



UNIVERSITY OF
CAMBRIDGE

Department of Engineering

Compiling Physical Invariants to Hardware for a Secure & Private Sensor Interface

Author Name: Gregory Brooks

Supervisor: Phillip Stanley-Marbell

Date:

I hereby declare that, except where specifically indicated, the work submitted herin is my own original work.

Signed _____ *date* _____

IIB Project Report:

Compiling Physical Invariants to Hardware for a Secure & Private Sensor Interface

Gregory Brooks, gb510, Christ's College

29 May 2019

Contents

1	Technical Abstract	1
1.1	System Specification	1
1.2	System Diagram	1
2	Introduction and Background	2
2.1	Differential Privacy	2
2.2	The Newton Language	4
2.3	Random Number Generation	5
3	Theoretical Development	7
3.1	Brainstorming Example Applications	7
3.1.1	Digital Camera/Viola-Jones Facial Detection	7
3.1.2	Microphone	8
3.1.3	<i>Intelligent</i> Noising	8
3.2	Multi-Sensor Differential Privacy Loss	9
4	System Design and Architecture	9
4.1	Privacy Budget Management System Architecture	9
4.2	Hardware	11
4.3	Uniform Random Number Generator	12
4.4	Inversion Method Random Number Generator	12

5 Implementation: Technologies and Techniques Used	13
5.1 GitHub	13
5.2 iCE40 FPGA	13
5.3 Software Prototyping	13
5.4 FPGA Synthesis Toolchain	13
5.5 Repository Structure and Build System	13
5.6 Verilog Development Techniques	13
5.7 Hardware Entropy Source	13
5.8 Uniform Random Number Generator Implementation	15
5.9 Inversion Method Random Number Generator Implementation	15
6 Evaluation and Characterisation	16
6.1 Evaluation of Hardware Entropy Source	16
6.2 Scaling of Logic Implementations	16
6.2.1 Uniform Random Number Generator	17
6.2.2 Inversion Method Random Number Generator	17
7 Conclusions	18
7.1 Future Work	18
A Risk Assessment Retrospective	21
B Derivation of an Upper Bound on <i>Indirect</i> Differential Privacy Loss	21
C Hardware Entropy Source Schematic and PCB	26
D EuroSys 2019 Poster: Safeguarding Sensor Device Drivers Using Physical Constraints	28
D.1 Derivation of Bimodal Beta Distribution	29

1 Technical Abstract

Compiling Physical Invariants to Hardware for a Secure & Private Sensor Interface

Gregory Brooks, gb510, Christ's College

⟨ TODO: write this last⟩

This project investigates the idea of using the Newton [1] physics description language as part of a multi-sensor embedded local differential privacy system implemented on an iCE40 FPGA. *⟨ TODO: tidy up the above and expand on it if necessary⟩*

1.1 System Specification

⟨ TODO: move to technical abstract?⟩

1.2 System Diagram

⟨ TODO: insert block diagram of system and describe which parts have been the focus of this project⟩

2 Introduction and Background

2.1 Differential Privacy

Differential privacy can be thought of as a constraint applied to queries for information from a database whereby information about individual members of the database is obscured whilst still releasing useful aggregate information about a population/demographic as a whole. A simple method of implementing this would be the addition of zero mean noise to database entries — whilst the noise would obscure the *true* value of individual data points, the mean value of the dataset would remain intact. More formally, ϵ -*differential privacy* is defined by Dwork and Smith [2] as follows:

“A randomized function K gives ϵ -*differential privacy* if for all data sets x and x' differing on at most one element, and all $S \subseteq Range(K)$,

$$Pr[K(x) \in S] \leq \exp(\epsilon) \times Pr[K(x') \in S], \quad (1)$$

where the probability space in each case is over the coin flips of the mechanism K . ”

Throughout this project, the chosen randomised function K is the Laplace mechanism i.e. the addition of zero mean Laplace distributed random noise to *private* data such as a measurement from a sensor¹ to produce a *noised output* that can be released to untrusted observers (the outside world). This technique provides privacy by ensuring that every possible measurement value (i.e. any value within the sensor measurement range) has a similar posterior probability of being the *true private* value given the observed *noised output*. An outside observer can use the *noised output* to estimate the value of *private* data with some uncertainty but should never be able to deduce it with complete certainty.

The Laplace distribution used by this method has parameter $\frac{\lambda}{\epsilon}$. λ refers to the *global sensitivity* of the database query function; in this scenario this is simply the maximum difference between possible sensor measurements i.e. a sensor’s measurement range. ϵ is a privacy scaling parameter, where a smaller ϵ value results in greater privacy by applying noise with a greater variance, thereby reducing the amount of information each *noised output* value reveals about the *private* data. For a particular application, deciding on an ϵ value is

¹the *databases* in this scenario are simply single sensor measurements hence any two *databases* will always differ by at most a single element.

a tradeoff between utility and privacy since, in the limiting case, an ϵ value of zero means that the system outputs no useful information (just random noise), consequently keeping the *private* data completely private.

Within the context of embedded electronic systems, recent work (for example, this project builds heavily on the work of Choi et al. [3]) has investigated the implementation of differential privacy techniques on low power hardware such as the processors found in smartphones and Internet-of-Things devices. In contrast to the conventional differential privacy model, where a trusted database stores *private* data collected from local hardware, a local differential privacy architecture involves *private* data being masked at the source (e.g. an embedded sensor system) before being sent to an untrusted database [4]. This local approach can result in a more secure system since *private* data does not need to be recorded anywhere nor transmitted from local hardware to a remote server.

One of the key advantages of generating random noise with known parameters is that it allows a quantitative *privacy loss* [3] to be calculated for each piece of information revealed to the outside world (e.g. with the release of each *noised output* value). Privacy losses can be deducted from a total *privacy budget* allocation — once the budget has been depleted, a differential privacy system will not release further information to the outside world. An intuitive justification for such a system comes from the fact that if an attacker were allowed to obtain an unlimited number of noised values, they would be able to take the mean of these responses to gain an arbitrarily accurate estimate for the private data value being masked by zero-mean random noise.

One challenge associated with implementing local differential privacy is that the randomised function K used to mask *private* data can be difficult to implement on low power hardware — one of the findings presented by Choi et al. [3] is that a finite precision implementation of the Laplace mechanism can lead to infinite privacy loss, since certain specific values for *noised output* data can reveal the *true* values for *private* unnoised data with complete certainty. This problem is most apparent on low power hardware utilising low precision fixed point arithmetic. This problem can be overcome by ensuring *noised output* remains within a safe range so that privacy loss can never be infinite; this can be achieved by truncating values if they fall outside the range or by resampling the random noise if it would result in a *noised output* value outside the acceptable range [3].

