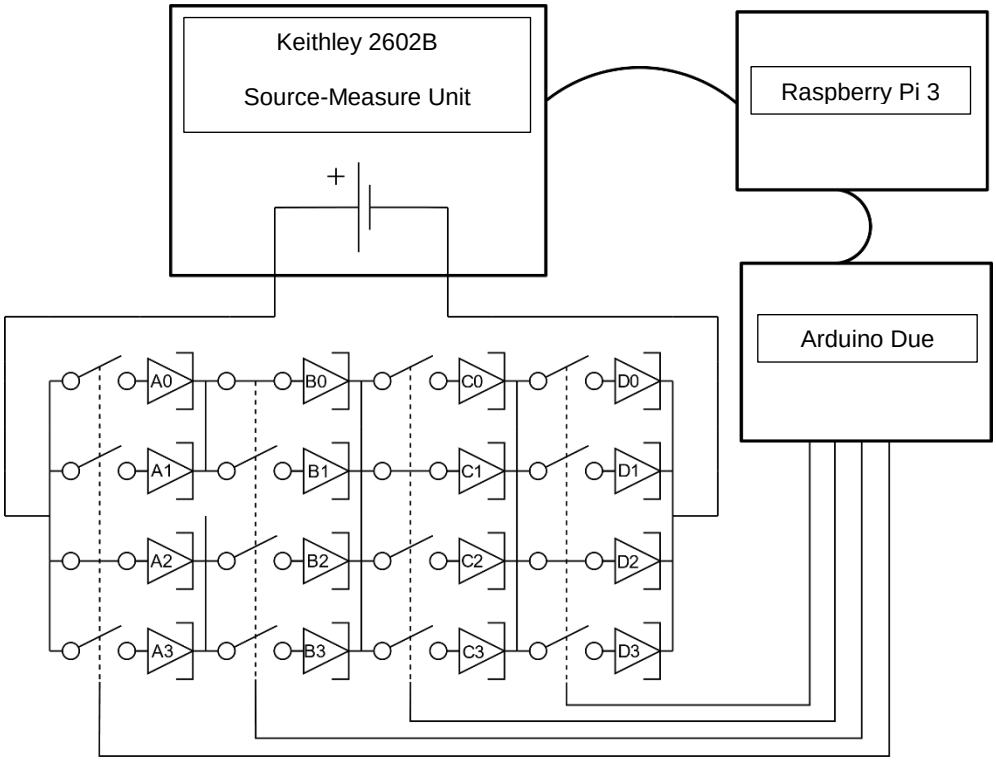


*Figure 19: Schematic of Array system including the labelled Keithley 2602B Source-Measure Unit (SMU), Raspberry Pi 3, Arduino Due and the array structure. The array structure here is interchangeable for each of the previously discussed array representations.*

### 3.3.4 System Reductions

This system has much higher power consumption than an ideal system, yet it is possible to decrease the power consumption of the system. The ADC and DAC lines of the Arduino Due can be used in lieu of the function of the SMU. This can be achieved by programming a sweep within the Arduino coding to output through the DAC pin acting as a normal sweeping mechanism. The ADC is connected to the output of the array, which is also connected to the ground plane by a resistor. The ADC is then able to measure the voltage drop across the resistor with respect to the ground plane, hence allowing the current to be deduced as the voltage is proportional to the current. By tweaking the resistance, the traces of data can be shifted to allow more peaks to be measured in the 3.3V range of the Arduino Due.

Further reductions can be accomplished by removing the need for either the Raspberry Pi or the Arduino by combining the duties of one into the other. The Arduino can already handle the entirety of the sweeping and measuring mechanics of the system, so saving the data to an SD card, via Bluetooth to a laptop or mobile or outputting a peak position to a small screen would not be a huge step. Guides to allow this storing/transmission of data can be found on the internet very easily and has been shown to be possible before [24-26]. The Raspberry Pi, on the other hand, can use the GPIO pins as the digital I/O pins to activate the switches. With the use of a digital-to-analogue/analogue-to-digital converter (DAC/ADC) [27] breakout board for a Raspberry Pi, it is possible to replicate the DAC/ADC setup of the Arduino Due.

### 3.3.5 Experimental Values

The values of current are taken using a source-delay measurement (SDM) method which is summarised as follows. Once the source is turned on, a period of latency is given for the output to make the transition from off to on. The trigger latency is the time taken to set the voltage to the source value and is fixed at 100 µs. If the output stays on, the latency is not repeated in subsequent cycles.

A delay between the outputs to settle is then implemented to allow the source enough time to settle to the required value. The delay can range from 0 – 9999 s because often circuitry with high capacitance often requires longer priods of time to settle. Due to the complex nature of the RTDs, lower bound for the delay time is given by looking at the capacitance of the device and assuming a 99% of this for the circuit to stabilise. Capacitance can be approximated via a parallel plate capacitor, $C = \varepsilon_0 \varepsilon A/d$. A is defined as the area of the device, 36µm$^2$, ε as the dielectric constant of the material, 13.9, and d as the distance between the plates, estimated at 46.9nm. This results at 9,45 fF which used within $V = V_0(1-e^{t/RC})$, with a 99% reduction, gives a value of 5RC. The delay time is therefore very small for the maximum resistance, given from the plateau region in the NDR at 50kΩ. Hence with a negligible delay time required, 0.1ms is used for the

measurements herein conducted. The value used for all measurement is taken at 0.02s.

Due to the nature of the resonant tunnelling diode, and the instability of the negative differential region, the system needs to settle before a measurement can be taken. The measurement time is given by the NPLC, power line cycle, which is related to the measurement speed. The value is given the time in which a measurement is taken and averaged. The range is given as 0.01 to 10, which in turn corresponds to a measurement time of 0.01/f to 10/f, where f is 50Hz (power line frequency). This can be calculated to have a measurement time of between 0.2 – 2 ms.

The NDR, as shown in 3.4.1, begins to disappear due to added resistance and conduction mechanisms become very stable. The NPLC value can be reduced for measurements in the system requiring only peak position values. An NPLC value of 0.1 is taken corresponding to 2ms measurement time.

The reductions in time are necessary to reduce the time taken for each permutation such that the fastest overall measurement time is achieved with such a large set of responses. With one of the purposes of the research to create a large set of responses which cannot be measured in a reasonable time, the time in which all measurements can be taken should be explored.

Finally, a small amount of time between each permutation measurement is taken to allow the switches time to settle and turn on each of the required devices. Given to be 1s, thus time allows all switches to be turned off and each switch required to be switched back on.

For an array with 256 permutations, a full characterisation of the array between 0 – 4.0V with 100 points every 0.5V takes ~ 1 hour 20 mins. With each permutation taking 18.6 seconds. While this may seem like a reasonable time, a 5 x 5, 25-device array would take ~16 hours. As seen before, using a rectangular array to tune the number of

responses, this time to measure increases beyond control. Each permutation was run 200 times to get the best average peak position and large enough data sets for statistical measurements.

## 3.4 Deconvolution of Characterisation

### 3.4.1 In-Series Resistance

By placing a device in-series, with an added resistance, will cause the peak position to shift towards higher voltages. However, the current at the peak can be seen to be equal for all resistances. The curve can be approximately de-convoluted into added resistance of RTDs in series and the original RTD spectrum using the following relation:

$$\frac{{dI_{rtd}}/{dV} - {dI_{total}}/{dV}}{{dI_{rtd}}/{dV}\,{dI_{total}}/{dV}} = R \tag{11}$$

Where $dI_{rtd}$ and $dI_{total}$ are the change in current outputs for the original RTD and the total curve for an RTD in series with a resistor, R. dV is denoted as the change in voltage by which the peak occurs from zero. This equation can be rearranged to extract peak position for a single RTD or for an in-series resistance RTD.

Resistance shifts for a single device are shown in Figure 20. Interestingly, the increase in resistance removes the plateaus from the NDR region. Additionally, while the peak current remains invariant, the valley increases. This is due to the in-series resistance in the complete circuit.

