Multi-level Random Telegraph Noise Analysis Using Machine Learning Techniques

by

AJ Malcolm

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of

Master of Applied Science
in
Electrical and Computer Engineering

Waterloo, Ontario, Canada, 2020

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

RTN (Random telegraph noise) is a noise process which occurs in solid-state electrical devices such as MOSFETs (metal-oxide-semiconductor field effect transistors) and Josephson Junctions. Defects in the crystal structure of these devices trap charge carriers, resulting in modulations of the devices electrical transport properties such as mobility μ or threshold voltage V_{th}. This is observed as sudden transitions between two discrete current, voltage, or resistance levels, corresponding to the occupied/unoccupied states of the trap. The magnitude of a traps effects can be linked to its physical location in devices, and so we sought to apply RTN analysis during the characterization of devices such as MOSFETs to learn more about its structure and any trap dependence on temperature or bias levels. However, device measurements demonstrated a large proportion of RTN signals with more than two levels, which strongly suggests the presence of multiple charge defects. This scenario is commonly avoided in published research on RTN analysis, with most literature focusing on measurements showing the effects of only a single trap. The frequency with which multiple traps was observed in our measurements motivated the development of an algorithm to better characterize multi-level RTN, and to avoid discarding large swathes of measurement data.

The developed algorithm applies mixture models formed with Gaussians to identify, isolate, and analyze RTN signals. This is accomplished through the use of machine learning techniques to maximize the likelihood that a constrained combination of these models describes the RTN components at every step. When multi-level RTN is present, it is further decomposed into its constituent components which allows characterization of each independent defects. The algorithm is applied to a set of cryogenic MOSFET measurements taken from near cutoff and into saturation, which demonstrates the ability to characterize trap count, trap amplitude, and state occupation distribution for each trap. Although development of the algorithm has precluded an in depth exploration of the effects of parameters such as temperature on these traps, that type of analysis could not have been achieved to a suitable level of accuracy without it.

Acknowledgements

I would like to thank all the people who made this thesis possible: from past and present members of our group, the various staff in QNFCF, IQC, and WIN for teaching me how to use equipment and assisting when I had issues, and anyone else who took time to teach me something, answer a question, or just generally be there.

I would like to further extend my thanks to my supervisor, Dr. Na Young Kim, who was helpful in making the transition to graduate studies and provided motivation, guidance, and support in navigating research projects like this.

I would also like to say thank you to my friends and family who supported me through this, and who were always understanding when I was busy working but were always there to lend an ear or just to chat.

Finally, to Kersty, thank you – you were on board to drop everything and move here the moment I brought it up, even in the face of uncertainty. You picked up so much slack that I am not really sure how to even properly thank you – were the roles reversed, I would not have managed as well as you. So thank you – not only for being there, but for all your support and, more importantly, for being the person you are.

Contents

List of Tables

List of Figures

List of Acronyms

BJT bipolar junction transistor 78

CDF cumulative density function 4, 17, 86–89

CMOS complementary metal-oxide-semiconductor 1, 49

EM expectation-maximization 9

GMM Gaussian mixture model 32

HMM hidden Markov model 20, 21, 25

KDE kernel density estimation 29

MLE maximum likelihood expectation 9

MOS metal-oxide-semiconductor 37

MOSFET metal-oxide-semiconductor field effect transistor 2, 32, 36–38, 42, 43, 62, 78

NMOS n-MOSFET 36–38

PDF probability density function 4, 8, 22, 86, 88

PMF probability mass function 4, 17, 32, 87, 89, 90

PMOS p-MOSFET 36–38, 49

PSD power spectral density 7, 8, 43, 44, 46

RTN random telegraph noise 1, 2, 11, 14, 16, 20–23, 25, 27–29, 31, 32, 36, 38, 41, 43, 48–51, 55, 62, 63

Chapter 1

Introduction

Moore's Law, the name given to the 1965 prediction by George Moore [39], has been driven largely by shrinking the size of transistors through improvements in fabrication techniques. However, it was recognized at least 20 years ago that as device dimensions approach their natural limits, particularly for ubiquitous CMOS technologies, more significant innovations would be required to prevent Moore's law from slipping away. At the feature sizes necessary for Moore's law to continue, improvements in device geometry [21], materials [7, 37], and fabrication technologies have all been realized and integrated into production workflow. Despite this effort Moore's law is beginning to fail, as the dimensions of modern CMOS devices are at the limits of classical physics and quantum phenomena such as tunneling are a legitimate concern. As a result, this has spurred on extensive research into so-called *post-silicon* technologies. Silicon-based devices are often deployed alongside these new technologies, to act as control or interfacing circuitry due to their well known behaviour. Still, device miniaturization and operational conditions, such as at cryogenic temperatures, means there is still knowledge to be gained from characterizing these classical devices.

1.1 Random Telegraph Noise

One technique for characterization is the study of fundamental noise processes that affects devices during operation, such as RTN. RTN is a noise process involving random and instantaneous transitions between two (or more) discrete levels of an observable such as current, and which tends to affect smaller devices more strongly and has a larger impact at cryogenic temperatures. RTN has three defining characteristics:

1. The transition amplitude: ΔRTN (when measuring current).
2. The capture lifetime $\overline{\tau_c}$, the average time a carrier is trapped for.
3. The emission lifetime $\overline{\tau_e}$, the average time between trapping events.

These quantities define the signals time domain procession, as shown in Figure 1.1.

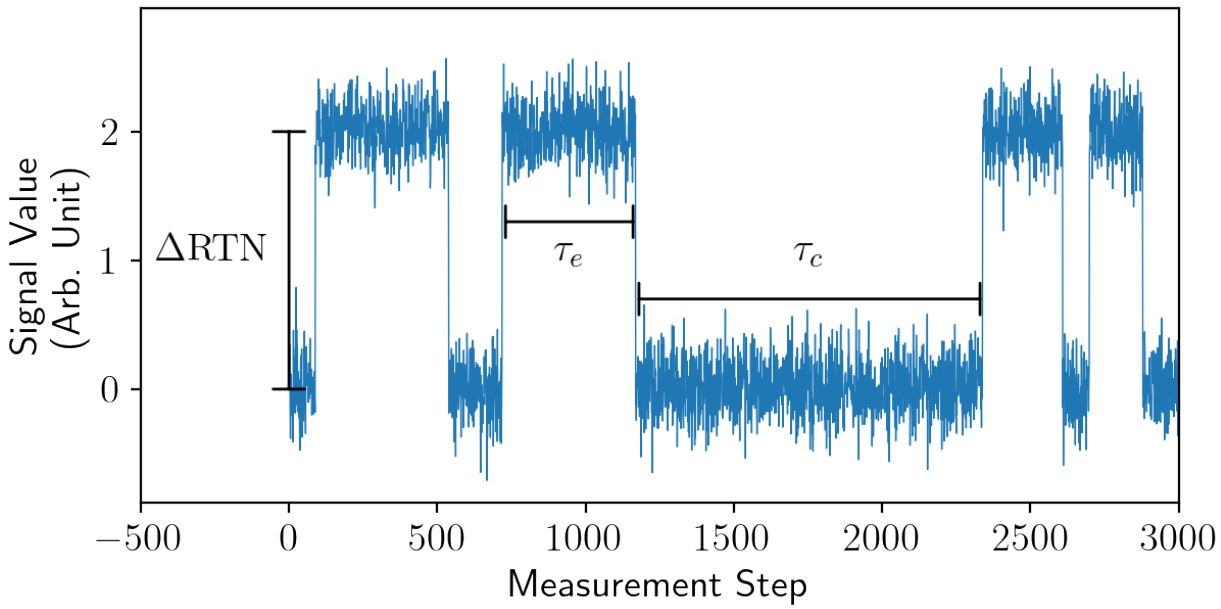


Figure 1.1: The main parameters in an RTN signal are the trap amplitude ΔRTN and the time constants $\overline{\tau_c}$ and $\overline{\tau_e}$.

My research has focused on the study of RTN in MOSFET devices for device characterization purposes, and has involved the direct measurement of devices at cryogenic temperatures and the development of an algorithm for analysis of the measurement data.

This thesis will cover the scope of this research, starting in Chapter 2 with a discussion of the theoretical background for modelling stochastic processes and a derivation of RTN models. Chapter 3 will discuss various implementations of modelling techniques, including a description of the developed algorithm. Following that will be Chapter 4 with a physical description of RTN in MOSFETs, and finally Chapter 5 will demonstrate the application of the algorithm on measured data and the resulting characterization.

Chapter 2

Stochastic Processes

When performing experimentation, one can see only the outcome of the underlying processes - for example, we do not observe *force* but rather the outcome of force acting on an object. In many ways, experimentation is like observing the output of a black-box. One approach to understanding what happens inside the black-box is to use inference to describe the process that leads to the outcome. Complicating the matter is many processes have random output values, which are represented as a random variable. When each observation can be uniquely indexed (such as with points in space or more commonly time) the process is called stochastic. Put another way, a process whose observed outcome is a random variable indexed over a mathematical set of points is called a stochastic process. We can define the distribution of outcomes of a random variable as $f(x)$, a function of each outcome. Then the distribution $f_x(t)$ of a stochastic process $x(t) = x_1, x_2, \ldots, x_N$ can be described as a result of the sequence of random variables and their distributions over time:

$$f_x(t) = f(x(t_1), x(t_2), x(t_3), \ldots, x(t_N)) \tag{2.1}$$

where the right-hand side term represents the distribution of the random variable at each point, from $t = 1$ to $t = N$.