Existing research regarding embedded systems has focused on single sensor systems [3]. Introducing multiple related sensors introduces complications when quantifying privacy loss, since releasing a noised value of one particular measurement can indirectly reveal information about others. For example, if a particular embedded system contains a GPS and accelerometer, accelerometer measurements can be integrated over time to gain an estimate for GPS position — some privacy loss quantity would have to be deducted from the privacy budget allocated to the GPS as a result of releasing noised acceleration values. One of the aims of this project was to analyse this problem in detail — this work is presented in Section 3.2, with detailed mathematical analysis in Appendix B.

2.2 The Newton Language

Newton is a physics description language developed by Jonathan Lim and Phillip Stanley-Marbell [1]. A Newton description can specify dimensionally annotated *signals* i.e. physical quantities such as measurements from a sensor, and invariants/constraints relating these signals (such as the physical laws governing a particular system). Figures 1 and 2 and Equation 2 illustrate a basic example of the langage being used to describe the physical resonance of a circuit board.

```

1 include "NewtonBaseSignals.nt"
2 board : invariant(Q:dimensionless, d:distance, f:frequency, a:acceleration, t:time) = {
3   a ~ Q*d*(2*kNewtonUnitfree_pi*f)**2*((2*kNewtonUnitfree_pi*f*t)
4     - (2*kNewtonUnitfree_pi*f*t)**3/6 + (2*kNewtonUnitfree_pi*f*t)**5/120)
5 }
```

Figure 1: Example of a Newton description for a physically resonating circuit board (modelled in two dimensions) governed by Equation 2. Figure sourced from a poster produced in parallel to this project [5].

$$a = Q_{bd}d_{bd}\omega^2 \sin(\omega t), \quad (2)$$

where:

- a represents board acceleration,
- Q_{bd} represents quality factor,
- d_{bd} represents board displacement amplitude,

- ω represents angular resonant frequency,
- t represents time.

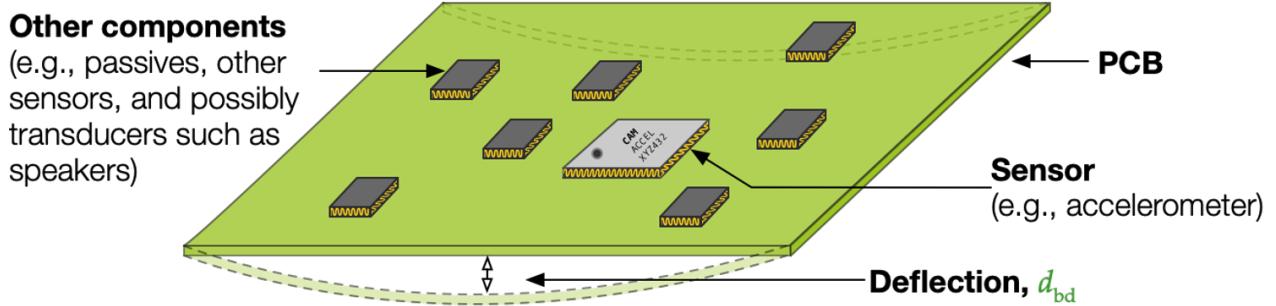


Figure 2: The system modelled by Equation 2. Figure sourced from a poster produced in parallel to this project [5].

At the time of writing the Newton language and compiler is still under development by the Physical Computation Laboratory [6]; the Newton compiler frontend is able to transform a human-readable Newton description into an intermediate representation, a tree structure that can be easily traversed by a computer. The motivation for this project was the idea that this information could be used by a compiler backend to generate a hardware description (e.g. Verilog).

Within the context of this project, Newton provides the ability to describe the relationships between the sensors in a multi-sensor embedded system. This information is necessary to be able to quantify privacy losses in a multi-sensor system.

2.3 Random Number Generation

One important component of a local differential privacy system is a true random number generator (TRNG) to generate the random noise used to obscure *private* data². The Laplace mechanism requires random noise to be Laplace distributed, introducing additional complexity compared to uniform or Gaussian distributed RNGs. The resource constraints imposed by the iCE40 FPGA dictated the need to implement a hardware efficient RNG to generate this Laplace distributed noise. A research article by De Schryver et al. [7] outlines a possible architecture for such a system, based on a lookup table indexed by a floating point

²In a security related application, pseudorandom number generators such as linear feedback shift registers are unsuitable since an attacker would be able to predict their output

representation of the output of a uniform random number generator. This architecture was implemented in Verilog and analysed as part of this project.

3 Theoretical Development

The early stages of the project involved brainstorming and investigating potential research problems that would involve compiling Newton descriptions into a hardware description language. It soon became apparent that the Newton language was well suited to describing the relationships between sensors in a multi-sensor embedded system, thereby providing a basis to build upon the single-sensor state-of-the-art [3] [8].

3.1 Brainstorming Example Applications

Below are some example applications of an iCE40 based differential privacy system that eventually led to the project’s focus on multi-sensor systems.

3.1.1 Digital Camera/Viola-Jones Facial Detection

iCE40 FPGAs are able to drive digital cameras and perform facial recognition, as illustrated by the iCE40 Ultraplus Mobile Development Platform [9]. In this context a differential privacy system might aim to apply random noise to an image in such a way that the identity of an individual can be obscured whilst still preserving enough information for facial detection algorithms to function correctly. A Python script available in this report’s GitHub repository³ demonstrates this idea using OpenCV [10] — the script converts a photo of a face to greyscale and subsamples to 32x32 size (effectively low pass filtering to remove high spatial frequencies) before Laplace distributed random noise ($\lambda = 25$, mean = 0) is added to the pixel intensity values, similarly to the Laplace method described in Section 2.1. The script then uses a Viola-Jones [11] classifier to detect faces in the images before and after the application of noise, demonstrating that a face can be detected in both cases provided the variance of the noise distribution is not too great to drown out all relevant information (analogous to the ϵ value in Section 2.1 being too small). The results of running this script can be observed in Figure 3; the addition of random noise somewhat obscures the identity of the subject in the photo whilst still allowing a Viola-Jones facial detection to function. The demonstration featured here is a simple proof-of-concept; random distortion could be applied to the image in many different ways e.g. random displacement of pixels.

³See Section 5.1.