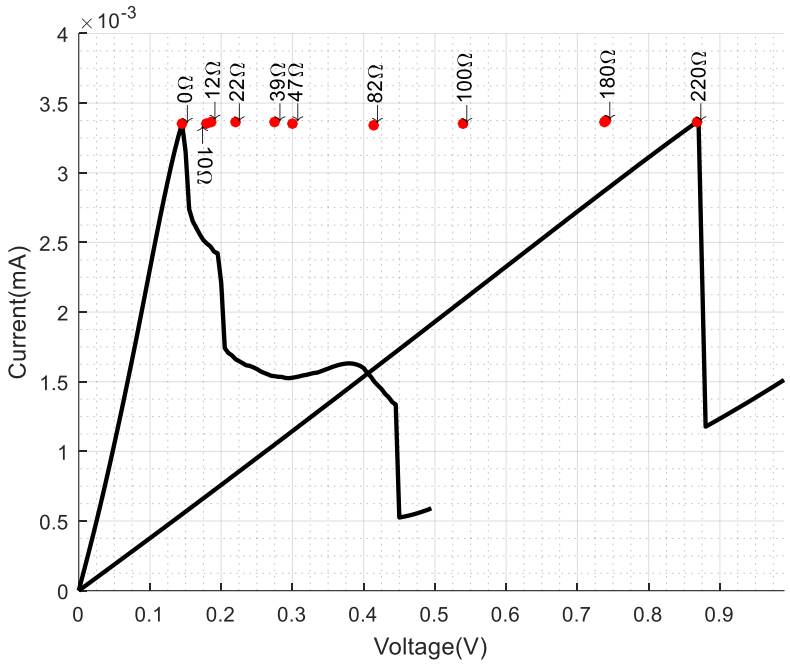


*Figure 20: Shift in peak for varying resistances, starting from left. 0, 10, 12, 22, 39, 47, 82, 100, 180, 220. Average peak positions for 50 loops for each resistance measurement.*

### 3.4.2 In-Series Devices

Each permutation through an array system will give a signature which is influenced by many factors, the first of which being the devices own signatures, which provide the increased number of peaks when in series and the increased size of a peak when in parallel. As each device is unique, each peak can be traced back to its base RTD but with other influences on the device, it has a new trace from the interactions between the devices. The devices peaks can be traced based upon the invariance of the peak current such that the peak with the lowest current will occur first and the largest current last.

The second contributor to the shift in a peak is the resistance at different sections on the trace of the individual devices. The first peaks shift is affected by the resistance of the tunnelling region of the devices which occurs after it in a series measurement. Whereas, the second peak is affected by the tunnelling region of every device after it and the thermionic region of the first peak. This will continue until the final peak is only affected by the thermionic region of every device before it. This is shown in Figure 21.

Testing to see if the permutations are indeed unique compared to every other permutation and to each of its constituent devices, each permutation will be de-convoluted down, removing the added resistance from the devices in parallel and series with it. This will distil the data down such that each peak can be evaluated against its base RTD.

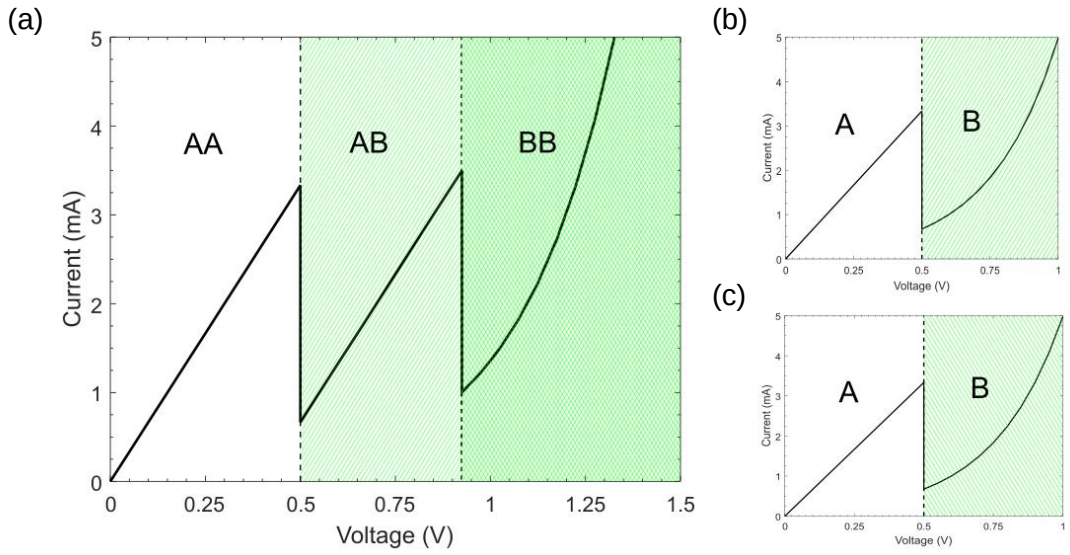


*Figure 21: (a) Current-Voltage spectra of an average two-series system of devices. The different regions are split to show the conduction mechanisms most prevalent and thereby the resistance acting upon the system at any given moment. A and B refer to the conduction mechanism and their resulting resistance which are acting upon the device from (b) and (c). (b/c) Example Current-Voltage spectra of the average single device which make up the two-series system in (a) divided up by their conduction mechanisms and their resulting resistances.*

Using the previous section, 3.4.1**,** we can see how an added resistance affects the devices and similarly, we can use this to extract the added resistance from the device in the case of multiple devices. Using equation (12), we can use resistances from the constituent devices mechanisms to calculate the approximate position of each peak in a combined device. Such that for each mechanism denoted by 1 and 2, which is acting upon the spectrum the combination can be approximated by:

$$\mathrm{dI_{Com}}/_{\mathrm{dV}} = \frac{\mathrm{dI_1}/_{\mathrm{dV}} + \mathrm{dI_2}/_{\mathrm{dV}}}{\mathrm{dI_1}/_{\mathrm{dV}}\,\mathrm{dI_2}/_{\mathrm{dV}}} \tag{12}$$

Where $\mathrm{I_{Com}}$ is the combined mechanism in the spectrum of the combined device and $\mathrm{I_1}$ and $\mathrm{I_2}$ are the individual mechanisms acting together resulting in the individual conduction mechanisms for the combined device. The derivatives are taken only from the valley current of the previous device to switch for the next conduction mechanisms average resistance as the current stays constant through the entire circuit.

### 3.4.3 Combining In-Series Devices

Initially, to understand how the devices interact and give unique responses, the combination of devices should be explored. While a set of parallel devices cause a superposition (and cause the devices to increase to a larger current), series devices keep a lower current with multiple peaks at increasing voltage. Peaks are then caused to shift from their independent positions by the added mechanisms as described in 3.4.2. To be able to de-convolute this data such that the uniqueness of this interaction can be seen would provide useful insight into the mechanisms of the structure by which a strong array PUF interacts. Ideally, the term by which the peak shifts by should be statistically distributed about a central value. It can be observed through combining devices in series and noting the shift from the single device to the devices in series.

For the purposes of consistency, each device used will be of the same size, 36μm$^2$. By using devices which are as similar as possible, it ensures that no extra information can be gained out the system such as some devices having a lower peak than others or outputting a shift which will occur in a specific area. Furthermore, as devices which have different sizes have peaks that cluster in separate locations, it is advantageous to have all peaks occur within the same area so that the overlap of the closest devices can be observed.

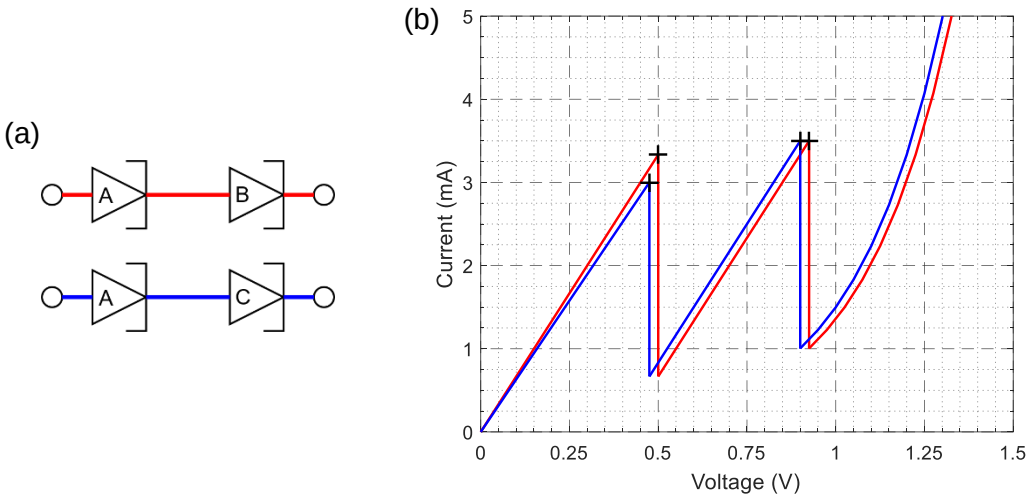


*Figure 22 (a) Example of 2 devices in series, AB and AC (Red/Blue respectively) (b) Current/Voltage trace of 2 devices in series AB and AC, Red and Blue respectively. It can be seen here that B and C occur as the first peak in their respective traces, while A occurs second in both traces, yet the peak of A occurs at a slightly different point in each trace.*

As previously discussed, devices in series can be distinguished between which device is causing the peak to occur. It can be known which device's peak is being shifted by that of the combination and which order they will occur in. It should be noted that this does not break the system of uniqueness as the shift voltage is unique and each device is unique, the combination of which generates our unique responses. Therefore, knowing which devices will occur in which position shouldn't allow an attacker to decipher where any peak in an array will occur in the voltage range.