2.1 Properties of Stochastic Processes

2.1.1 Probability Density and Mass

The distribution of outcomes for a random variable depends on whether the outcomes are defined over continuous space or over discrete space. For continuous random variables the

distribution is defined using a PDF, while the PMF defines the distribution of a discrete random variable. Both functions express the random variable as a function of the outcome x defined over the sample space of possible outcomes $\mathscr{X}$. A PDF is written $f(x)$ and defines the probability of a an observation in an arbitrary interval $[a, b]$:

$$P[a \leq x \leq b] = \int_a^b f(x)dx \tag{2.2}$$

To ensure that probability is properly behaved everywhere, the following properties are given:

$$f(x) \geq 0 \tag{2.3}$$

and

$$\int_{\mathscr{X}} f(x)dx = 1. \tag{2.4}$$

In other words, the PDF is non-negative everywhere and the integral over all outcomes must equal one.

The PMF is similarly defined, though with changes to account for the discrete outcomes. It is denoted using $p_X(x)$, with

$$P(X = a) = p_X(a). \tag{2.5}$$

Likewise, it must be non-negative everywhere and the sum over all outcomes must equal one:

$$p_X(x) \geq 0 \tag{2.6}$$

$$\sum_{\mathscr{X}} p_X(x)dx = 1 \tag{2.7}$$

Both the PDF and PMF can also be expressed with a CDF denoted with $F(X)$, where:

$$P(X \leq x) = F(x) \tag{2.8}$$

In other words, it is the cumulative probability of all outcomes up to and including x. For a continuous random variable,

$$F(a) = \int_{-\infty}^{a} f(x)dx \tag{2.9}$$

and for a discrete random variable,

$$F(a) = \sum_{x \leq a} p_X(x) \tag{2.10}$$

2.1.2 Moments and Statistics

The expected value and variance are descriptors of the models shape, and because it is expressed as a mathematical function they are easily calculated as mathematical moments. The r-th moment of a function $f(x)$ around a value c is:

$$\mu_r = \int_{-\infty}^{\infty} (x-c)^r f(x) dx \tag{2.11}$$

though $c = 0$ unless otherwise explicitly stated. The expectation value, or mean, $E[x]$ of a distribution is defined as the first moment:

$$\mathrm{E}[x] = \mu = \int_{-\infty}^{\infty} x f(x) dx. \tag{2.12}$$

The second moment represents the variance, expressed as $\mathrm{Var}(x)$ or σ^2:

$$\sigma^2 = \mu^2 = \int_{-\infty}^{\infty} x^2 f(x) dx \tag{2.13}$$

and represents the spread of the distribution from the expected value.

2.1.3 Multivariate Distribution

When examining two or more random variables, additional statistics are used to describe their relationship. For two random variables X and Y, covariance (sometimes written as σ) is the deviation of the expected value of their product from their individual expected values:

$$\begin{aligned} \mathrm{cov}(X,Y) = \sigma_{X,Y} &= \mathrm{E}[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])] && (2.14) \\ &= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] && (2.15) \end{aligned}$$

A similar statistic is the correlation ρ, which scales the the covariance by the product of the individual variances:

$$\mathrm{corr}(X,Y) = \rho_{X,Y} = \frac{\mathrm{cov}(X,Y)}{\sigma_X \sigma_Y} \tag{2.16}$$

where σ_X and σ_Y are the variances of X and Y respectively.

The two random variables need not be from separate processes — it is equally valid to define them from different indices in the stochastic process. For instance, let X represent

the random variable X at time t_1 and Y represent the random variable X at time t_2. When the correlation function ρ is written using time delayed variables from the same process, it is called the autocorrelation function ϕ:

$$\rho_{X,Y} = \rho_{X(t_1),X(t_2)} = \phi_x(t_1, t_2) \tag{2.17}$$

More formally, the autocorrelation is defined as

$$\phi_x(t_1, t_2) = \overline{x(t_1)x(t_2)} = \int_{\infty}^{-\infty} \int_{\infty}^{-\infty} x_1 x_2 f_x(x_1, x_2) dx_1 dx_2 \tag{2.18}$$

Finally, the multivariate distribution z of independent random variables $f, g, h, \ldots$ is the convolution of their distributions:

$$z(x) = f(x) * g(x) * h(x) * \ldots \tag{2.19}$$

One method for solving a convolution is the Fourier transform. A convolution in the original domain becomes multiplication under a Fourier transform, resulting in

$$\mathscr{F}\{z(x)\} = \mathscr{F}\{f(x)\}\mathscr{F}\{g(x)\}\mathscr{F}\{h(x)\} \tag{2.20}$$

2.1.4 Stationarity

A stochastic process is said to be stationary if

$$f(x(t_1), x(t_2), \ldots, x(t_N)) = f(x(t_1 + t_0), x(t_2 + t_0), \ldots, x(t_N + t_0)) \tag{2.21}$$

where t_0 is an arbitrary delay. That is to say, the distribution of the process in some window of length N is invariant under a shift in time - it is called stationary to the order N.

When a process is stationary to the second order, then the autocorrelation function (Equation 2.18) is a function of the difference $\tau = t_2 - t_1$:

$$\phi_x(t_1, t_2) = \phi_x(\tau) = \overline{x(t)x(t + \tau)} \tag{2.22}$$

From this, we can define a stochastic process as wide-sense stationary when

1. $\phi_x(t_1, t_2) = \phi_x(\tau)$, and
2. $\overline{x(t)}$ = constant.

That is, a stationary process of order 2 or greater is a wide-sense stationary process.

2.2 Spectral Density

Accurate identification of a stochastic process in the time domain can be challenging and depends strongly on the observation "window". An alternative approach which is highly effective is through studying the spectral density, or more often the PSD, of the stochastic process in the frequency domain. The PSD, $S(f)$, of a signal describes the density of power in a frequency range, with units of watts/hertz. Depending on what is being measured, it may alternatively have units of amps^2/hertz or volts^2/hertz. Typically, an arbitrary noise process is classified as low-frequency noise, high-frequency noise, or white noise based on the frequency response of its PSD. Low frequency noises have a PSD with an inverse power law relation $S(f) \propto 1/f^{\alpha}$ and high frequency noise processes have a PSD which follows a power law relation $S(f) \propto f^{\alpha}$ (where α determines the degree of frequency dependence). In comparison, white noise is frequency independent; see Figure 2.1 for a visualization of the differences.

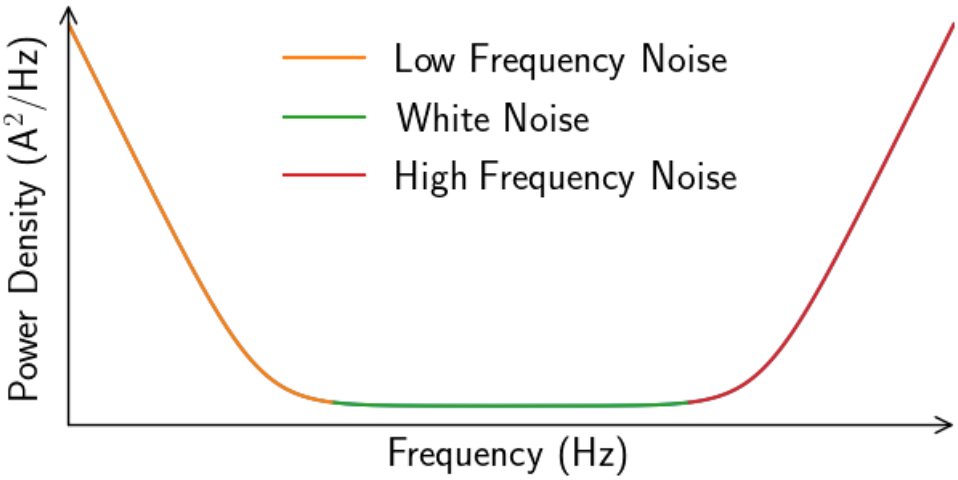


Figure 2.1: The spectral densities of stochastic process can typically be grouped based on whether they are low-frequency dominant, high-frequency dominant, or frequency independent (i.e. white noise).

For a random process $x(t)$, the PSD is given by

$$S_x(f) = \lim_{T\to\infty}\left(\frac{\overline{|X_T(f)|^2}}{T}\right) \tag{2.23}$$

where

$$X_T(f) = \int_{-T/2}^{T/2} x(t) e^{-i2\pi ft} dt. \tag{2.24}$$

As $x(t)$ can be difficult to define in such an analytical form, an alternative approach for obtaining the PSD is through the Wiener-Khinchin Theorem, which states that the PSD of a wide-sense stationary process $x(t)$ is given by the Fourier transform of its autocorrelation function:

$$S_x(f) = \mathscr{F}[\phi(t)] = \int_{-\infty}^{\infty} \phi(t) e^{-i2\pi ft} dt. \tag{2.25}$$

2.2.1 Likelihood

As shown, a statistical model is defined not only by its shape but also by its parameters, denoted as a vector θ. To calculate the probability of the output that has value x, the parameters θ are assumed fixed and known and the equation is expressed:

$$P(X = x; \theta) \tag{2.26}$$

In experimentation however, the output is known (through measurement) while the parameters are not. But, just as the outputs can be expressed as a function of the fixed parameters, we can express the parameters as a function of the known output. This is known as the likelihood function, and is a goodness-of-fit metric for a model. There are slight differences depending on the sample space and whether is is discrete or continuous. For a discrete model the likelihood function is defined as:

$$\mathscr{L}(\theta|x) = P(x|\theta), \tag{2.27}$$

and for a continuous model:

$$\mathscr{L}(\theta|x) = f(x|\theta), \tag{2.28}$$

The likelihood function estimates how well the model describes the observations x for a given set of intervals θ. Thus, the goal in modelling is to maximize the likelihood function - this can be done with any number of optimization techniques. Typically people work with the *log-likelihood* function rather than the likelihood function, as it simplifies various aspects of optimization. For example, the logarithmic function is a strictly increasing function, so maximization of the log-likelihood function is equivalent to maximizing the likelihood function. In addition, finding a global maximum for a function requires concavity - a number of PDFs are only log-concave, such as the exponential and normal distributions.