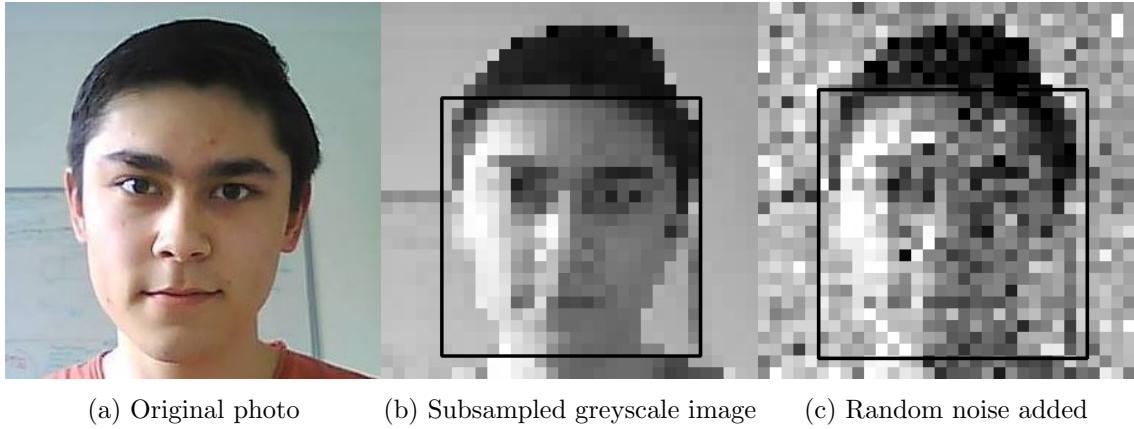


Figure 3: Proof-of-concept demonstration of Laplace distributed random noise being used to obscure an individual’s identity whilst preserving a detectable facial structure. A square has been drawn around the faces detected by the pre-trained Viola-Jones classifier bundled with OpenCV. The photo features, and was taken by the author.

The Viola-Jones classifier works by applying rectangular filters to detect facial features. A Newton description could contain invariants to describe these filters in terms of pixel values; this information could, for example, be used by a noising system to ensure that random noise is applied in a way that does not significantly impact an image’s response to the facial detection filters (i.e. if an image contains a face, then the Viola-Jones classifier described in Newton should still detect the face after the noising process, and vice versa if no face is present in the un-noised *private* image).

3.1.2 Microphone

(TODO)

3.1.3 *Intelligent* Noising

(TODO: allowing the compiler to choose how to add noise, rather than explicitly stating where to add noise) This project investigates a system which allows a user to specify which measurement signals they wish to add random noise to. It is possible to generalise this idea to a hypothetical system which has control over how it applies random noise to measurement data in order to mask a more abstract form of information. One concrete example of this would be when working with images (e.g. the Viola-Jones example above) — given the requirement that individuals’ identities should be obscured whilst still providing enough

information for facial detection algorithms to work reliably, a system might choose to apply noise using, for example:

- Random noise added to pixel intensities.
- Convolution with a random filter.
- Random rearrangement of pixels.
- A combination of multiple techniques.

GIVEN HAAR-LIKE FEATURES USED BY VIOLA JONES, APPLY NOISE IN SUCH A WAY THAT THOSE CONSTRAINTS ARE NOT INVALIDATED

3.2 Multi-Sensor Differential Privacy Loss

4 System Design and Architecture

4.1 Privacy Budget Management System Architecture

In the following section, I propose a privacy management system architecture that could be used to manage multi-sensor privacy on an FPGA. Due to project time constraints, the system has not been completely implemented, since this would divert time and attention away from investigating the more important research questions posed by the project.

(TODO: describe what will have been implemented by the submission date e.g. use privacy yaml file and Newton AST to generate some of the logic implementation by filling in a Verilog template)

The proposed system architecture is as follows:

- The FPGA acts as a sensor interface, allowing external circuitry to request noised sensor values. Noised measurement signals are referred to as *protected*, as their true value has been masked by the addition of zero-mean Laplace distributed random noise to preseve privacy; the system also allows *unprotected* signals to be forwarded to the outside world with no random noise added.
- A *privacy file* (e.g. a YAML file) is used to specify protected measurement signals for a particular embedded system.
- In order to apply random noise to a particular signal, a range of parameters must be defined for each protected signal:

- Range of possible sensor outputs.
- *Privacy budget* value allocated to the sensor.
- *Privacy budget* replenishment rate.
- Value of ϵ , a measure of privacy where a smaller value results in greater privacy by increasing the variance of applied noise [3].

These values can be provided for each protected signal within the *privacy file*.

- The FPGA maintains a differential privacy *budget* for each protected signal. Querying a value for a protected signal (i.e. requesting a noised measurement from the FPGA) causes that signal to incur a privacy loss. The FPGA quantifies this loss (a function of the random noise sample added to the raw measurement [3]) and subtracts it from the signal’s privacy budget. Once the budget is depleted, further queries of the signal in question are ignored — a nonzero budget replenishment rate can be specified in the privacy file so that the signal can eventually be queried again once the budget has replenished sufficiently.
- Invariants in the Newton description define relationships between signals. This is important as privacy is lost if a measurand’s value is calculated by taking measurements from related sensors, without directly measuring the measurand. The system must account for this *indirect* privacy loss:
 - Each Newton invariant containing protected measurement signals has an associated bitfield register on the FPGA. Each bit in the register acts as a read flag for each measurement signal contained in the associated invariant — the flag is set when the sensor is queried (i.e. when a measurement is taken and the noised measurement value becomes *known* to the outside world) and is not cleared until the sensor’s privacy budget is completely replenished some time after a query (at which point the measurement value can be considered *unknown* to the outside world).
 - As long as any two flags remain unset, there are two variables in the equation that are *unknown* to an attacker. The equation cannot be solved and so no information⁴ can be gained about the *unknown* values.

⁴This assumes that the probability distributions of sensor outputs are independent, see Appendix B for an investigation into correlated measurements.

- If only one bit remains unset, then the remaining *unknown* value can be calculated using the other *known* values with set flags — the *unknown* signal incurs a privacy loss. If all flag bits are set then responding to a measurement query results in a privacy loss for every other signal in the invariant. These scenarios can be summarised by saying that, for signals contained within a particular Newton invariant, a signal experiences an indirect privacy loss if any of the other signals in the invariant are queried *and* all other signals are *known* i.e. have been recently queried.
- If responding to a query would exceed the privacy budget for any protected signal (not just the one being queried) then a response should not be granted.
- The mathematics behind the calculation of a quantitative value for privacy loss is investigated in Appendix B; whilst calculation of an exact value for privacy loss (Equation 6 in Appendix B) requires the use of a mathematical optimisation technique such as Lagrange multipliers, a simpler calculation can be used to obtain an upper bound on the value. This simpler calculation can be performed more easily on a small FPGA such as the iCE40, and can even be pre-calculated during compile-time since it is a function of the random noise output only.

4.2 Hardware

As discussed previously, this system is based on the iCE40 FPGA, selected for its small size and low power consumption. These qualities make it ideal for inserting into an embedded device without significantly impacting the size or power consumption/battery life. Using FPGA hardware rather than a software implementation running on a processor has several advantages:

- Increased speed, to the point where data rate is limited by the SPI/I2C bus clock rather than the privacy system.
- Increased energy efficiency.
- Potential for integration into a sensor’s package (due to the FPGAs small size).
- Security - a hardware implementation is difficult for an attacker to manipulate or modify.