To categorise the shift of the peak, we can define the value of the difference between the original peak and the peak shifted by the combination of two devices. $\delta_A$ defines the shift of the peak in terms of the voltage shift at which point the peak occurs. (Figure 23)

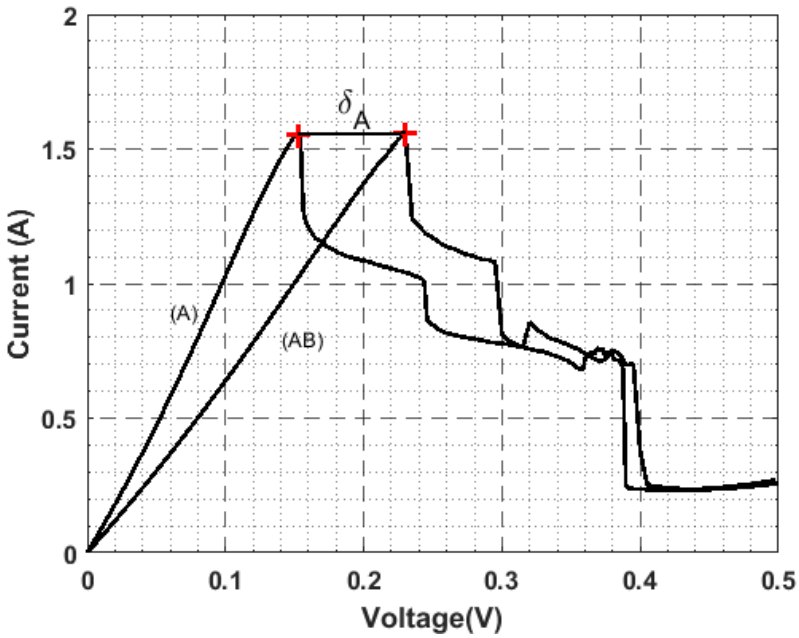


*Figure 23: (A) I/V curve of a single devices, (AB) I/V curve of two devices, A and B in a series combination. $\delta_A$ categorises the difference in the peak between the single devices and a combination of devices.*

As each device will be of similar size, all device can be combined with one another and the shift in the peak should remain centred on the same point. This would allow it to be possible to combine the peak shift of all the different devices starting at the smallest until the penultimate device is only shifted by a single device.

After the data is de-convoluted and the added resistance is accounted for, the data can be collated such that the range of the peak shift can be observed plotting the shift of the peaks. When plotted in a histogram the shift of the peak should have a normal distribution about the average shift of the peak.

# 3.5 PUF Categorization

## 3.5.1 Bit Output

Due to the way in which voltage is input into the system and current is measured, voltage is already divided into discrete values however current is a continuous set of data points with no discernible discretisation of the scale. This is due to the voltage sweep being a staircase sweep where the voltage is held at a specific value for the measurement to be taken, but the measurement of current taken is the highest degree of accuracy allowed. Hence to determine a bit output, some further discretisation of the system is required, mainly upon the current measurement.

A discretisation can be performed by looking at the spread of the data and initially choosing bins with a uniform width such that if a data value falls within that bin, the data value is superseded by the median point of the bin itself. This would allow a pseudo-bit output of the system without the output being reduced to a binary system used specifically for the characterisation of the robustness and uniqueness of the system.

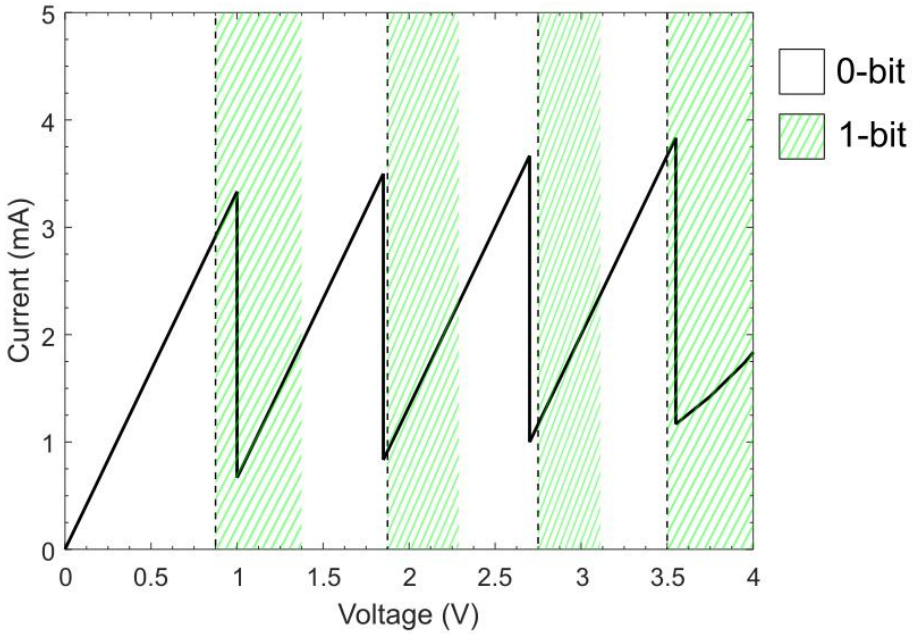


*Figure 24: Graphical representation of an average point bit output for 4 in series devices. The dotted line denotes the average voltage for each peak, and the 0 or 1 bit is decided upon which side of the average the peak appears on.*

To reduce the data down to a binary bit output a few approaches can be taken, however, these methods of reducing the output into bit format are purely speculation and are based on a theory which is untested. Hence, the algorithms described here may not be used in the finished product and are the subject of future work on the project. The first method uses an average point for each peak position calculated before the measurement of the peak to define a divide between a 0 and 1 bin. Upon measurement, it is compared to the average point whereby a negative distance from the average point denotes a zero and a positive distance denotes a 1 (Figure 24). This is repeated for each peak in the measurement and hence the bit output combined is equal to the number of peaks and therefore the number of peaks in series.

The next possible solution to conversion into a bit output uses the spread of the data points to give a uniform output into each bin such that each bin becomes equally probable. Using a uniform binary bin system, the attacker cannot gain further information to allow a brute-force attack to be easier. The bit output is therefore variable depending on how many bins are chosen over the 2D Gaussian shown in Figure 25 (a).

The final solution to be postulated here uses the two previously denoted solutions but by using a different bit output described as the difference between the original peak and the peak when combined with different devices in series (Figure 25(b)). This shift can then be categorised in the same way the peak position is categorised. This would be by either using the average point to denote a divide between a two-bin system or using a binary bin system over the range of the peak-shift range.