Two common methods which use the likelihood function to form a statistical model are the MLE and the more involved EM algorithm. MLE involves finding the set of parameters which maximizes the likelihood of the statistical model matching the observed data:

$$\max_{\theta} \mathscr{L}(\theta|x) \tag{2.29}$$

If the function is differentiable, the simplest technique to accomplish this is likely by performing root-finding on the function:

$$\frac{\delta \mathscr{L}}{\delta \theta_i} = 0 \tag{2.30}$$

for $\theta_i \in \theta$. The EM algorithm calculates the expectation of the likelihood function $\mathscr{L}$ based on a set of parameters, and then finds new parameters by maximizing that expectation. This process is repeated until some a maximum is found, although it is not guaranteed to be a global maximum. Thus, EM uses MLE in the maximization step of the algorithm.

2.3 Markov Models

A Markov model is used to describe a stochastic process which transitions between discrete states in either a discrete time domain or continuous time domain - Figure 2.2 shows the state diagram and transition probabilities for a two-state Markov model. They differ from other stochastic models in that they satisfy the Markov property, which states that the future states depend only on the current state. In other words, they are *memoryless*. In this section, I will derive the statistical models which describe a continuous Markov *process* or discrete Markov *chain*.[1]

2.3.1 Continuous Markov Process

Consider an arbitrary system with the following properties:

1. The system state can occupy one of two discrete states, labelled x_0 and x_1.
2. The system state can be written as $X(t)$.

[1]The words *process* and *chain* are often used to differentiate Markov models in continuous or discrete time.

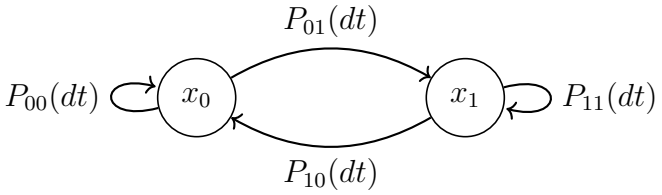


Figure 2.2: State diagram for a Markov model with possible states x_0 and x_1. Over a step in time dt, the probability of transitioning from state $i \in x_0, x_1$ to $j \in x_0, x_1$ is given by P_{ij}.

3. $\mathbf{E}(X(t_0)) = x_0 P\{X(t_0) = x_0\} + x_1 P\{X(t_0) = x_1\}$
4. $P_{ij}(t) = P\{X(t_0 + t) = j \mid X(t_0) = i\}$
    - $i, j \in \{x_0, x_1\}$
    - The system is memoryless.
5. $P\{X(t_0 + t) = j \mid X(t_0) = i\} = P\{X(t) = j \mid X(0) = i\}$
    - The system is time homogenous — in other words, the starting point does not matter, only the window t.

A system which satisfies these properties is known as a continuous time Markov process (CTMP), or continuous Markov process. For clarity moving forward, we will force the condition that if a transitions occurs at time t, then $X(t)$ is the new state and $X(t) \neq X(t^-)$ (where t^- is the time just before t).

In meeting the requirements of a continuous Markov process, we can see that the behaviour of the transition events will also meet that of a Poisson point process. Given a time window t and random variable T for interarrival time, we will have:

$$1 = P\{T \leq t\} + P\{T > t\} \tag{2.31}$$

or

$$P\{T \leq t\} = 1 - P\{T > t\} \tag{2.32}$$

$P\{T > t\}$ is equal to the probability that zero transition events occur in time t, which

is easily calculated with the Poisson distribution:

$$\begin{aligned} P\{k,t\} &= \frac{(\lambda t)^k e^{-\lambda t}}{k!} \\ P\{0,t\} &= \frac{(\lambda t)^0 e^{-\lambda t}}{0!} \\ P\{0,t\} &= e^{-\lambda t} \end{aligned} \tag{2.33}$$

where the transition events occur at a rate $\lambda = 1/\tau$. Thus we can write

$$P\{T <= t\} = 1 - e^{-t/\tau} \tag{2.34}$$

which is the CDF of the exponential distribution (as $t \to \infty, P \to 1$), and would thus lead to the PDF through a simple differentiation. This supports our assertion of a memoryless system, as the exponential distribution has the same property. This treatment also agrees with one experimental observation, that the distribution of the state lifetimes is exponential with a state dependent mean lifetime $1/\tau$.

Because there are only two possible states our RTN system can occupy, we can also assert the probability that the system will be in either state under inspection through a ratio of their average lifetimes:

$$P\{X(t) = x_0\} = \frac{\tau_0}{\tau_0 + \tau_1} \tag{2.35}$$

$$P\{X(t) = x_1\} = \frac{\tau_1}{\tau_0 + \tau_1} \tag{2.36}$$

These probabilities form the stationary distribution of the Markov model, and together satisfy the need for

$$P\{X(t) = x_0\} + P\{X(t) = x_1\} = 1$$

From our earlier condition, we can write the expectation value of the signal at any arbitrary time t_0 as:

$$\begin{aligned} \mathbf{E}(X(t_0)) &= x_0 P\{X(t_0) = x_0\} + x_1 P\{X(t_0) = x_1\} \\ \mathbf{E}(X(t_0)) &= x_0 \frac{\tau_0}{\tau_0 + \tau_1} + x_1 \frac{\tau_1}{\tau_0 + \tau_1} \end{aligned} \tag{2.37}$$

After a delay t in which transitions can occur, we can write the expectation value of the signal as

$$\mathbf{E}(X(t_0 + t)) = x_1 P_{11}(t) + x_1 P_{01}(t) + x_0 P_{10}(t) + x_0 P_{00}(t), \tag{2.38}$$

where $P_{ij}(t)$ is the probability of beginning in state i and ending in state j after time t. We would like to use these equations to calculate the autocorrelation of the signal after a delay t:

$$\phi(t) = \langle X(t_0)X(t_0+t)\rangle \tag{2.39}$$

However, we must impose the following condition:

$$X(0)P_{ij} \neq 0 \iff X(0) = i$$

In words, the initial state of the system must be the same as the initial state in the transition probability to be non-zero. Using this condition, and a simplification of the math by setting $x_1 = \Delta X$ and $x_0 = 0$, we can explicitly write the autocorrelation function:

$$\begin{aligned}
\phi(t) &= \langle X(t_0)X(t_0+t)\rangle \\
\phi(t) &= \left\langle \Delta X \frac{\tau_1}{\tau_0+\tau_1} \Delta X P_{11}(t) \right\rangle \\
\phi(t) &= \left\langle (\Delta X)^2 \frac{\tau_1}{\tau_0+\tau_1} P_{11}(t) \right\rangle
\end{aligned} \tag{2.40}$$

We will need to solve for $P_{11}(t)$ to continue. Starting in state x_1, we can write

$$P_{11}(t) + P_{10}(t) = 1 \tag{2.41}$$

Furthermore, we can split the time window t into another window $t' + dt$. Writing

$$P_{11}(t+dt) = P_{11}(t)\left(1 - \frac{dt}{\tau_1}\right) + P_{10}(t)\left(\frac{dt}{\tau_0}\right) \tag{2.42}$$

In words, the probability of ending in the same state as you started after $t + dt$ is the combination of two mutually exclusive events:

1. The probability of returning to the same state after time t and not transitioning in dt.

2. The probability of being in the other state after time t, and then transitioning back in dt.

Substituting (2.41) in and taking $dt \to 0$, we get

$$\begin{aligned}
P_{11}(t+dt) &= P_{11}(t)\left(1-\frac{dt}{\tau_1}\right)+(1-P_{11}(t))\left(\frac{dt}{\tau_0}\right)\\
P_{11}(t+dt) &= P_{11}(t)-\frac{P_{11}(t)dt}{\tau_1}+\frac{dt}{\tau_0}-\frac{P_{11}(t)dt}{\tau_0}\\
\frac{P_{11}(t+dt)-P_{11}(t)}{dt}&+P_{11}(t)\left(\frac{1}{\tau_1}+\frac{1}{\tau_0}\right)=\frac{1}{\tau_0}\\
\frac{dP_{11}(t)}{dt}&+P_{11}(t)\left(\frac{1}{\tau_1}+\frac{1}{\tau_0}\right)=\frac{1}{\tau_0}
\end{aligned} \tag{2.43}$$

which is a differential equation in the form of Kolmogorov's Forward Equation, with the solution

$$P_{11}(t)=\frac{\tau_1}{\tau_1+\tau_0}+\frac{\tau_0}{\tau_1+\tau_0}e^{-\left(\frac{1}{\tau_0}+\frac{1}{\tau_1}\right)t}. \tag{2.44}$$