4.3 Uniform Random Number Generator

⟨ TODO: insert block diagram ⟩

4.4 Inversion Method Random Number Generator

⟨ TODO: insert block diagram ⟩

5 Implementation: Technologies and Techniques Used

5.1 GitHub

Below is a list of GitHub repositories containing work done as part of this project:

- Main project repository containing Python simulations/tests, Verilog templates and C code to fill in templates:

<https://github.com/Gregox273/Newton-Verilog-Compiler>

- Verilog repository for iCE40 differential I/O based uniform random number generator:
<https://github.com/physical-computation/iCE40-LVDS-RNG>

- Schematic and PCB design for hardware (zener diode based) noise generator:

<https://github.com/Gregox273/zener-rng>

⟨ TODO: describe how GitHub was used ⟩ ⟨ TODO: include an example of commit history for a repository ⟩

5.2 iCE40 FPGA

5.3 Software Prototyping

5.4 FPGA Synthesis Toolchain

5.5 Repository Structure and Build System

5.6 Verilog Development Techniques

⟨ TODO: describe good practices picked up over the course of this project ⟩

5.7 Hardware Entropy Source

As discussed in section 2.3, one of the key components of a security/privacy application is a true random number generator (TRNG). Unlike a pseudorandom number generator (PRNG), which produces a predictable deterministic outcome, a true random number generator’s output cannot be anticipated by an attacker, preserving system security and users’ privacy. The output bit rate of TRNGs is often limited so can be used to seed a cryptographically secure PRNG (CSPRNG) algorithm to increase output bit rate without significantly compromising security. This extra step was not required for this project, since the rate at which random

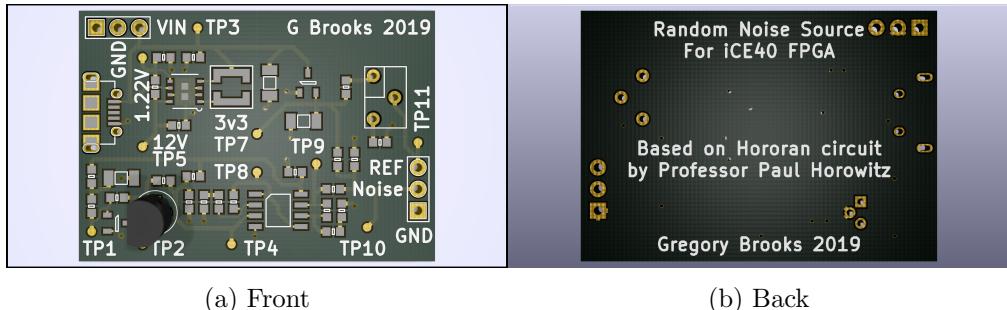
numbers are consumed is relatively low.

Initial investigations focused on using the iCE40 FPGA's differential I/O hardware to generate random bits which could be fed into a TRNG to generate random samples from an arbitrary distribution. Initially, it was hoped that simply leaving two comparator inputs floating would cause the output to fluctuate randomly — in practice this did not occur as the comparator hardware requires one of the inputs to be biased within a narrow common mode voltage range (roughly half the I/O supply voltage, allowing the other input to swing about this reference voltage) [12]. I decided to design a PCB to set this bias voltage, as well as provide an alternative entropy source in the form of a reverse biased zener diode producing avalanche noise. In theory, this circuit would provide a better i.e. less predictable source of entropy since, unlike the floating input pin, the avalanche noise (and therefore random number output) produced is independent of the device's external environment⁵.

The circuit used in this noise generator is based on a design by Professor Paul Horowitz [13, p. 984], modified to operate at 3.3V and 0V power rails (rather than $\pm 5V$). Avalanche noise produced by a reverse biased zener diode is amplified through a dual op-amp amplifier stage before being modulated about a 1.65V DC bias ($\frac{V_{cc}}{2}$); this output can then be fed into one of the iCE40's comparator inputs. A second output on the PCB comes from a simple potentiometer circuit to set the common mode reference voltage on the second comparator input (with series resistors to allow for fine control about $\frac{V_{cc}}{2}$). Since the comparator output depends on whether the instantaneous voltage of the noise waveform lies above or below this reference voltage, this reference point can be adjusted to set the ratio of ones to zeros in the random comparator output.

(TODO: finish e.g. talk about theoretical circuit specs (e.g. noise bandwidth))

⁵A floating input pin can easily be influenced by its external environment e.g. picking up 50Hz noise from nearby electrical mains, resulting in a non-uniform frequency spectrum for random output.



(a) Front

(b) Back

Figure 4: 3D renders of PCB (in KiCad).

< TODO: finish>

The full circuit schematic and PCB layout can be found in Appendix C, as well as the GitHub repository listed in Section 5.1.

5.8 Uniform Random Number Generator Implementation

< TODO: insert block diagram>

5.9 Inversion Method Random Number Generator Implementation

< TODO: insert block diagram>

6 Evaluation and Characterisation

6.1 Evaluation of Hardware Entropy Source

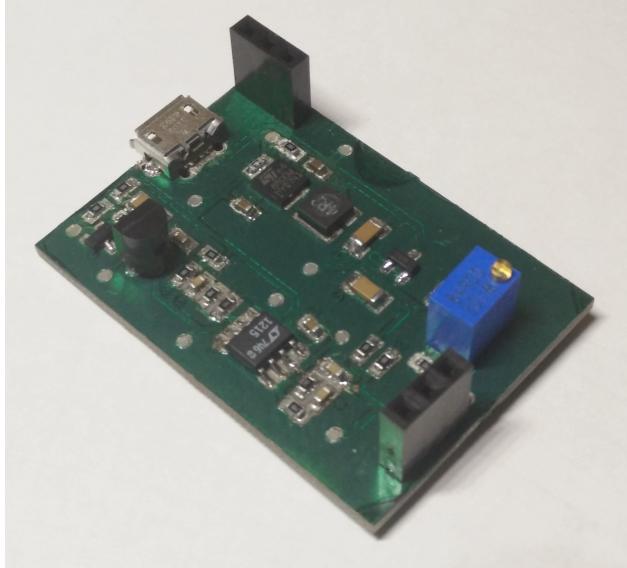


Figure 5: The physical PCB, assembled by the CUED Dyson Centre’s Electronics Development Group.

⟨ *TODO: finish e.g. talk about measured circuit specs (e.g. noise bandwidth)* ⟩

⟨ *TODO: include all measurements (put large figures in appendix)* ⟩

One problem identified during testing was the tendency for the 12V boost regulator’s inductor to get very hot during operation; whilst the circuit still functions in a room temperature environment, a further prototype iteration would use an inductor with a higher current rating to improve reliability. This might require a redesign of the PCB to accommodate a larger inductor package and footprint, although similarly sized alternatives are available⁶.