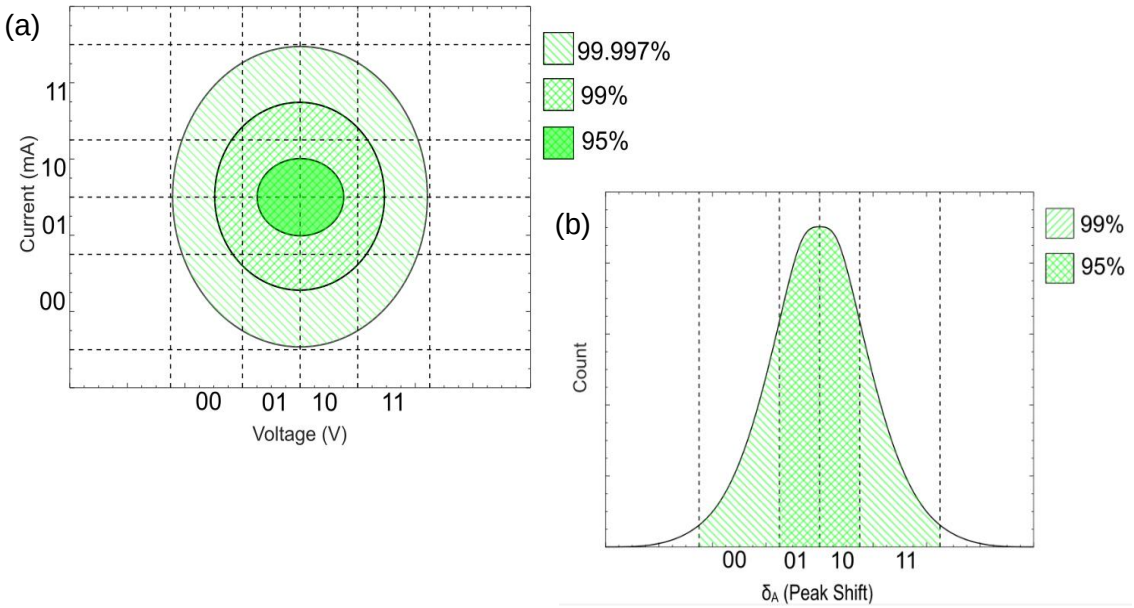


*Figure 25: (a) 2D Gaussian Approximation to an ideal spread of data points across the Current voltage plane. Each axis split into 4, with 16 boxes covering the plane where points can be seen. Each point outputs a 4-bit signature. This can be increased/decreased based upon the spread of data. (b) 1D Gaussian approximation to spread of voltage shift of peak compared to original device. Axis split into 4, so each device outputs 2 bits, but can be varied based on data spread.*

## 3.5.2 Characterization

To characterise each device as a PUF without directly calculating a bit output of the system, each peak position from each spectrum should be categorised by a region of the plane in which it occurs. The plane in which the peaks occur is divided up into uniform-width bins such that permutations can be compared. A method for this binning process is that any value falling within the bin is superseded by the median value of the bin. Hence, this allows the permutations peaks to be described by the bin in which it falls into.

The measurements here give a description of the entire system's properties. These are denoted by robustness and uniqueness. Robustness is defined here as the similarity between repeat measurements of a single permutation of the system. Uniqueness is similarly defined as a measurement of how distinct the permutations of a system are and there by the inversion is the measurement of how likely two distinct challenges are to give the same response.

Robustness and uniqueness can be calculated in a similar way and is based around what data points are compared. Upon comparing two data points a zero or one response is given based on if the data points are equal. The range of responses is binned by axis to create a way to categorise all peaks by where it occurs. A peak which lies within a bin is superseded by the median value of the bin such that all peaks which occur within a bin are equal. The value is compared to other measurements using equation 13 to calculate how distinct each permutation or subsequent measurement is. The distinction of permutations is shown by (13)

$$diff(R_i, R_j) = \begin{cases} 0 \; if \; R_i = R_j \\ 1 \; if \; R_i \neq R_j \end{cases}$$

Where $\mathrm{R_i} = (\mathrm{V_i}, \mathrm{I_i})$, denoting the voltage and current position of the $i^{th}$ iteration.

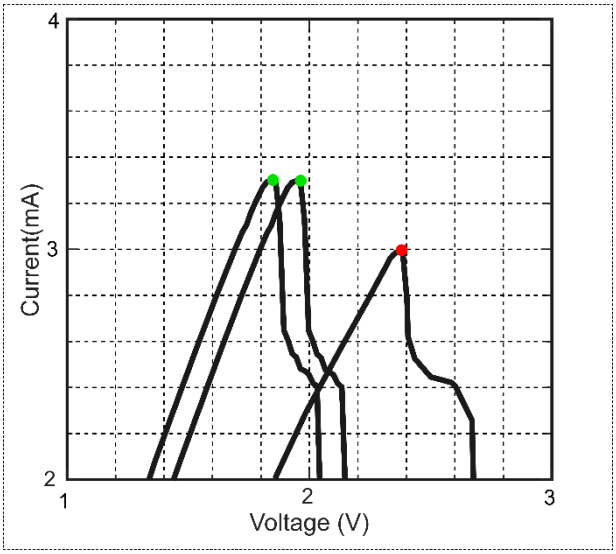


*Figure 26: Binning system for comparing peaks using equation 13 on how distinct each permutation is. Using equation 13, Green points would be equal and red point will be different.*

The robustness of the system can be found for each peak of a permutation individually or more importantly, can be found for the permutation as a whole. Summing $diff(R_i, R_j)$ of subsequent measurements and dividing by the number of measurements taken would give the probability that a subsequent measurement will be different to the expected output. A measure of robustness would therefore be given by the inverse of the described function. This is denoted by the equation:

$$1 - \frac{1}{\mathrm{k}} \sum_{\mathrm{t}=1}^{\mathrm{k}} diff(R_i, R'_{j,t}) \tag{14}$$

Where $R_i$ and $R_{i,t}$ gives the initial output and output at time t, respectively and k gives the number of total measurements. The ideal robustness for the given equation is 1, i.e. each subsequent measurement is guaranteed to have an output equal to that of the expected output and is reliable in its measurement.

The uniqueness is calculated in much the same way, relying on the distinctness of the expected output of a system. The chance that any two random inputs would produce an equivalent output can be found by calculating an average distinction across a set of permutations. A uniqueness measurement is calculated by taking $diff(R_i, R_j)$ of a permutation compared to all subsequent permutations. A sum of $diff(R_i, R_j)$ is multiplied by a coefficient which normalizes by the number of comparisons made. For the following equation:

$$\frac{2}{\mathrm{k}(\mathrm{k}-1)} \sum_{\mathrm{i}=1}^{\mathrm{k}-1} \sum_{\mathrm{j}=\mathrm{i}+1}^{\mathrm{k}} diff(R_i, R_j) \tag{15}$$

An ideal uniqueness for this is 1, where each permutation is distinct and has a unique output from all previous and subsequent measurements. The total number of permutations is given by k. Subsequent permutations are denoted with i and j.

Both uniqueness and robustness are used here to show the effectiveness of a system on providing reliable yet distinct responses from all of its possible permutations. The metrics of robustness and uniqueness correlate directly to the probabilities of false positives and false negatives. False positives are defined as the probability that a random incorrect challenge is accepted i.e when two responses to different challenges are too similar. False negatives are defined as the correct challenge is declined i.e. when subsequent measurements give a different response to the expected output.

As robustness affects the minimum size of bins in the plane, to keep robustness high, means that it also affects the uniqueness of the responses. When bin size decreases, uniqueness increases whereas robustness decreases unless the system outputs a perfectly robust system. Hence, uniqueness and robustness need to be balanced to keep them both as high as possible as these values are connected to the probability of false positive and false negatives.

Uniqueness is difficult to calculate in terms of its relationship to subsequent permutations as it cannot be known if an output is truly unique or just deeply convoluted. For a true uniqueness to be calculated an infinite number of outputs would need to be tested and proven to be difficult or impossible to deconvolute. Devices can be fully tested to be entirely unique from one another, such that one device cannot be found from any of its predecessors, but would require and infinite number of measurements. However, an approximation can be found by using a large subset of the possible outputs and the previous equation. Hence in this research, a range of the different outputs are categorized to simulate an increasing subset which would tend towards showing that the system is truly unique for all permutations possible.

A further analytic measurement is given by Perm, which is a robustness of each permutation. This is defined by using each peak in a permutation in the same $diff(R_i, R_j)$ measurement where one peak being different would constitute a 1 for that permutation measurement.

# Chapter 4 - Results and Discussion

A maximum of 16 similar 36μm$^2$ resonant tunnelling diodes are used in this work to create an exponential array representation. The arrays measured herein are created with the capacity to be characterised in a realistic time frame. Larger systems of arrays would have complete characterisation time which would render the need for protected access to be unnecessary. This property stems from the large set of CRPs and physical limitations present in a system and devices cause a single measurement time to be limited. The motivation being that a system can be challenged as many times in a reasonable time frame but still retain its security.

Responses are manipulated such that analysis can be performed to compare all permutations and peak positions. The peak outputs of each representation are normalised with respect to the point at which the tunnelling region of each constituent device begins. Therefore, using a linear approximation to the tunnelling gradient, extrapolation to zero current gives an approximation to where the peak would start from. Hence it is possible to shift the peak to a normalized position to compare a permutation peak to constituent devices.

## 4.1 Combinations of Two Devices

As previously discussed, the combination of devices can be used outside of an array structure. The method allows investigation of the change in properties and how the unique devices interact. The 16 similar 36μm$^2$ devices are used here to provide 120 unique combinations of devices. Each device is combined with every subsequent device exactly once. Each permutation of devices has 2 combinations, however, both give the same output due to device position being ordered with respect to the peak current of a single device.