After solving $P_{11}(t)$ we can substitute it back into (2.40), resulting in

$$\begin{aligned}
\phi(t) &= \left\langle (\Delta X)^2\frac{\tau_1}{\tau_0+\tau_1}P_{11}(t)\right\rangle\\
\phi(t) &= \left\langle (\Delta X)^2\frac{\tau_1}{\tau_0+\tau_1}\left(\frac{\tau_1}{\tau_0+\tau_1}+\frac{\tau_0}{\tau_0+\tau_1}e^{-\left(\frac{1}{\tau_0}+\frac{1}{\tau_1}\right)t}\right)\right\rangle\\
\phi(t) &= \left\langle \left(\frac{\Delta X\tau_1}{\tau_0+\tau_1}\right)^2+\frac{(\Delta X)^2\tau_0\tau_1}{(\tau_0+\tau_1)^2}e^{-\left(\frac{1}{\tau_0}+\frac{1}{\tau_1}\right)t}\right\rangle.
\end{aligned} \tag{2.45}$$

We make use of the Wiener-Khinchin theorem to calculate the power spectral density (PSD) of the signal from (2.45). Because $X(t)$ is a stationary process, the Wiener-Khinchin theorem uses a Fourier transform in the form

$$S_x(f)=4\int_0^\infty \phi_x(t)\cos(2\pi ft)dt, \tag{2.46}$$

which after substituting $\phi_x(t)$ with Equation 2.45 yields

$$S_x(f)=\frac{1}{\pi}\frac{(\Delta X)^2\tau_0\tau_1}{(\tau_0+\tau_1)^2}\frac{1}{1+\left(\frac{2\pi f\tau_0\tau_1}{\tau_0+\tau_1}\right)^2}+S_x(0), \tag{2.47}$$

where $S_x(0)$ is a DC term and does not contribute to overall noise.[2] Because the sum of two independent Poisson rates $\lambda_0 + \lambda_1 = \lambda_A$, we can offer an alternative form using $\frac{1}{\tau_0} + \frac{1}{\tau_1} = \frac{1}{\tau_A}$.

$$S_x(f) = \frac{1}{\pi} \frac{(\Delta X)^2 \tau_A}{\tau_0 + \tau_1} \frac{1}{1 + (2\pi f \tau_A)^2} \tag{2.48}$$

As we are working with frequency, it can be useful to use the frequency versions of the time constants. Forgoing the previously used notation λ, we will now use f_i, where i is the state. With this notation, we can write f_0 and f_1 for the two unique states, and another f_c for the combined time constant τ_A (the notation f_c is inspired by cutoff frequency in filter design).

$$S_x(f) = \frac{f_0 f_1}{\pi} \left(\frac{\Delta X}{f_c} \right)^2 \frac{1}{1 + (2\pi f / f_c)^2}$$

There are two interesting cases when considering the combined frequency f_c:

$$S_x(f) = \begin{cases} \dfrac{f_0 f_1}{\pi} \left(\dfrac{\Delta X}{f_c} \right)^2 & \text{for } f \ll f_c; \\ \dfrac{f_0 f_1}{\pi} \left(\dfrac{\Delta X}{2\pi f} \right)^2 & \text{for } f \gg f_c; \end{cases}$$

For $f \ll f_c$ the system has a white noise spectrum, and for $f \gg f_c$ the system has a Brownian spectrum.

2.3.2 Discrete Markov Chain

In contrast to the continuous time model description RTN can also described using a discrete time model, which is more suited to describing a finite-length measurement rather than the physical phenomenon behind RTN. Consider a system with the following properties:

1. The system state can occupy one of two discrete states, labelled x_0 and x_1.
2. The system state can be written as $X_s, \quad s = 0, 1, 2, \ldots$
   - s corresponds to a step in time.
   - The system can be expressed over time as: $X_0 \to X_1 \to X_2 \to \ldots$

[2] More information on the PSD derivation and the DC term can be found in Appendix B.1.2.

3. $p_i(s) = P\{X_s = i\}$
    - The probability of being in state i at step s
4. $p_{ij} = P\{X_{s+1} = j \mid X_s = i\}$
    - $i, j \in \{x_0, x_1, x_2, \dots\}$
    - Probability of transitioning from state i to state j is independent of the states before time s.

This is a discrete-time version of the Markov chain that was used for the continuous time model. We can combine the last two Markov conditions to express the probability of being in state more explicitly:

$$p_j(s) = P\{X_s = j \mid X_{s-1} = i\}P\{X_{s-1} = i\}$$

Which naturally leads to:

$$\begin{aligned} p_i(s-1) &= P\{X_{s-1} = j \mid X_{s-2} = i\}P\{X_{s-2} = i\} \\ &\vdots \end{aligned}$$

This expression for $p_j(s)$ thus leads to a recursion relation. We can describe such a relation in the following way:

$$p_j(0) = p_j(0) \tag{2.49}$$

$$p_j(s) = \sum_i p_i(s-1)p_{ij} \tag{2.50}$$

However, if we can assign a definite state value i to the system at $s = 0$, then (2.49) vanishes for any state other than i. The recursion equation (2.50) can then be rewritten as:

$$p_j(s) = P\{X_s = j \mid X_0 = i\} = p_{ij}(s) \tag{2.51}$$

In words, it is the probability of transition from state i to state j after s steps, with the following case for $s = 0$:

$$p_{ij}(0) = \begin{cases} 1 & \text{if } i = j; \\ 0 & \text{if } i \neq j; \end{cases} \tag{2.52}$$

The form of the recursion naturally leads to formation of a transition probability matrix for a single step, denoted by P:

$$P = ||p_{ij}|| = \begin{bmatrix} p_{11} & p_{12} & \dots \\ p_{21} & p_{22} & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

And an accompanying s-step transition probability matrix:

$$P(s) = ||p_{ij}(s)|| = \begin{bmatrix} p_{11}(s) & p_{12}(s) & \dots \\ p_{21}(s) & p_{22}(s) & \dots \\ \vdots & \vdots & \ddots \end{bmatrix}$$

It can be shown[3] that the following relation holds:

$$P(s) = P^s \tag{2.53}$$

While this approach works for n possible states, because RTN is a two-state system we can also treat the system as a series of Bernoulli trials. That is to say, at each point in the measured signal we can perform a Bernoulli test on X to find its value. Let $P\{X_s \neq X_{s-1}\} = p_i$ be the probability of a "success", and $P\{X_s = X_{s-1}\} = q_i$ the probability of a "failure". A success (p_i) is defined as a transition from state i, while a failure (q_i) is defined as remaining in state i. In other words, a successful outcome of the Bernoulli test means the system has changed states. We require that if $X_{s-1} = i$ and a test at step s registers a failure, then $X_s = j$ (for $i, j \in \{x_0, x_1\}, i \neq j$).

Consider a signal in state i — let a transition (i.e. a successful Bernoulli test) occur after S number of steps and let s be an arbitrary number of steps. We define two outcomes:

1. No transition occurs in s steps: $S > s$
2. A transition does occur in s steps: $S <= s$

which leads to

$$P\{S <= s\} + P\{S > s\} = 1$$
$$P\{S <= s\} = 1 - P\{S > s\}$$

$P\{S > s\}$ is the probability that no transitions occur in s steps, which can be equivalently described using a binomial distributed random variable[4] S' which describes the

[3] See Appendix B.2.1

[4] Represented by the symbol $\text{Bin}(s, p)$, where s is the number of trials and p is the probability of success.

number of Bernoulli successes r (if a success occurs with probability p_i) in s independent trials:

$$S' \sim \text{Bin}(s, p_i) \tag{2.54}$$

with PMF

$$P\{S' = r\} = P\{r; s, p_i\} = \binom{s}{r} p_i^r q_i^{s-r}. \tag{2.55}$$

Then,

$$\begin{aligned}
P\{S <= s\} &= 1 - P\{S > s\} \\
P\{S <= s\} &= 1 - P\{S' = s\} \\
P\{S <= s\} &= 1 - P\{r = 0; s, p_i\} \\
P\{S <= s\} &= 1 - \binom{s}{0} q_i^s \\
P\{S <= s\} &= 1 - q_i^s
\end{aligned}$$

which yields the geometric distribution CDF, which counts failures until the first success. This is to be expected as the geometric distribution is the discrete equivalent to the exponential distribution, which we found in the continuous time model. It has the same important property of being *memoryless*, which is vital to our assertion of a discrete Markov chain.

We will use σ to denote the expected number of successes before the first failure (from geometric distribution):

$$\sigma_i = \text{E}(S_i) = 1/q_i \tag{2.56}$$

We will also introduce a new notation for the possible states:

$$\vec{x} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \tag{2.57}$$

We will then define[5] the starting probability distribution as a vector π_0:

$$\pi_0 = \begin{bmatrix} P\{X_0 = x_0\} \\ P\{X_0 = x_1\} \end{bmatrix} = \begin{bmatrix} \dfrac{\sigma_0}{\sigma_0 + \sigma_1} \\ \dfrac{\sigma_1}{\sigma_0 + \sigma_1} \end{bmatrix} = \begin{bmatrix} \dfrac{q_1}{q_0 + q_1} \\ \dfrac{q_0}{q_0 + q_1} \end{bmatrix} \tag{2.58}$$

[5]Requiring that if $X_{s-1} = x_0$ and there is a success at step s, $X_s = x_1$ (and vice-versa). This requirement forces $\sigma_0 + \sigma_1 = 1$.

and corresponding expectation value:

$$\mathrm{E}(X_0) = \pi_0^T \vec{x} \tag{2.59}$$

More generally, we can define a probability distribution vector at any state s:

$$\pi_s = \begin{bmatrix} P\{X_s = x_0\} \\ P\{X_s = x_1\} \end{bmatrix}$$

from which it follows:

$$\pi_s^T = \pi_0^T P^s \tag{2.60}$$

$$\mathrm{E}(X_s) = \pi_s^T \vec{x} \tag{2.61}$$

Tying this back into the Markov chain reduces the transition probability matrices to:

$$P = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix} = \begin{bmatrix} p_0 & q_0 \\ q_1 & p_1 \end{bmatrix} \tag{2.62}$$

$$P^s = \begin{bmatrix} p_{00}(s) & p_{01}(s) \\ p_{10}(s) & p_{11}(s) \end{bmatrix} \tag{2.63}$$

We can explicitly solve for P^s through matrix manipulation. We begin by diagonalizing matrix P. Instead of using the characteristic equation to solve for the eigenvalues ($\det(P - \lambda I) = 0$), two useful properties make it much easier:

1. $\lambda_1 = \sigma$ if for all rows i and columns j:

$$\sum_j p_{ij} = \sigma \tag{2.64}$$

   - In words, if the sum of the elements in each row is the same for all rows, that sum is an eigenvalue of the matrix.