⟨ *TODO: quantify randomness of output, with and without zener diode* ⟩

6.2 Scaling of Logic Implementations

⟨ *TODO: evaluate e.g. number of LUTs etc.* ⟩

⟨ *TODO: investigate how URNG and RNG logic implementations scale with various para-*

⁶e.g. Coilcraft XAL4030-472MEC with a 5.1A RMS current rating, compared to the 1.2A for the Wurth Elektronik 744031004 currently being used. The datasheet for the ST8R00 step-up regulator states an absolute maximum inductor current of 3A [14].

meters (and how this compares to a naive implementation)⟩

⟨ TODO: investigate scaling of adder hardware (inferred from Verilog vs iCE40 hardware accumulators)⟩

6.2.1 Uniform Random Number Generator

6.2.2 Inversion Method Random Number Generator

7 Conclusions

7.1 Future Work

The most obvious path for future work is to continue with the implementation of the system outlined above, in order to reduce a working prototype. As a research project, the characterisation work described in Section 6 was prioritised over working towards a working prototype, for the purposes of inspiring future research.

One potential area for future research is the application of the Newton language to safeguarding against *transduction attacks* (attacks on a computer system performed by manipulating transducers i.e. sensors in the system). This idea was briefly investigated by the author in parallel to this project, resulting in the publication of a poster [5] presented by the author at EuroSys 2019 in Dresden. The overall idea behind the poster was that Newton could be used to describe plausible system behaviour (e.g. constraints imposed by physical laws governing the system), thereby allowing implausible behaviour to be recognised and interpreted as a possible transduction attack in progress. The poster is summarised in Appendix D.

References

- [1] Jonathan Lim and Phillip Stanley-Marbell. “Newton: A Language for Describing Physics”. In: *CoRR* abs/1811.04626 (2018). arXiv: 1811.04626. URL: <http://arxiv.org/abs/1811.04626>.
- [2] Cynthia Dwork and Adam Smith. “Differential privacy for statistics: What we know and what we want to learn”. In: *Journal of Privacy and Confidentiality* 1.2 (2010).
- [3] Woo-Seok Choi et al. “Guaranteeing Local Differential Privacy on Ultra-Low-Power Systems”. In: *2018 ACM/IEEE 45th Annual International Symposium on Computer Architecture (ISCA)* (2018), pp. 561–574.
- [4] Peter Kairouz, Sewoong Oh and Pramod Viswanath. “Extremal Mechanisms for Local Differential Privacy”. In: *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2*. NIPS’14. Montreal, Canada: MIT Press, 2014, pp. 2879–2887. URL: <http://dl.acm.org/citation.cfm?id=2969033.2969148>.
- [5] Gregory Brooks, Youchao Wang and Phillip Stanley-Marbell. *Safeguarding Sensor Device Drivers Using Physical Constraints*. Poster presented at EuroSys 2019, Dresden, Germany.
- [6] *Physical Computation Laboratory*. <http://physcomp.eng.cam.ac.uk/>. Accessed 2019-05-18.
- [7] Christian De Schryver et al. “A Hardware Efficient Random Number Generator for Nonuniform Distributions with Arbitrary Precision”. In: *Int. J. Reconfig. Comput.* 2012 (Jan. 2012), 12:12–12:12. ISSN: 1687-7195. DOI: 10.1155/2012/675130. URL: <http://dx.doi.org/10.1155/2012/675130>.
- [8] Cynthia Dwork et al. “Calibrating Noise to Sensitivity in Private Data Analysis”. In: *Theory of Cryptography*. Ed. by Shai Halevi and Tal Rabin. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006, pp. 265–284. ISBN: 978-3-540-32732-5.
- [9] *iCE40 Ultraplus MDP*. <http://www.latticesemi.com/Products/DevelopmentBoardsAndKits/iCE40UltraPlusMobileDevPlatform>. Accessed 2019-05-18.
- [10] *opencv-contrib-python*. <https://pypi.org/project/opencv-contrib-python/>. Accessed 2019-05-19.
- [11] Paul Viola and Michael Jones. *Rapid object detection using a boosted cascade of simple features*. 2001.

- [12] Lattice Semiconductor. *Common Analog Functions Using an iCE40 FPGA*. Tech. rep. 2012.
- [13] Paul Horowitz and Winfield Hill. *The Art of Electronics*. 3rd ed. Cambridge University Press, 2015. ISBN: 978-0-521-80926-9.
- [14] *Micropower 1 A synchronous step-up DC-DC converter*. ST8R00. Doc ID 13449 Rev 2. ST Microelectronics. Feb. 2012.
- [15] Kevin Fu and Wenyuan Xu. “Risks of Trusting the Physics of Sensors”. In: *Commun. ACM* 61.2 (Jan. 2018), pp. 20–23. ISSN: 0001-0782. DOI: 10.1145/3176402. URL: <http://doi.acm.org/10.1145/3176402>.
- [16] J. Liu, C. Yan and W. Xu. *Can you trust autonomous vehicles: Contactless attacks against sensors of selfdriving vehicles*. In *DEFCON24* (Aug. 2016). URL: <http://bit.ly/2EQNOLs> (visited on 11/03/2019).
- [17] Analog Devices. *ANALOG DEVICES ADVISORY TO ICS ALERT-17-073-01*. Tech. rep. 14th Mar. 2017.
- [18] Gregory Brooks. *issue-25, Physical Computation Measurement Data Repository*. URL: <https://github.com/physical-computation/measurement-data/tree/issue-25> (visited on 11/03/2019).

A Risk Assessment Retrospective

The risk assessment submitted at the start of the project does not mention any specific hazards besides office (computer) work, since the project is predominately software/firmware based (all project hardware operated at low voltages i.e. 12V or less). No other hazards were encountered during the course of the project, since the random noise generator PCBs were manufactured by the Dyson Centre’s Electronics Development Group. In retrospect, although not part of the initial project specification, the risk assessment could have anticipated the possibility of manufacturing PCBs for the project. The hazards associated with this activity include high temperatures (from a soldering iron/oven/hot air gun) as well as chemical hazards associated with solder, fume extraction etc.

B Derivation of an Upper Bound on *Indirect Differential Privacy Loss*

Let \hat{X} be a random variable denoting a sensor measurement, including any measurement error.

Let $N_{Laplace}^\lambda$ be a random variable following a Laplace distribution with parameter λ . This represents random noise added to sensor measurements.

Let X be a random variable representing a *noised* sensor measurement i.e. the masked value that the differential privacy system provides to the outside world after applying Laplace distributed noise to a measurement:

$$X = \hat{X} + N_{Laplace}^\lambda. \quad (3)$$

A measurement event comprises the variable \hat{X} taking on value \hat{x} . Similarly, a noising event can be defined by X taking value $x = \hat{x} + n$. We define an embedded system as having n sensors with measurements distributed as \hat{X}_i , $1 \leq i \leq n$.