Interesting properties necessary in the aforementioned IoT security include the robustness and reliability of a device and the uniqueness of the full set of permutations. Further, the uniqueness and spread of the permutation from its constituent device is rather useful to categorise.

### 4.1.1 Robustness

Robustness is defined as the similarity between multiple readings of a single permutation. Here, as stated in 3.5.2, a difference (diff) between peak position is the measure of similarity between permutations and is divided into Peak 1 and Peak 2. Figure 27 gives a graphical representation of the robustness for each combination of devices. Peak 1 is seen to be much more stable than that of Peak 2, this can most likely be down to the dependence of the Peak 1 affecting Peak 2. This is surmised from that Peak 2 being not robust in similar permutations to peak one with few differences. Hence, Peak 1 being unreliable is likely to cause peak 2 to be unreliable.

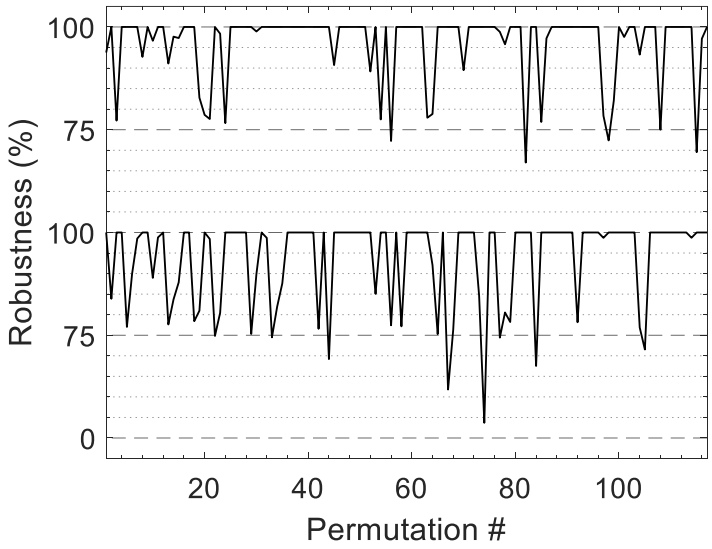


*Figure 27: Robustness for all 2-series combination of 16 devices for 256 bins. (Top) Robustness for peak 1 (Bottom) Robustness for peak 2. Robustness is given by the equation given in 3.5.2*

As can be seen in Figure 27, the combinations show a semi-stable system of devices with a sizeable amount of permutations with 100% robustness. Even with some permutations having a large probability of false negative, the average across each peak are seen to be 92.3% and 92.7%. However, in Table 1, we observe that bin size plays a large role in the robustness measurement, where smaller bins show a decrease in the reliability of the measurements.

| *Bins* | *Identifier* | *Uniqueness (%)* | *Robustness (%)* |
|---|---|---|---|
| *32* | Peak 1 | 97.8 | 92.3 |
| | Peak 2 | 98.6 | 92.7 |
| | Average | 98.1 | 92.5 |
| | **Perm** | **99.6** | **84.0** |
| *16* | Peak 1 | 93.8 | 96.0 |
| | Peak 2 | 94.5 | 93.7 |
| | Average | 94.2 | 94.9 |
| | **Perm** | **98.5** | **87.8** |

*Table 1: Uniqueness and robustness percentages for an average across all peak and peaks 1 and 2 individually for all combinations of 16 RTD devices in series. The measurements are made for multiple variations on the number of boxes per current/voltage output across the range of data.*

### 4.1.2 Uniqueness

Uniqueness, much like robustness, uses a difference measurement to define the similarity or difference between permutations. Table 1 shows the uniqueness between the combinations for individual peaks and total system for 2 distinct bin sizes. The uniqueness of the system for 16 bins is given as 98.5, with a 1.5% chance for two permutations to be the same. However as can be seen, by increasing the bins to 32 over the same range, 99.6% is observed. The uniqueness is dependent on the robustness of the measurements and needs to be balanced. It can be seen that the more robust a system, the more unique and distance each of its permutations can be.

While it cannot be indicated if the system has permutations be truly unique from its constituent device, it is shown that there is no one place where permutations are completely concentrated on. This makes the combination of devices difficult to predict without further information about the system.

It can be assumed that the robustness in these measurements is not as reliable as required, as a 100% robust system would allow for a highly unique and distinct system for a large number of bins over the range of responses. Thus allowing small deviations from permutation to permutation to be visible and make them entirely distinct.

### 4.1.3 Voltage Shift

Voltage shift of a peak due to an in-series device/resistance is described in 3.4.3. The shift is voltage is characterised by a $\delta_A$ term which is a combination of all added resistances acting upon the shifted peak. This shift in peak is the secondary source, besides that of the individual devices, of uniqueness in the system which makes it suitable for security purposes.

Figure 28(a/b) shows the peak voltage shift for each combination from its derivative devices peak. From this, it is obvious that the shift for each device is concentrated around a similar point for all peaks, however, this is due to the devices all being of the same size. Devices being the same size causes a similar resistance hence causing similar shifts due to the added resistances of in-series RTDs. What can be seen is that not all devices are shifted by the same amount indication that the slight differences in resistances are enough to cause combinations of similar devices to give different responses.

Figure 28(c/d) shows the shift with respect to the device which causes the shift of the derivative RTD. It can be seen here that the devices don't cluster perfectly such that each device seems not to shift the derivative RTD by the same amount each time. The shift is more clustered in Peak 1; however, this would be due to the larger variation in resistance in the thermionic region as current increases. The larger distinction in the thermionic region can be explained by the fact that the valley point is subject to the energy level causing the tunnelling region. Thereby, the Thermionic region has some dependence on the tunnelling region and at which point the transport regime switches.

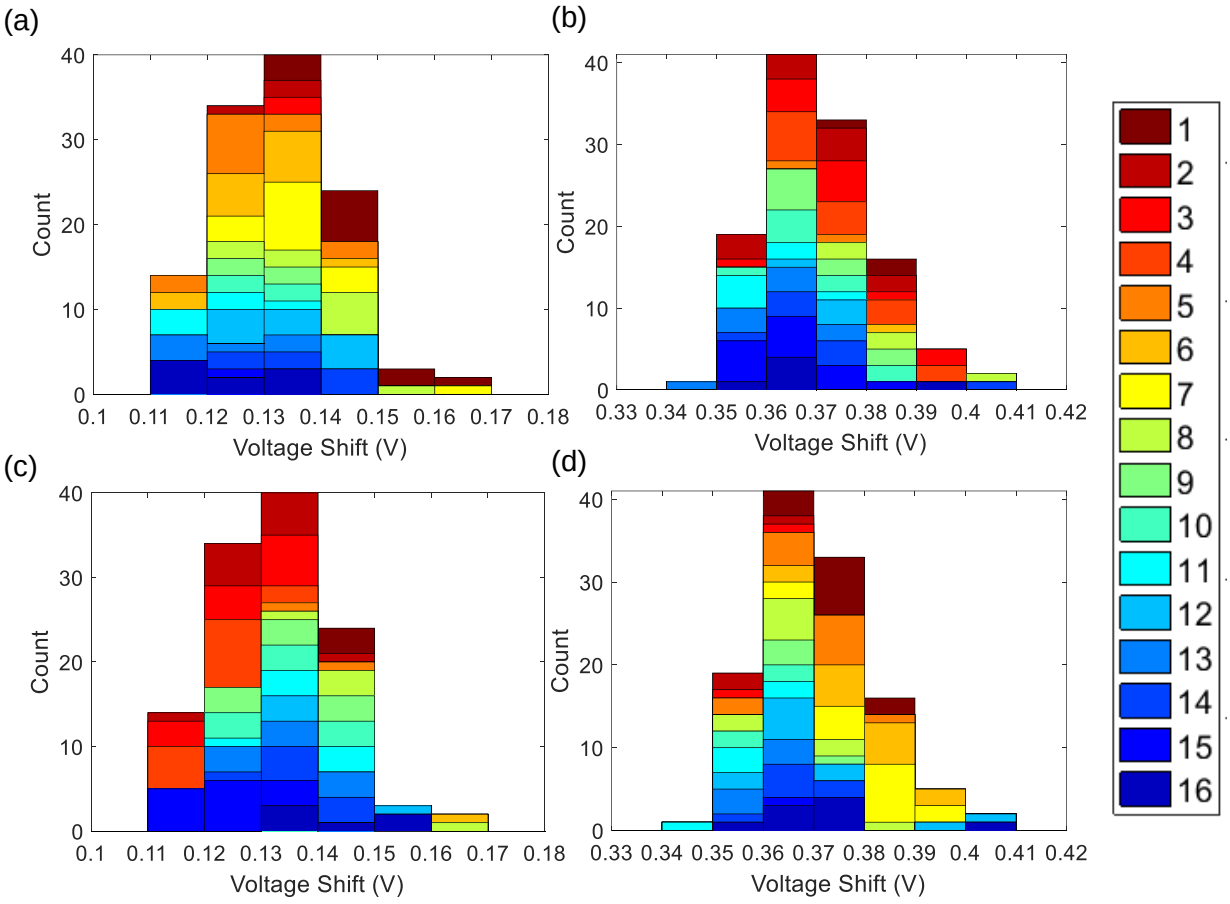


*Figure 28: (a/b) Histogram of voltage shift for of 2-series devices, colour coded by the derivative device from which each peak in question is derived from. (c/d) Histogram of voltage shift where colour coding refers to the device which acts upon the derivative device, i.e. the device which causes the shift in peak. (a/c) Peak 1 (b/d) Peak 2 (Legend) Colour codes for 16 devices in complete combination of 2-series devices. Each peak has 15 shifts acting upon it.*