2. $\sum_i \lambda_i = \mathrm{Tr}(P)$

   - The trace of the matrix (sum of the diagonal) is equal to the sum of the eigenvalues.

For P, the sum of the rows $p_0 + q_0 = q_1 + p_1 = 1$, which means $\lambda = 1$ is an eigenvalue of P. Furthermore, for a 2×2 matrix, there can be 2 eigenvalues. There:

$$\mathrm{Tr}(P) = p_0 + p_1 = \lambda_1 + \lambda_2$$

Which we can use to solve for λ_2. That makes the eigenvalues of P:

$$\begin{aligned}\lambda_1 &= 1\\ \lambda_2 &= p_0 + p_1 - 1\end{aligned}$$

The corresponding eigenvectors[6] $\vec{u}$ and $\vec{v}$ are:

$$\vec{u} = \begin{bmatrix}1\\1\end{bmatrix} \qquad \vec{v} = \begin{bmatrix}1\\\alpha\end{bmatrix} \tag{2.65}$$

where $\alpha = -\frac{q_1}{q_0}$. We can then diagonalize P using $S = [\vec{u}\ \ \vec{v}]$:

$$S^{-1}PS = \begin{bmatrix}1 & 0\\0 & g(s)\end{bmatrix}$$

where

$$g(s) = \left(\frac{1 - q_0 - 2q_1 - (1 - q_1)\alpha}{1 - \alpha}\right)^s \tag{2.66}$$

$$g(s) = \left(\frac{1 - q_0 - 2q_1 - (1 - q_1)(-q_1/q_0)}{1 - (-q_1/q_0)}\right)^s \tag{2.67}$$

$$g(s) = \left(\frac{1 - q_0 - 2q_1 + q_1/q_0 - q_1^2/q_0}{(q_0 + q_1)/q_0}\right)^s \tag{2.68}$$

$$g(s) = \left(\frac{q_0 - q_0^2 - 2q_1q_0 + q_1 - q_1^2}{q_0 + q_1}\right)^s \tag{2.69}$$

$$g(s) = \left(\frac{-(q_0 + q_1 - 1)(q_0 + q_1)}{q_0 + q_1}\right)^s \tag{2.70}$$

$$g(s) = (1 - q_0 - q_1)^s \tag{2.71}$$

[6] See B.2.3 for calculation.

Then:

$$
\begin{aligned}
P^s &= (S(S^{-1}AS)S^{-1})^s \\
P^s &= S(S^{-1}AS)^s S^{-1} \\
P^s &= S \begin{bmatrix} 1 & 0 \\ 0 & g(s) \end{bmatrix}^s S^{-1} \\
P^s &= \frac{1}{1-\alpha} \begin{bmatrix} -\alpha + g(s) & 1 - g(s) \\ \alpha(g(s) - 1) & 1 - \alpha g(s) \end{bmatrix}
\end{aligned} \tag{2.72}
$$

With P^s defined, we can begin the process to calculate the PSD in the same way as in the continuous model. To start, the autocorrelation function of the discrete time signal is:

$$
\begin{aligned}
R_X(s_0, s_0 + s) &= \overline{X_{s_0} X_{s_0+s}} && (2.73) \\
&= \overline{X_0 X_s} && (2.74) \\
&= \overline{(\pi_0^T \vec{x})(\pi_s^T \vec{x})} && (2.75)
\end{aligned}
$$

The first product is:

$$
\pi_0^T \vec{x} = \Delta X \frac{q_0}{q_0 + q_1} \tag{2.76}
$$

and the second product is:

$$
\begin{aligned}
\pi_s^T \vec{x} &= \pi_0^T P^s \vec{x} \\
&= \Delta X \frac{q_0}{q_0 + q_1}
\end{aligned} \tag{2.77}
$$

2.3.3 Hidden Markov Model

A very well established technique for modelling Markov chains is the HMM. HMM is a tool used to model stochastic processes in cases where the sequence of observations Y_T is generated by a hidden state sequence S_T, demonstrated in Figure 2.3, and where the hidden sequence of states is assumed to satisfy the Markov property.

HMM can be used to model the time domain parameters of such a system, such as transition probability and time constants, and has been applied for characterizing two-level RTN data in the past. However, a limitation of HMM is the mutual exclusivity of states, which leads to an exponential relation between the state space needed to describe a system and the number of possible states. For N independent objects which can each occupy

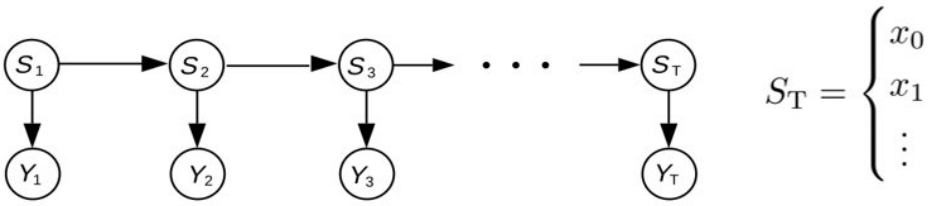


Figure 2.3: The sequence of observations described by the random variable Y_T is generated by the hidden states S_T.

S unique states, an HMM would need N^S total states to describe the state progression. By decomposing the system into N independent subsystems, the state space is reduced to $N \times S$. Ghahramani discussed some variations on the traditional HMM that address this limitation, such as the tree structure HMM, but they are not as readily available in common analysis software. [13]

2.4 RTN as a Mixture Model

A mixture model is a statistical model used when the total sample population is described by a mixture of multiple submodels of the same family, but with different parameters. A time-discrete, two-level RTN signal transitions between two mutually exclusive and independent states. The observed state of the RTN signal at any point in time is given by the random variable ξ, with parameters τ for average time spent in state and transition amplitude $\Delta\xi$. The two states themselves are represented by the random variables X and Y, with mean values μ_X and μ_Y respectively. While the distribution of signal values for pure RTN will consist only of two delta functions located at μ_X and μ_Y, noise factors arising from measurement inaccuracy and other noise sources will result in Gaussian distributed subpopulations. Hence, we can model the overall RTN population as a Gaussian Mixture model,

$$f_\xi = f_X + f_Y \tag{2.78}$$

$$f_\xi = \frac{1}{\sigma_X\sqrt{2\pi}} \exp\left[\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right] + \frac{1}{\sigma_Y\sqrt{2\pi}} \exp\left[\frac{1}{2}\left(\frac{x-\mu_Y}{\sigma_Y}\right)^2\right] \tag{2.79}$$

However, we need to relate these model parameters to the RTN signal parameters for the model to make sense. As the measurement time and number of samples increases, the

probabilities associated with observing either state X or Y when randomly observing the signal will approach the stationary distribution of the signal:

$$P(\xi = X) = \frac{\tau_X}{\tau_X + \tau_Y} \qquad P(\xi = Y) = \frac{\tau_Y}{\tau_X + \tau_Y} \tag{2.80}$$

We use these probabilities to scale the Gaussians in the Gaussian mixture model accordingly:

$$f_\xi = P(\xi = X) f_X + P(\xi = Y) f_Y \tag{2.81}$$

Next, the Gaussian mean values are defined using a base value b and separation value s,

$$\mu_X = b \qquad \mu_Y = b + s \tag{2.82}$$

where b and s represent an offset affecting both peaks (such as a DC offset) and the trap amplitude respectively.

Finally, as white noise processes will affect both states equally, we set $\sigma_X = \sigma_Y = \sigma$. Thus, the total PDF of the mixture model is

$$f(x) = P(\xi = X)\ f_X(x; \mu = b, \sigma) + P(\xi = Y) f_Y\ (x; \mu = b + s, \sigma) \tag{2.83}$$

and plotted in Figure 2.4.

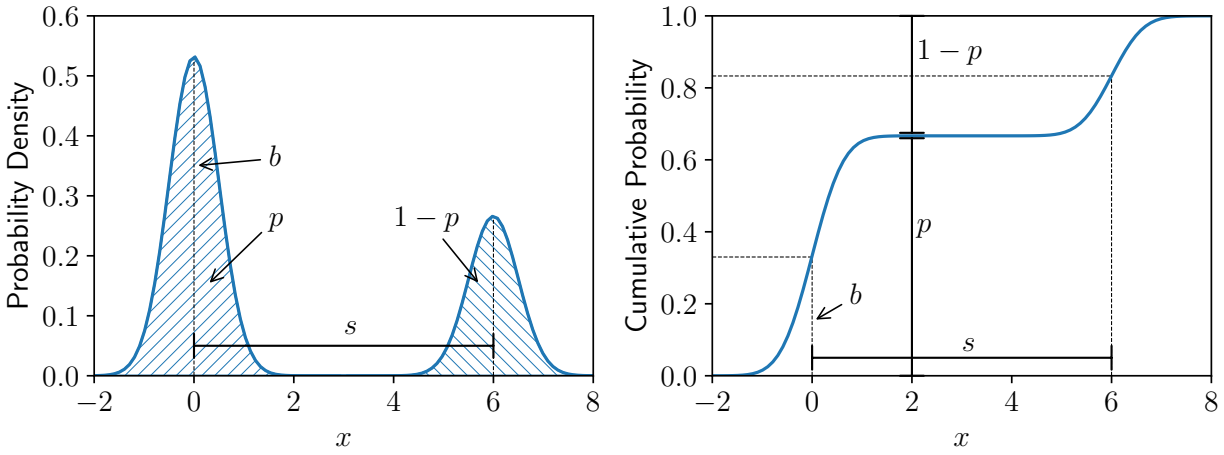


Figure 2.4: Gaussian mixture model with parameters $b = 0$, $s = 6$, and $p = 1/3$.