Let Y denote a variable derived from one or more *noised measurements*, X , such as measurements from different sensors in an embedded system:

$$\begin{aligned} Y &= f(X_1, X_2, \dots, X_n) = f(\mathbf{X}), \\ y &= f(x_1, x_2, \dots, x_n) = f(\mathbf{x}). \end{aligned}$$

This same mapping function $f(\mathbf{x})$ can take unnoised measurements as arguments:

$$\begin{aligned}\hat{Y} &= f(\hat{X}_1, \hat{X}_2, \dots, \hat{X}_n) = f(\hat{\mathbf{X}}), \\ \hat{y} &= f(\hat{x}_1, \hat{x}_2, \dots, \hat{x}_n) = f(\hat{\mathbf{x}}).\end{aligned}$$

Using this function, an untrusted party can compute an instance value y by taking measurements x_1 to x_n — variable \hat{Y} has experienced some privacy loss, $l_{\hat{Y}}$, since a noised instance value y has effectively been released to the outside world (i.e. untrusted parties), revealing some information about the private noise-free value \hat{y} taken by variable \hat{Y} . The privacy loss can be quantified using a log-likelihood ratio [3] (Equation 5). This privacy loss function requires two parameters, \hat{y}_a and \hat{y}_b , which are the possible *true* i.e. noise-free values for \hat{Y} that maximise the privacy loss (Equation 4).

$y_{obs} = f(\mathbf{x}_{obs})$ = value for y calculated from noised measurements \mathbf{x}_i ,

$$\hat{y}_a, \hat{y}_b = \underset{\hat{y}_a, \hat{y}_b}{\operatorname{argmax}} l_{\hat{Y}}(\hat{y}_a, \hat{y}_b), \quad (4)$$

$$l_{\hat{Y}}(\hat{y}_a, \hat{y}_b) = \log \left(\frac{Pr\{Y = y_{obs} | \hat{Y} = \hat{y}_a\}}{Pr\{Y = y_{obs} | \hat{Y} = \hat{y}_b\}} \right). \quad (5)$$

Equation 5 can be rewritten as follows, where $\hat{\mathbf{x}}_a$ and $\hat{\mathbf{x}}_b$ are chosen to maximise $l_{\hat{Y}}$ as before, but subject to constraints $\hat{\mathbf{x}}_a = f^{-1}(\hat{y}_a)$ and $\hat{\mathbf{x}}_b = f^{-1}(\hat{y}_b)$:

$$l_{\hat{Y}}(\hat{\mathbf{x}}_a, \hat{\mathbf{x}}_b) = \log \left(\frac{Pr\{\mathbf{X} = \mathbf{x}_{obs} | \hat{\mathbf{X}} = \hat{\mathbf{x}}_a\}}{Pr\{\mathbf{X} = \mathbf{x}_{obs} | \hat{\mathbf{X}} = \hat{\mathbf{x}}_b\}} \right). \quad (6)$$

This constrained optimisation problem could be solved to calculate an exact value for privacy loss. Alternatively, an upper bound on privacy loss can be determined using a far simpler calculation (the sum of privacy losses for each individual sensor):

$$\begin{aligned}l_{\hat{Y}}(\hat{\mathbf{x}}_a, \hat{\mathbf{x}}_b) &= \log \left(\prod_{u=1}^n \frac{Pr\{X_u = x_{u,obs} | \hat{X}_u = \hat{x}_{u,a}\}}{Pr\{X_u = x_{u,obs} | \hat{X}_u = \hat{x}_{u,b}\}} \right), \\ &\leq \log \left(\prod_{u=1}^n \frac{Pr\{X_u = x_{u,obs} | \hat{X}_u = \hat{x}_{u,c}\}}{Pr\{X_u = x_{u,obs} | \hat{X}_u = \hat{x}_{u,d}\}} \right),\end{aligned} \quad (7)$$

where:

$$\hat{x}_{u,c} = \underset{\hat{x}_{u,c}}{\operatorname{argmax}} Pr\{X_u = x_{u,obs} | \hat{X}_u = \hat{x}_{u,c}\},$$

$$\hat{x}_{u,d} = \underset{\hat{x}_{u,d}}{\operatorname{argmin}} Pr\{X_u = x_{u,obs} | \hat{X}_u = \hat{x}_{u,d}\},$$

i.e. the constraint $f^{-1}(\hat{y})$ has been removed. Equation 7 can be interpreted as the sum of the privacy losses incurred by the X variables i.e. $\sum_{u=1}^n l_{X_u}$. $Pr\{X_u = x_{u,obs} | \hat{X}_u = \hat{x}_u\}$ is simply the Laplace distribution of the noise applied to the measurement represented by \hat{X} , centred on value \hat{x} .

Note that this privacy loss is incurred when all values $x_{1,obs}$ to $x_{n,obs}$ are provided to the outside world (i.e. the event where an attacker transitions from having no information to the system to being provided with $x_{1,obs}$ to $x_{n,obs}$); for each subsequent individual query response $x_{u,obs}$, the resulting privacy loss is only:

$$l_{\hat{Y}}(\hat{\mathbf{x}}_a, \hat{\mathbf{x}}_b) = \log \left(\frac{Pr\{X_u = x_{u,obs} | \hat{X}_u = \hat{x}_{u,a}\}}{Pr\{X_u = x_{u,obs} | \hat{X}_u = \hat{x}_{u,b}\}} \right), \quad (8)$$

where values for $\hat{\mathbf{x}}_a$ and $\hat{\mathbf{x}}_b$ are the same as those used in Equation 6 (if calculating exact privacy loss) or Equation 7 (if calculating an upper bound).

Since the Newton language recently gained a mutual information operator, I attempted to factor mutual information into this calculation. Unfortunately, it appears that mutual information alone does not provide enough information to calculate indirect privacy loss — the exact nature of the correlation between two (or more) random variables is required i.e. the conditional distribution for the unknown variable given known ones:

Define \hat{X} and X as before and let \hat{Y} be a random variable correlated with \hat{X} . Both \hat{X} and \hat{Y} can take values within some range (e.g. due to finite sensor precision):

$$\hat{X} \in R_{\hat{X}}, \quad \hat{Y} \in R_{\hat{Y}},$$

The mutual information $I(\hat{X}; \hat{Y})$ is defined as:

$$I(\hat{X}; \hat{Y}) = \int_{R_{\hat{Y}}} \int_{R_{\hat{X}}} p(\hat{X}, \hat{Y}) \log \left(\frac{p(\hat{X}, \hat{Y})}{p(\hat{X})p(\hat{Y})} \right) d\hat{X} d\hat{Y}, \quad (9)$$

where $p()$ denotes the probability density function for a random variable. I have been unable to insert this value into the privacy loss equation, however a value for privacy loss can be obtained if the conditional distribution of \hat{X} given \hat{Y} is known, as this allows the conditional distribution of X given \hat{Y} to be calculated:

$$Pr\{X = x | \hat{Y} = \hat{y}\} = \int_{R_{\hat{X}}} Pr\{X = x | \hat{X} = \hat{x}\} Pr\{\hat{X} = \hat{x} | \hat{Y} = \hat{y}\} d\hat{x}. \quad (10)$$