## 4.2 Dependant-Switch Array

The dependant-switch array is created as defined in 3.1.4 and 3.2.3, whereby its ability to vary in size allows measurements ranging from a 4x4 16-device array down to a 2x2 4-device array. The RTDs inserted into the array are similar 16 devices of size 36 $\mu m^2$ for maximum overlap in the permutation spread.

For each system here, we took a simple binning algorithm such that all bins are of uniform width. This can be changed to a variable bin width giving the occupation of each bin to be uniform giving equal probability to all possible outcomes. The outcome would give an even spread of data points. As we saw previously, that the number of bins doesn't affect the system properties greatly and hence the number of bins per axis is set at the same as the number of devices in the system.

The noteworthy experimental values for the measurement of this system are main influences on the time to characterise. The NPLC of the system is set at 0.1, to give a fast measurement of each point at the cost of accuracy. This gives that a new value can be reported no faster than at 2ms intervals. The delay between measurements is then set at 20ms such that the system has time to reduce capacitance. Finally, 200 loops are performed to give plenty of data to perform a statistical analysis.

### 4.2.1 Robustness

The 4x4 array system, being the largest of the arrays given here, has an output of 256 permutations and 4 peaks. The robustness can be calculated for each permutation divided into individual peaks, an average across each peak for all permutations (denoted as Total) and finally a total robustness for the array (Perm). Each analytical measurement is computed by equation (14). A similar approach is taken for each of the smaller 3x3 and 2x2 array sizes with 27, 3-peak and 4, 2-peak permutations, respectively.

As expected, the stability of each peak is not independent of the whole sweep. This is evident from the shift in robustness to a more stable system for later permutations. This could be due to the system settling during measurement or due to external vibrations reducing across the system. The pins of the system, to allow for easy measurement of the system and prototyping needs have jumper cables, not hard-wired in the system, hence a shift in the connections can cause resistances to shift. Either of these explanations accounts for the changing resistances possibly seen within the system, especially within the early permutations.

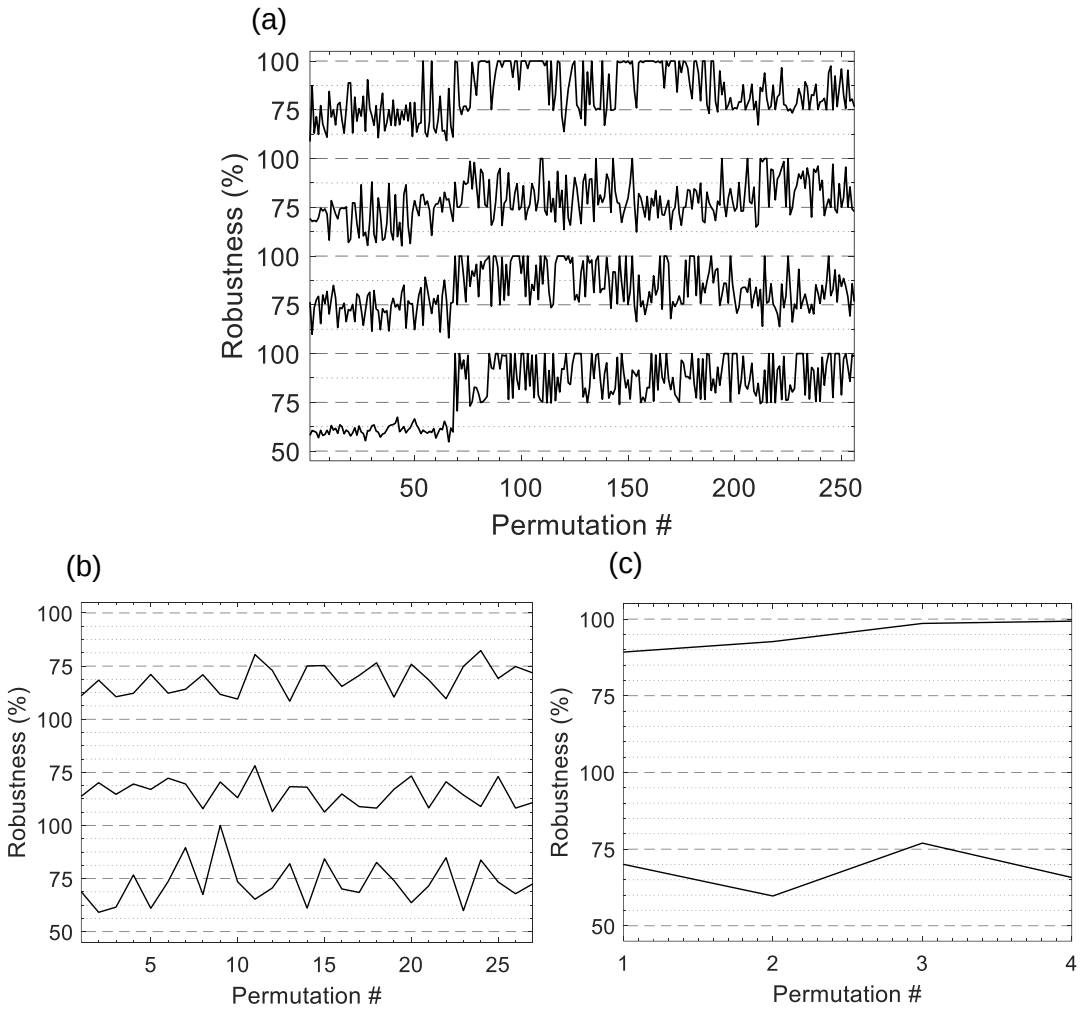


*Figure 29 (a) Percentage robustness measurements over 200 loops for all 256 permutations of a 4x4 dependant-switch array. Each plot is the robustness of a different peak, with the first peak occurring at the top and the fourth peak occurring at the bottom (b) Robustness measurements over 200 loops for all 27 permutations of a 3x3 array. Peak 1 occurs at the top, and Peak 3 occurs at the bottom. (c) Robustness measurement for the 4 permutations of a 2x2 array. Peak 1 occurs at the top of the plot.*

For the 4x4 array, as seen in Figure 29(a), has a robustness above 75% for most peaks after permutation 70. It can be seen that some permutations do have 100% robustness, owing to the notion that a perfectly reliable system is possible. A hard-wired system with less movement due to external influences can negate this and increase overall robustness of the entire set of permutations.

Across all peaks, the average robustness measures to be 85.6% (Table 2). However, as seen in the majority of points, it can be noted that this could easily tend to a higher reliability in the event that early permutations are settled and vibrations through the system's connections are negated. Hence for a dedicated system which required all hardwired elements, a robustness of 90%+ or higher is achievable.

The 3x3 and 2x2 array, have ~70% and ~80% robustness, respectively. While these values are calculated in the same way as previously, the lack of permutations make these values have a high uncertainty. Due to the large number of measurements required for a complete statistical analysis of the system, 27 and 4 permutations are not enough to give an accurate average for the system. This is compounded by the large deviations between peak robustness. Larger arrays of devices would give a more accurate measurement of the system.