Modelling a multi-level RTN signal follows from this definition under the assumption that it is the result of multiple independent two-level RTN signals, which will be referred

to as RTN subcomponents. This assumption comes from the physical origins of RTN in devices, and will be explained further in Chapter 4. For each subcomponent $i, j, k, \ldots$ there are associated random variables $\xi_i, \xi_j, \xi_k, \ldots$. Hence, the multi-level distribution is a convolution of multiple two-level distributions. From the associative property of the convolution operator ($(f * g) * h = f * (g * h)$), the operation itself is not affect by the number of signals and the order in which the convolution is performed.

For a single trap signal (with states X, Y), it can be shown

$$\mathscr{F}\{f(x)\} = \mathscr{F}\left\{\frac{1}{\sigma_X\sqrt{2\pi}}\exp\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2 + \frac{1}{\sigma_Y\sqrt{2\pi}}\exp\frac{1}{2}\left(\frac{x-\mu_Y}{\sigma_Y}\right)^2\right\} \tag{2.84}$$

$$= \exp\left(-\sigma_X\sigma_Y\omega^2/2\right)\left[\exp\left(-i\omega\mu_X\right) + \exp\left(-i\omega\mu_Y\right)\right] \tag{2.85}$$

As the measurement and white noise contributions should be invariant under state switching, let $\sigma_X = \sigma_Y$ for a slight simplification in the model. Using this result for a single trap, Equation 2.20 can be easily expanded for multiple subsignals. In the case of two RTN signals, it would become:

$$\begin{aligned}\mathscr{F}\{z(x)\} = \exp\left(-3\sigma^2\omega^2/2\right)\big[&\exp\left(-i\omega(\mu_{X_1}+\mu_{X_2}+\mu_{X_3})\right) + \exp\left(-i\omega(\mu_{Y_1}+\mu_{X_2}+\mu_{X_3})\right)\\ &+ \exp\left(-i\omega(\mu_{X_1}+\mu_{Y_2}+\mu_{X_3})\right) + \exp\left(-i\omega(\mu_{X_1}+\mu_{X_2}+\mu_{Y_3})\right)\\ &+ \exp\left(-i\omega(\mu_{Y_1}+\mu_{Y_2}+\mu_{X_3})\right) + \exp\left(-i\omega(\mu_{Y_1}+\mu_{X_2}+\mu_{Y_3})\right)\\ &+ \exp\left(-i\omega(\mu_{X_1}+\mu_{Y_2}+\mu_{Y_3})\right) + \exp\left(-i\omega(\mu_{Y_1}+\mu_{Y_2}+\mu_{Y_3})\right)\big]\end{aligned} \tag{2.86}$$

where μ_{X_i} and μ_{Y_i} are the two states for the i-th subsignal. The convolution of three Gaussian mixture models is depicted graphically in Figure 2.5d. In this case, the convolution is solved numerically from three vectors each representing a mixture model density function (Figure 2.5a, Figure 2.5b, and Figure 2.5c).

After developing a statistical model which can be used for describing an RTN signal, the next step is application of the model to extract parameters of interest. To accomplish this, the measurement data must undergo processing to convert it from time-domain sampling data to density functions for each subcomponent. This required the development of an analysis algorithm, which will be discussed in the following chapter.

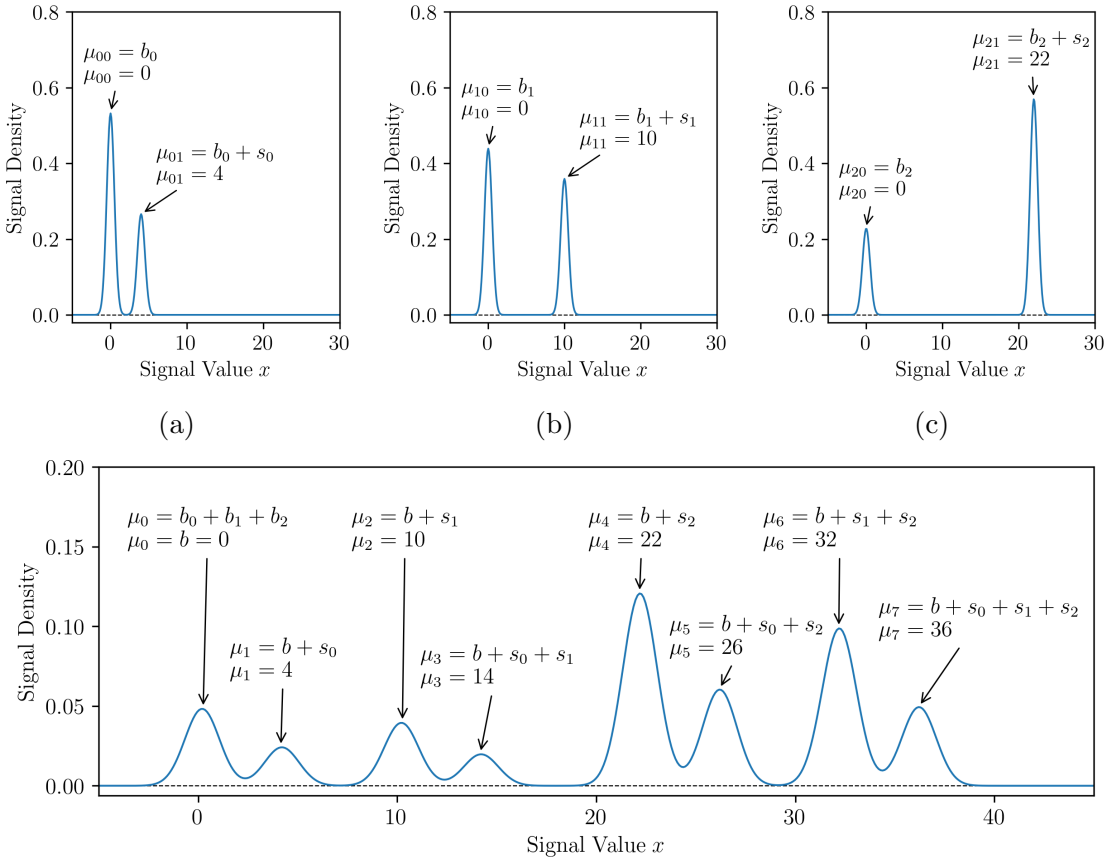


Figure 2.5: (a), (b), (c) Gaussian mixture models with parameters $b_0 = b_1 = b_2 = 0$ and $s_0 = 4$, $s_1 = 10$, $s_2 = 22$ respectively. (d) Numerical convolution calculated using the Python *NumPy* library.

Chapter 3

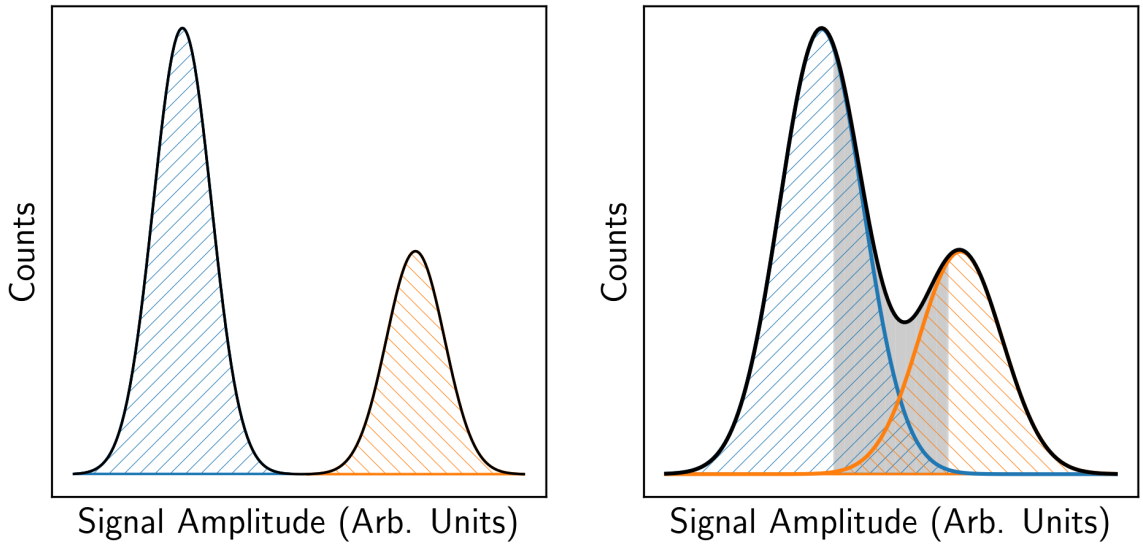
Noise Analysis Algorithm

Over the course of my research, I implemented and developed the analysis algorithm over many stages. The original algorithm used to analyze RTN noise was intended only for use with a two-level system, but when measurements resulted in a much larger number of datasets with multiple (> 2) levels than anticipated, a new algorithm was developed to account for this. This new algorithm uses a combination of machine learning techniques for modelling a multi-level RTN signal, and has been developed over time to provide more accurate analysis. In this chapter I will explain the initial algorithm as well as the evolution of the subsequent algorithm.