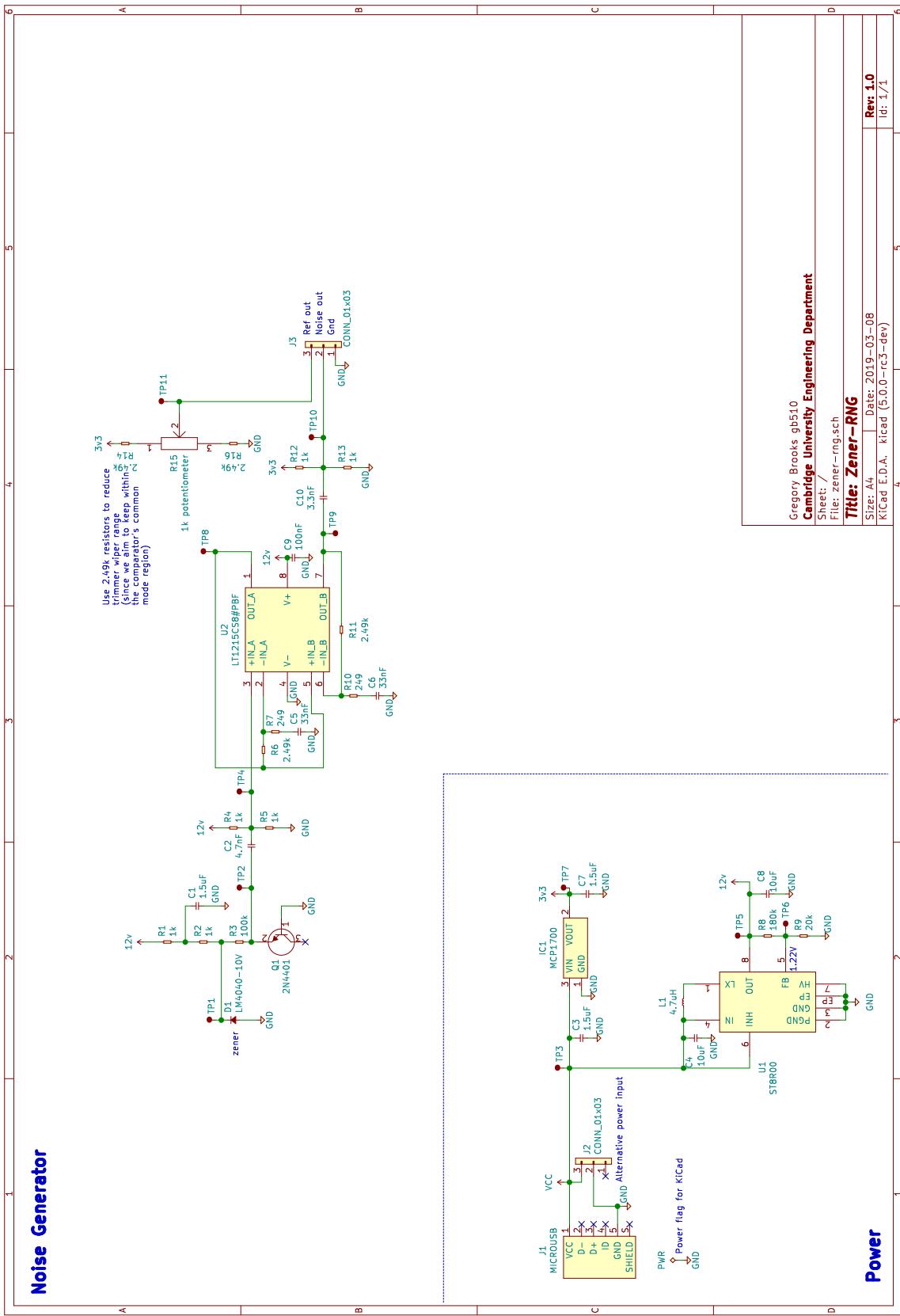
Equation 10 can be substituted into Equation 11, the formula for privacy loss incurred by \hat{Y} as a result of observing a value x_{obs} for X . This results in Equation 12:

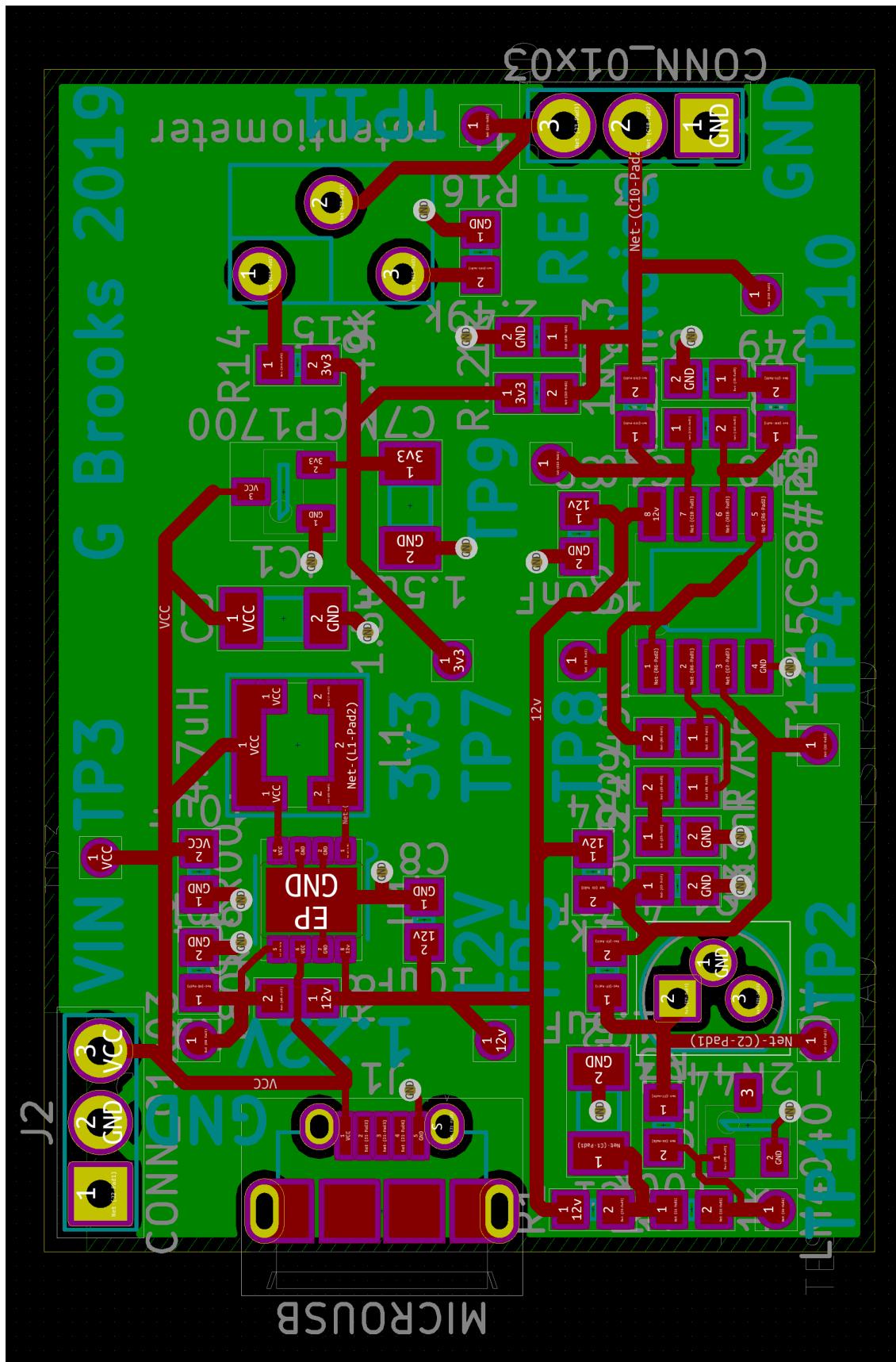
$$l_{\hat{Y}}(\hat{y}_a, \hat{y}_b) = \log \left(\frac{\Pr\{X = x_{obs} | \hat{Y} = \hat{y}_a\}}{\Pr\{X = x_{obs} | \hat{Y} = \hat{y}_b\}} \right) \quad (11)$$