Through extrapolation of these small arrays, it can be theorised that larger arrays will have similar robustness. However, for a practical application, a higher reliability may be required. It is likely for increased reliability to be found in a hard-wired, variation-less system and through slower more accurate measurements. Due to potential security flaws, a requirement of the system is the time taken for a characterisation, so fast measurements need to be explored to fully prove the system.

| *Array Size* | *Bins* | *Identifier* | *Uniqueness (%)* | *Robustness (%)* |
|---|---|---|---|---|
| *4x4* | 256 | Peak 1 | 76.8 | 83.9 |
| | | Peak 2 | 95.0 | 78.4 |
| | | Peak 3 | 94.4 | 82.8 |
| | | Peak 4 | 92.3 | 81.0 |
| | | Average | 85.6 | 81.5 |
| | | **Perm** | **99.7** | **60.7** |
| *3x3* | 64 | Peak 1 | 56.1 | 81.2 |
| | | Peak 2 | 87.5 | 66.3 |
| | | Peak 3 | 76.9 | 73.0 |
| | | Average | 73.9 | 73.5 |
| | | **Perm** | **97.4** | **54.4** |
| *2x2* | 16 | Peak 1 | 83.3 | 95.0 |
| | | Peak 2 | 66.7 | 68.1 |
| | | Average | 75.0 | 81.5 |
| | | **Perm** | **83.3** | **66.7** |

*Table 2: Uniqueness and robustness measurements divided into a total measurement and each peak of the system individually. Each set of measurements are detailed for a 4x4, 3x3 and 2x2 array with bins per axis corresponding to the number of devices per array.*

### 4.2.2 Uniqueness

Calculated in the same way to that of the combinations seen in 4.1.2, a 4x4 array has 99.7% uniqueness of the permutations. This gives that 0.3% of comparisons between the 256 permutations are the same. A high uniqueness allows the chance of any random output being successful or providing a false positive when compared to the expected output to be low. However, a requirement for an authentication system is that a single attempt has a $1 \times 10^{8}$ probability to succeed or false acceptance [28].

Again, a 3x3 and 2x2 are scarce on information due to lack of permutations such that uniqueness has large uncertainties. From the given analysis, it can be seen that the uniqueness of the 3x3 and 2x2 permutations is high for a small set of bins. It can be noted from the analysis that with extrapolated larger data array would produce a higher uniqueness, yet at the cost of reliability. Hence the main improvement in the system design and implementation would be increased reliability and accuracy of subsequent measurements.

### 4.2.3 Voltage Shift

The shift in voltage given in Figure 30 shows that the shift is peak shift is centred on a similar point for each individual peak. This is due to each device being of the same size, so resistances and hence the shift due to resistance is similar. The shift becomes more concentrated and detached with subsequent peaks, this is most likely due to the resistance of the thermionic region reducing as current increases. The thermionic region has a lower average resistance compared to the tunnelling region which would cause a less unique shift.

Due to the peak current affecting the order in which devices occur, devices tend to be probabilistic in terms of which position the device will occur in the response. Most notably this can be seen for RTD 7 (Yellow) which occurs only in peak 1 and RTD 5 which occurs only in peak 4.

By looking at a single device it would not be possible to accurately determine where that device will occur unless the output of every subsequent device is also known. A single device will have a probability of occurring in any position based on the range of possible RTD outputs.

Further, the exact output cannot be known unless a complete output of every device is known. As noted in 3.4, for a set of in series devices, without directly measuring each individual device. As each permutation is measured, more becomes known about the system. If the input is known, then the position of each device relative to other devices is revealed. Therefore, as more information about the system is gathered, this probabilistic device occurrence will become more defined as further permutations can be categorised.

An implication of similar devices is that a peak shift for a single device can cluster. This can be seen in RTD 1 (Dark Red) which clustered in the same bin in Peak 2. However, it can be that the shift across devices still shows a spread. Most notably this can be seen in RTD 16 (Dark Blue) by which the device seems to spread relatively uniformly across the range in peaks 2, 3 and 4.

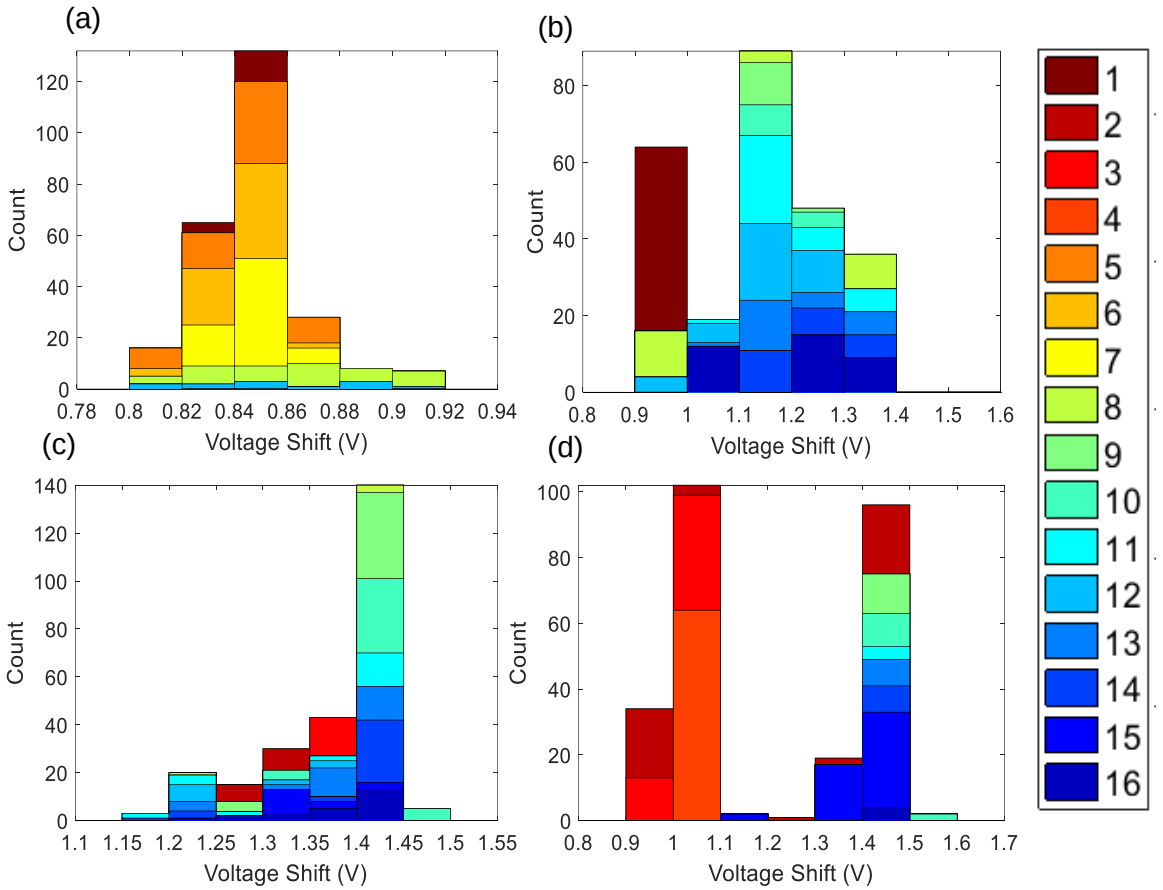


*Figure 30: Histogram of the voltage shift of the peak from its derivative device for (a) Peak 1 (b) Peak 2 (c) Peak 3 (d) Peak 4 of a 4x4 16-device dependant-switch array needs something to show which is the smallest device vs the biggest device. Colour coding shows the derivative peak for which the shift acts upon. Each device has 64 different voltage shifts acting upon it.*

## 4.3 Permutation Spread

Any spread of data is useful if the overlap between values is minimal and points are well defined. However, the spread should have little to no visible trend which would give extra information about the position in which a response would appear. This allows each data point to be distinct without allowing the response of the system to be guessed.

The 4x4 set of data shows a slight positive correlation, with areas which show clustering of points. This is to be expected with similar devices. In light of clustering permutations, a high reliability is required to allow each output to be distinct. Further, each permutation should give out a unique permutation where permutations may look similar but still output a distinct response.