3.1 Two-Level Model

The first technique was inspired by [60], and was intended for two-level RTN analysis. The technique is similar to a manually implement HMM, in that each sample is sorted into one of the two states based on the state of the previous sample. With two levels, it is relatively straightforward to sort the samples as the shape of distribution is known a priori to be a mixture of two Gaussians. However, one complicating factor is the degree of separation between the two Gaussians, which is determined by the RTN transition amplitude and amount of white background noise captured during measurement, with varying degrees of separation shown in Figure 3.1. For very well defined RTN, the Gaussians will be distinct and the separation between them will be much greater than their standard deviations, as in Figure 3.1a. On the other hand, if the trap amplitude is smaller in magnitude, or there is a large amount of white noise contributions, then the Gaussians will overlap, like in Figure 3.1b. Sorting samples in the former case is trivial, but sorting criterion needs

to be developed for the latter. More specifically, the criteria must determine the sample classifications in the overlap region, which would be the area defined by the tails of the Gaussians (the grey box in Figure 3.1b).



(a) (b)

Figure 3.1: (a) The signal peaks are well defined, so classification is straightforward. (b) Signal peaks suffer from large overlap, denoted by grey box. Samples in this region may belong to either peak.

Given mutually exclusive states i, j and the signal value at step s, two cases must be addressed for the signal at step $s' = s + 1$.

1. The signal at step s is equal to the signal at s', $x(s) = x(s')$, which occurs with probability:
$$P_{ii} = P(x(s') = i \mid x(s) = i)$$
2. The signal at step s is not equal to the signal at s', $x(s) \neq x(s')$, with probability:
$$P_{ij} = P(x(s') = j \mid x(s) = i)$$

With only two outcomes, $P_{ii} + P_{ij} = 1$, and we can define the sorting criteria based on their values. Given $x(s) = i$,

$$x(s') = \begin{cases} i & \text{if } P_{ii} \geq P_{ij} \\ j & \text{if } P_{ij} > P_{ii} \end{cases}$$

Without a priori knowledge of the RTN transition probabilities or time constants, the sorting criteria cannot be evaluated directly. Instead they are estimated using threshold values I_{TH} unique to each state, which denote the maximum variance of a sample from the state mean before it is considered to be in the other state. Give $x(s) = i$, the sorting criteria can then be rewritten:

$$x(s') = \begin{cases} i & \text{if} \quad P_{ii} \geq P_{ij} \equiv x(s') \leq I_{TH_i} \\ j & \text{if} \quad P_{ii} < P_{ij} \equiv x(s') > I_{TH_i} \end{cases} \tag{3.1}$$

The threshold values are illustrated for overlapping peaks in Figure 3.2.

The threshold is set as a function of the state separation ΔI and the variance σ_i of its associated state. For example, the threshold value of state $i = \{down\}$ may be defined as:

$$I_{TH_{down}} = \mu_{down} + \frac{\Delta I}{2} + \frac{\sigma_{down}}{2} \tag{3.2}$$

The second term is important because to prevent any undefined regions in the domain of the sample criteria. In other words, for two states *down* and *up* where $|\mu_{down}| < |\mu_{up}|$, the threshold values must satisfy:

$$|I_{TH_{down}}| >= |I_{TH_{up}}| \tag{3.3}$$

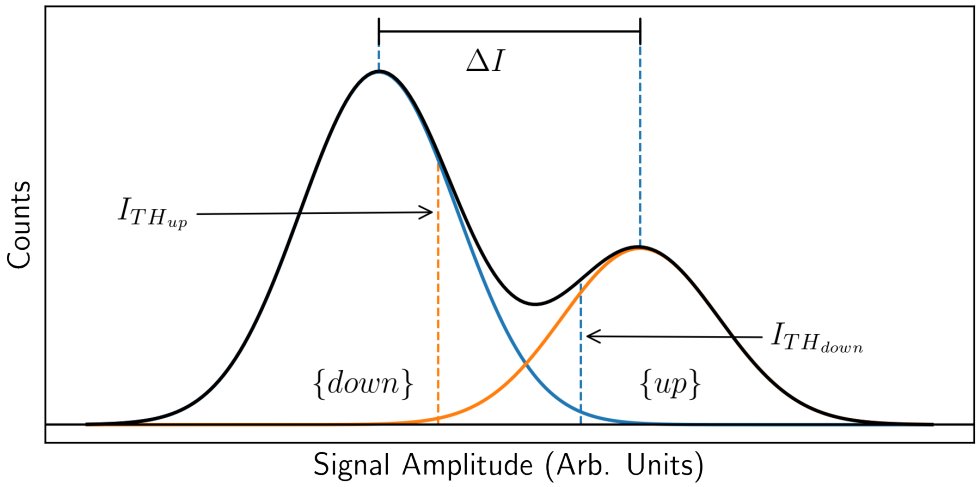


Figure 3.2: The lines $I_{TH_{down}}$ and $I_{TH_{up}}$ represent the threshold limit for a sample to remain in the same state for the left hand peak and right hand peak respectively.

To digitize the signal, the first sample in the measurement data is assigned to the state whose mean value it is closest to. All subsequent samples are then classified according to Equation 3.1 and their values set equal to the mean value of their corresponding state.. After iterating over the entire measurement sequence, a pure RTN digitized signal will have been generated. With the digitized signal, estimation of the RTN parameters can be determined using a variety of methods. The first is by analyzing the new posterior distribution based on the digitized signal and comparing the relative amplitudes of the two Gaussians to find the state occupation ratio. The second method is to manually count how long each state's lifetime is for both up and down states, and then calculate the average lifetime based on those counts. This can be supplemented by generating a distribution of the individual lifetimes and fitting it against the geometric distribution.

3.2 Multi-level Model

While the previous method worked well for two-level measurements, the thresholding technique is insufficient for multi-level RTN. With more complex data a more systematic approach was needed, which involves up to four stages:

1. Prior Estimation
2. Digitization
3. Decomposition
4. Parameter estimation

In prior estimation we seek to identify features of the signal and provide initial estimates for signal parameters. Because white background noise is still a factor the signal is first filtered using a rolling average. While computationally simple, the transfer function of the rolling average is very similar to that of a Gaussian filter — the result is the deconvolution, or at least the reduction, of any Gaussian noise contributions from the time domain signal [54]. While very simple, it is ideal for preserving step responses and minimizing white noise fluctuations. An ideal window length can be determined by comparing the RTN Lorentzian spectrum to the filters transfer function

$$H(f) = \frac{\sin(\pi f M)}{M \sin(\pi f)} \tag{3.4}$$

where M is the window length. The window length is chosen so that the drop off of the rolling average is similar to the Lorentzian - this is so as much white noise as possible is filtered without removing RTN components.

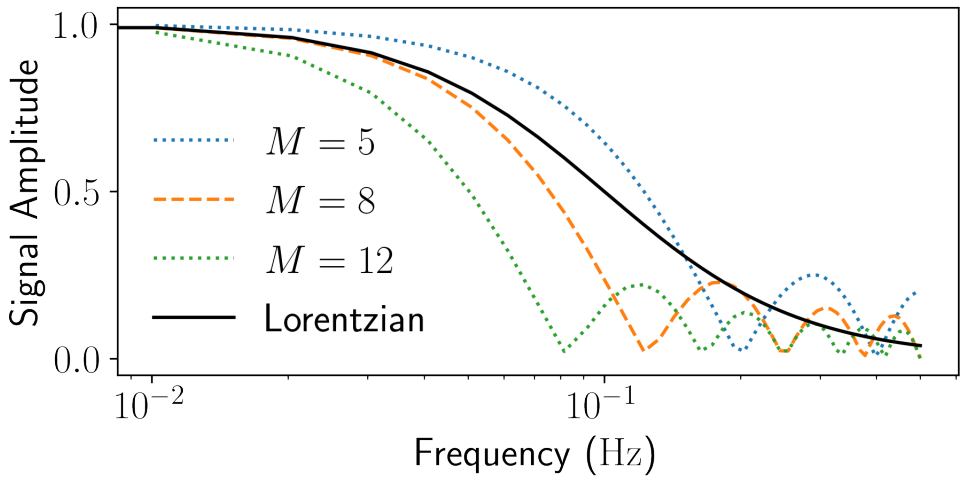


Figure 3.3: Transfer function of a rolling average filter with window size M, compared to the Lorentzian spectrum of RTN.

Following application of the rolling average to reduce white noise contributions, the density of the signal is estimated using a technique called KDE. While a traditional histogram counts the number of samples that fall into pre-defined bins, KDE creates a distribution by adding a kernel function at each point a sample falls. This has the effect of smoothing the signal which is very important in the case of multi-level RTN. If the distribution is not smooth, the chance of identifying roughness artifacts as RTN levels increases substantially. Figure 3.4 shows the difference in density estimation achieved by using KDE with Gaussian kernels in comparison to a standard histogram. The cumulative effect achieved by processing a measurement signal using these techniques is shown in Figure 3.5.