$$= \log \left(\frac{\int_{R_{\hat{X}}} \Pr\{X = x_{obs} | \hat{X} = \hat{x}\} \Pr\{\hat{X} = \hat{x} | \hat{Y} = \hat{y}_a\} d\hat{x}}{\int_{R_{\hat{X}}} \Pr\{X = x_{obs} | \hat{X} = \hat{x}\} \Pr\{\hat{X} = \hat{x} | \hat{Y} = \hat{y}_b\} d\hat{x}} \right), \quad (12)$$

for $\hat{y}_a, \hat{y}_b = \operatorname{argmax}_{\hat{y}_a, \hat{y}_b}(l_{\hat{Y}})$ as before.

C Hardware Entropy Source Schematic and PCB





D EuroSys 2019 Poster: Safeguarding Sensor Device Drivers Using Physical Constraints

This poster [5] was submitted to EuroSys 2019, to communicate the idea of using information about a physical system to condition electronic sensor measurements. The example discussed in the poster is the detection and safeguarding against transduction attacks [15], where an attacker manipulates a sensor’s output to gain control over a system. For example, research by Liu, Yan and Xu [16] has shown that the proximity sensors on a Tesla automobile can be fooled into providing erroneous (or even no) data using ultrasonic interference produced by a device built using off-the-shelf electronic components. An attacker is able to influence the system’s behaviour without needing to directly manipulate the execution of the vehicle’s firmware.

The poster illustrates how the Newton language can be used to describe a physical system, in this case an accelerometer mounted to a PCB without vibration isolation. In this scenario, vibrations of the PCB (e.g. due to a nearby loudspeaker or even, in the case of a smartphone, due to loudspeakers mounted to the PCB) are measured by the accelerometer, obscuring a *true* acceleration measurement i.e. acceleration of the entire system due to gravity. This effect is particularly pronounced if the board is driven at its resonant frequency, since the resulting oscillation will have a greater amplitude [17]. This phenomenon could, in theory, be used as part of a transduction attack e.g. where a smartphone’s loudspeaker is used to interfere with a software application’s estimate of the device’s physical orientation.

To test whether this phenomenon can be distinguished from regular measurement noise, I performed an experiment [18] using an MMA8451Q accelerometer mounted on an FRDM-KL03Z board. With the sensor resting stationary on a horizontal surface, five minutes worth of accelerometer samples were recorded at 10Hz in order to obtain a probability distribution for the accelerometer’s measurement noise (along the z axis, aligned with the vertical); this data was empirically observed to fit a Laplace distribution. This measurement was then repeated with the board resting on top of a smartphone playing a 440Hz audio tone — in theory, this data (random samples from a sinusoid) would fit a bimodal beta distribution (see section D.1 for derivation) allowing a log likelihood ratio to be computed:

$$LLR = -2 \sum_{i=1}^N \frac{\frac{1}{\pi} \left| \frac{1}{\sqrt{A^2\omega^4 - \ddot{x}_i^2}} \right|}{\frac{1}{2b} \exp\left(-\frac{|\ddot{x}_i - \mu|}{b}\right)} \quad (13)$$

The value of this log-likelihood ratio indicates whether the audio tone based transduction attack described here is likely to be occurring. For the data collected in the experiment, the log-likelihood ratio was found to be around -24600.08 whilst the tone was playing, and +3731 when it was not; the negative value indicates that the measurements taken whilst the tone was playing were indeed more accurately described by the bimodal beta distribution, compared to the Laplace distribution of sensor noise, and vice versa for the positive value.

D.1 Derivation of Bimodal Beta Distribution

Let T be a uniformly distributed random variable representing the time at which acceleration is sampled:

$$T \sim U\left(-\frac{\pi}{\omega}, \frac{\pi}{\omega}\right),$$

$$f_T(t) = \begin{cases} \frac{\omega}{2\pi} & -\frac{\pi}{\omega} \leq t \leq \frac{\pi}{\omega} \\ 0 & \text{otherwise} \end{cases}, \quad (14)$$

$$F_T(t) = \begin{cases} 0 & t < -\frac{\pi}{\omega} \\ \frac{\omega(t+\pi/\omega)}{2\pi} & -\frac{\pi}{\omega} \leq t \leq \frac{\pi}{\omega} \\ 1 & t > \frac{\pi}{\omega}. \end{cases} \quad (15)$$

Then let $X = g(T)$ represent the acceleration value at time T . By considering the dynamics of the system, we know that:

$$g(t) = A\omega^2 \sin(\omega t), \quad (16)$$

where A is a frequency-dependent constant equal to the product of quality factor and board displacement at resonance. We can also define a monotonically increasing function $h(x)$ as the inverse of $g(t)$:

$$h(x) = g^{-1}(x) = \frac{1}{\omega} \arcsin\left(\frac{x}{A\omega^2}\right). \quad (17)$$

The cumulative distribution function for X can therefore be written in terms of the cumulative

distribution function for T, accounting for the fact that $g(t)$ is not monotonic:

$$F_X(x) = F_T(h(x)) + \left(1 - F_T(\pi/\omega - h(x))\right). \quad (18)$$

Taking the derivative results in the probability density function for X:

$$f_X(x) = \left(f_T(h(x)) - f_T(-h(x))\right) \frac{d(h(x))}{dx} \quad (19)$$

$$= \begin{cases} \frac{1}{\pi\sqrt{A^2\omega^4-x^2}} & |x| \leq A\omega^2 \\ 0 & \text{otherwise} \end{cases}. \quad (20)$$