Permutation peak outputs given in Figure 31 are peaks which are normalized such that each peak position can be compared to one another. The permutations peaks occur within the same region of space at varying currents and voltages, which spread across most of the bin-able region

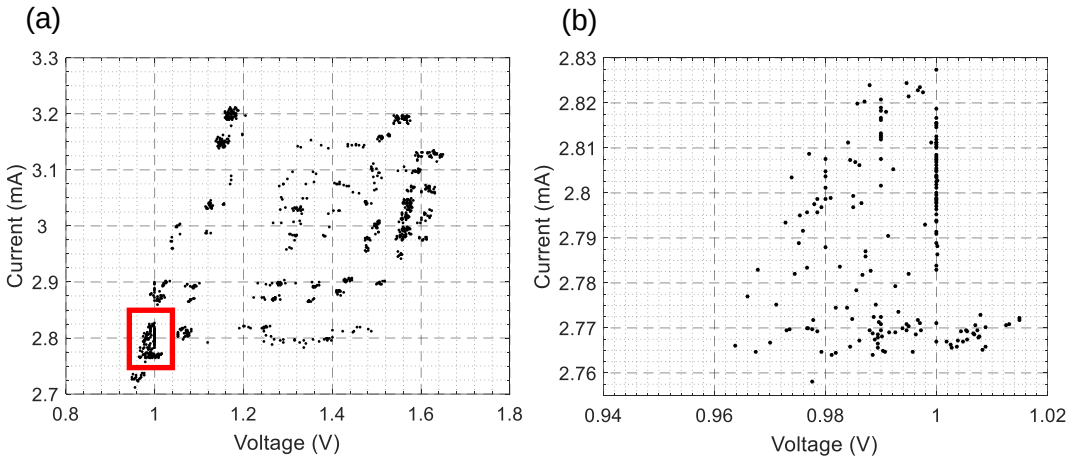


*Figure 31: (a) Peak position of 4 Normalized peak position for 4 peaks of 256 permutations for 4x4 array of 16, 36μm$^2$ RTD devices. (b) Red outlined area of (a) to show distinct points in clustered area.*

# Chapter 5 - Conclusion

In summary, the system of secure authentication proposed here shows merit as a possible strong physical uncloneable function. The design is low cost, easy to produce, integrate into small lightweight systems and contains a large set of authentication responses for a single system. A single 4x4 system has shown the ability to have 256 varying stable and reliable outputs based on 16 unique signatures combined. The responses show unique and distinct responses from other permutations of the system such that a variety of CRPs can be produced.

The combination of devices is shown to be distinct from other permutations through an initial characterisation of 2 in-series devices. The shift in peak is described as a result of the unique resistances from the set of unique devices in a combination. The unique resistance is a by-product of negligible differences in the physical structure of the devices which then cause the device acted upon to shift by an uncontrollable amount.

Further weaknesses come in the form of possibility to create a software-based model of the system once the constituent devices are fully categorised. The way around this issue is to use a fully internal measurement apparatus described in 3.3.4. Achievable by reducing the current system down into a small easy-to-use system with the bare minimum required to measure and control the resonant tunnelling diodes. A further possible development comes in the form of a timed authentication PPUF which used a possible software model as an authentication technique [29]. The representation works on the premise that permutations are difficult to reduce into constituent devices and a software model would take a longer time to respond with a correct answer than a physical system.

For a practical system meeting the requirements of an electronic authentication system [28], the robustness and reliability of subsequent measurement shall be increased, which will directly affect the uniqueness of the design. While further testing on rectangular arrays is necessary to maximise the output, a proof of concept has been provided. The current design serves as a strong PUF with opportunities to further develop the system into a practical application of a unique authentication system.

## 5.1 Future Work

Further research to advance the project of creating a strong PUF from devices employing quantum confinement as a source of unique responses are required. The system described herein needs to be characterised fully on its abilities to return responses. Its ability can be shown in two ways, the first being the size of the CRP array and secondly, the time it takes to return a single response. Further advancements in the size of possible arrays and CRP sets is easily explorable. However, the speed in which a single response can be output becomes a key component in reducing the exposure of the system.

A reduction in the measurement time to the limit of the performance of the system will give an approximation of the minimum time to fully characterise an array. The time is limited in several ways including bottlenecks in data transfer, charge build-up on devices and natural limitation of the software/ hardware controlling a single system.

As discussed previously, the reduction of the system down to a single system is required to make the representation viable for the applications described in Chapter 1, and to successfully prove that the system can be low cost and low power while maintaining security. A bit output needs to be successfully installed to keep the device unique while outputting a uniform spread of bits where no bit string is preferential over any other. Finally, the system needs to be successfully proved to be an authentication method. Whether it be creating a database of challenge-response pairs before use, or using a software model to predict responses at a slower rate than the physical system can respond.

# Bibliography


1. Roberts, J., et al., *Using Quantum Confinement to Uniquely Identify Devices.* Sci Rep, 2015. **5**: p. 16456.
2. Amazon. *Echo Personal Assistant*. 2017; Available from: www.amazon.co.uk/echo.
3. Microsoft, *Cortana Digital Assistant.* 2017.
4. Mycroft. *Open Source Artificial Intelligence*. 2017; Available from: https://mycroft.ai/.
5. Ring. *Ring Smart Doorbel*. 2016; Available from: www.ring.com/.
6. Nest. *Nest Smart Thermostat*. 2017; Available from: https://nest.com/uk/thermostat/install-and-explore/.
7. Phillips, *Hue Bulb.* 2017.
8. Guardian, T. *Major cyber attack disrupts internet service across Europe and US*. 2016.
9. Greenberg, A., *Hackers Remotely Kill a Jeep on the Highway—With Me in It*. 2015: WIRED. p. 5:07.
10. Rührmair, U., S. Devadas, and F. Koushanfar. *Security based on Physical Unclonability and Disorder*. 2011.
11. Skoric, B., et al., *Experimental hardware for coating PUFs and optical PUFs.* Security with Noisy Data, 2007: p. 255-268.
12. Holcomb, D.E., W.P. Burleson, and K. Fu. *Initial SRAM state as a fingerprint and source of true random numbers for RFID tags*. in *Proceedings of the Conference on RFID Security*. 2007.
13. Helfmeier, C., et al. *Cloning physically unclonable functions*. in *Hardware-Oriented Security and Trust (HOST), 2013 IEEE International Symposium on*. 2013. IEEE.
14. Gassend, B., et al. *Silicon physical random functions*. in *Proceedings of the 9th ACM conference on Computer and communications security*. 2002. ACM.
15. Rührmair, U., et al. *Modeling attacks on physical unclonable functions*. in *Proceedings of the 17th ACM conference on Computer and communications security*. 2010. ACM.
16. Kim, T., B. Choi, and D. Kim, *Zero bit error rate ID generation circuit using via formation probability in 0.18 μm CMOS process.* Electronics Letters, 2014. **50**(12): p. 876-877.
17. Brown, E., et al., *Oscillations up to 712 GHz in InAs/AlSb resonant-tunneling diodes.* Applied Physics Letters, 1991. **58**(20): p. 2291-2293.
18. Kanaya, H., et al., *Fundamental Oscillation up to 1.31 THz in Resonant Tunneling Diodes with Thin Well and Barriers.* Applied Physics Express, 2012. **5**(12): p. 124101.
19. Maekawa, T., et al., *Oscillation up to 1.92 THz in resonant tunneling diode by reduced conduction loss.* Applied Physics Express, 2016. **9**(2): p. 024101.
20. Orihashi, N., et al., *Experimental and Theoretical Characteristics of Sub-Terahertz and Terahertz Oscillations of Resonant Tunneling Diodes Integrated with Slot Antennas.* Japanese Journal of Applied Physics, 2005. **44**(11): p. 7809-7815.
21. Chung, S.-Y., et al., *Three-terminal Si-based negative differential resistance circuit element with adjustable peak-to-valley current ratios using a monolithic vertical integration.* Applied physics letters, 2004. **84**(14): p. 2688-2690.
22. Soderstrom, J. and T.G. Andersson, *A multiple-state memory cell based on the resonant tunneling diode.* IEEE electron device letters, 1988. **9**(5): p. 200-202.

23. Gavito, R.B., et al. *N-state random switching based on quantum tunnelling*. in *Nanoengineering: Fabrication, Properties, Optics, and Devices XIV*. 2017. International Society for Optics and Photonics.
24. Arduino. *Read/Write to a SD card with Arduino*. 2015; Available from: https://www.arduino.cc/en/Tutorial/ReadWrite.
25. Girish, M. *Arduino Bluetooth Basic Tutorial*. 2016; Available from: https://create.arduino.cc/projecthub/user206876468/arduino-bluetooth-basic-tutorial-d8b737.
26. Mathavan, H. *Control an Arduino with Bluetooth*. 2015; Available from: https://www.allaboutcircuits.com/projects/control-an-arduino-using-your-phone/.
27. ABElectronics. *ADC/DAC Raspberry Pi Breakout Board*. Available from: https://www.abelectronics.co.uk/p/74/ADC-DAC-Pi-Zero-Raspberry-Pi-ADC-and-DAC-expansion-board.
28. *Security Requirements for Cryptographic Modules (FIPS 140-3)*, N.I.o.S.a. Technology, Editor. 2009.
29. Herder, C., et al., *Physical unclonable functions and applications: A tutorial.* Proceedings of the IEEE, 2014. **102**(8): p. 1126-1141.