The location and variance of each peak is estimated by fitting the estimated signal density to a multi-Gaussian distribution. The number of traps N is estimated from the number of peaks n_{peaks} using

$$N = \text{ceil}(log_2(n_{peaks})) \tag{3.5}$$

where ceil means to round up to the next highest integer. However, unlike in the two-level system, where one peak corresponded to the distribution of one state, in the multi trap case one peak can correspond to the states of multiple traps. This means a further step is needed. The time-lag plot, which is commonly used in prior estimation, provides a clear

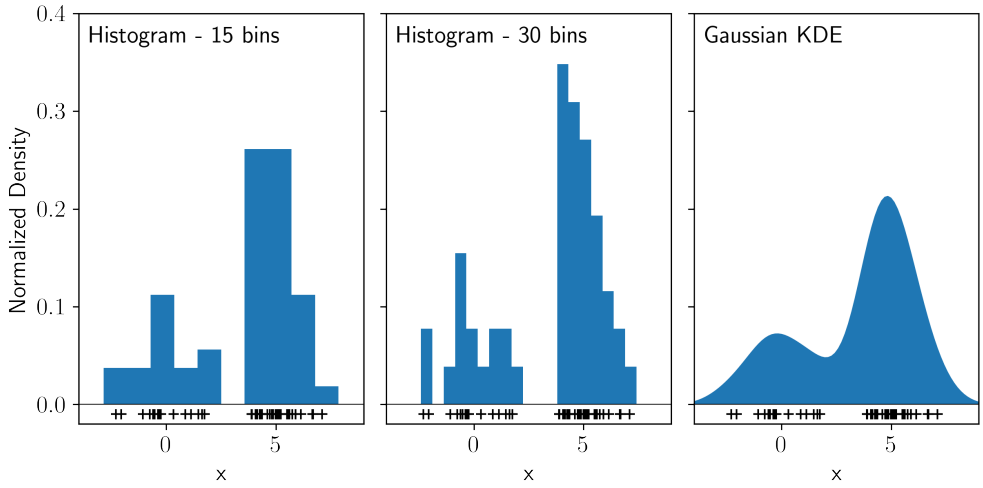


Figure 3.4: Estimated density for the same data found using a 15-bin histogram, a 30-bin histogram, and Gaussian KDE.

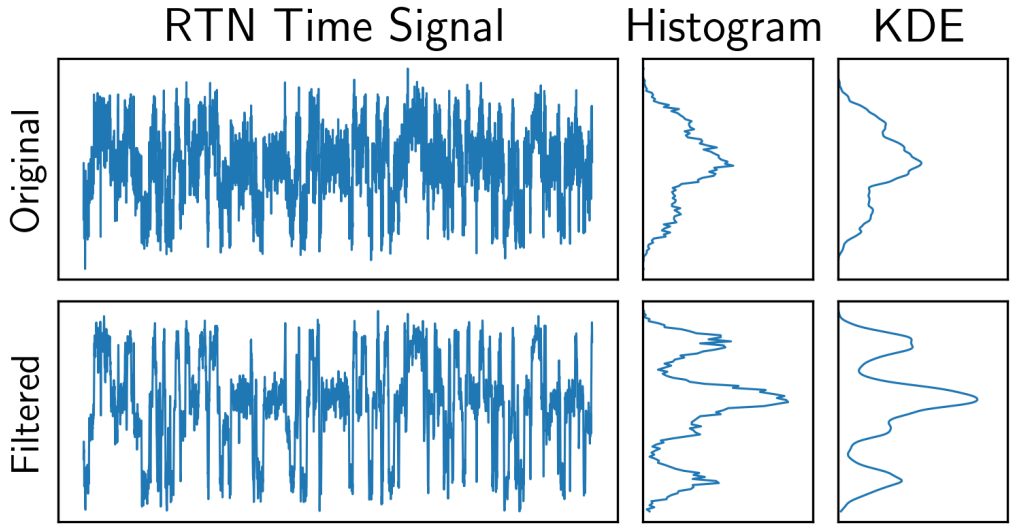


Figure 3.5: The top row shows the density of measurement values using both a histogram and KDE. The shapes are very similar, with no discernable RTN. The bottom row shows the density of values after applying a rolling average to remove white noise. Both the histogram and KDE show distinct RTN peaks, but peak finding of the histogram developed density could yield many false positives without careful constraints.

visualization that the states in RTN have high autocorrelation — when the signal is plotted against a slightly delayed copy of itself visible clusters form, as shown in Figure 3.6.

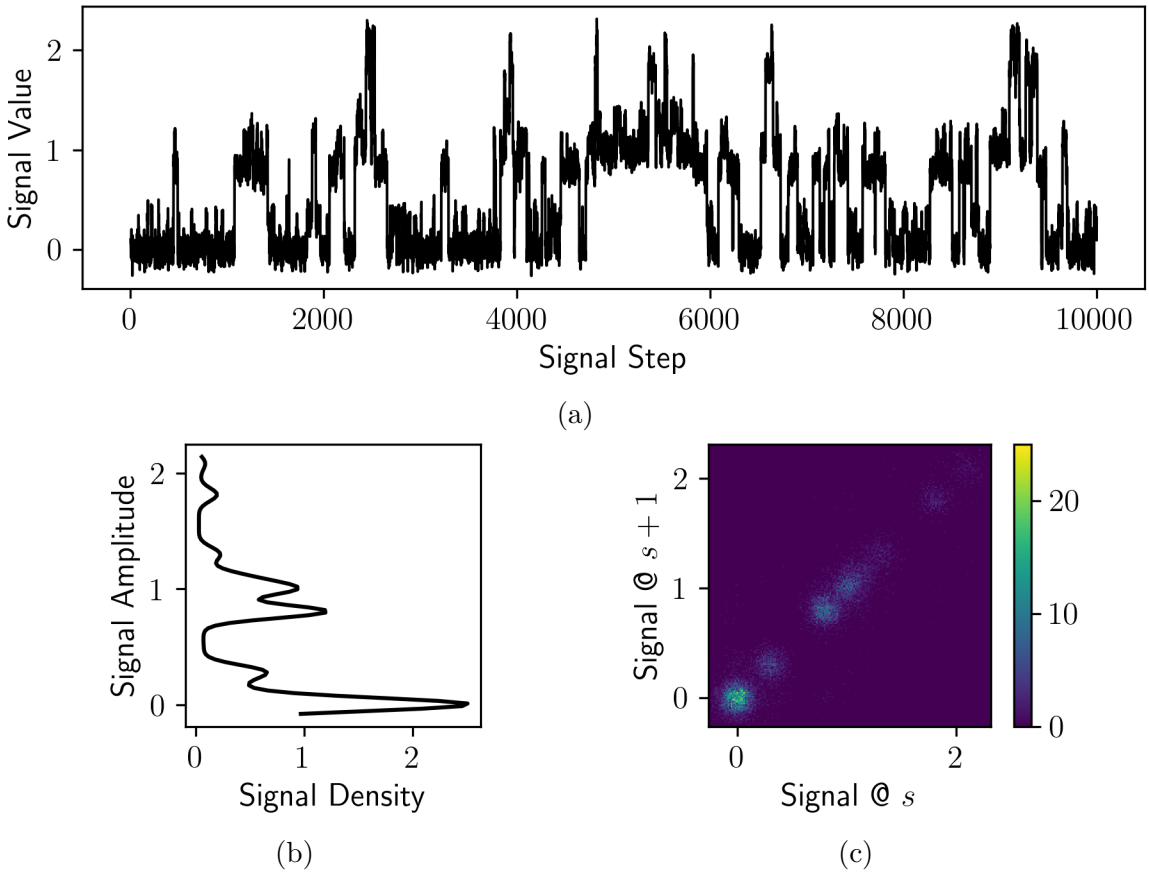


Figure 3.6: (a) Simulated multi-level RTN to demonstrate clusters formed from highly correlated states. (b) One dimensional signal density found using KDE. (c) Time-lag plot visualization of same data with a 1-step delay, plotted as a two-dimensional histogram. *Densities are not normalized to 1.

This motivated the use of clustering techniques from machine learning on the RTN data. Because the RTN states with high autocorrelation naturally form clusters, the value of the signal at time t and the value of the signal at time $t + dt$ are used as the "features" of the sample population. The most common technique for clustering is called K-means. In K-means, clusters are all treated as the same size and samples are assigned to each cluster based on proximity to the nearest cluster's average. While efficient, this is limited if the populations of one state is substantially different from the others or in cases where different

variances in the two features is required. To overcome these limitations, a two-dimensional GMM clustering technique which uses expectation-maximization was applied, using an implementation from the *scikit-learn* Python library. GMM clustering allows clusters of different sizes and can account for different variances in the Gaussian parameters for the two features of a state. While this means more free parameters are used, which increases the potential for overfitting, the results from prior estimation are used as constraints on the generated model. The result is that samples with high autocorrelation are clustered together, and assigned the mean value of their associated cluster. Figure 3.7 shows the same two-dimensional histogram of time-lag data from Figure 3.6c, plotted on a logarithmic scale to enhance contrast of the smaller peaks, alongside a flat time-lag plot with clusters marked by colour. By mapping the two-dimensional cluster data to the one dimensional time domain signal a digitized signal is generated, as shown in Figure 3.7c.

Once the signal is digitized, its PMF is described with

$$f_\xi = \sum_{i=1}^{n} \delta(c_i x) \tag{3.6}$$

where n is the number of observed levels and $\sum_{i=1}^{n} c_i = 1$. In other words, the distribution is described as a series of scaled delta functions, with the scaling coefficient c_i corresponding to the state frequency.

The next step is to decompose the multi-level signal into its independent RTN subcomponents. A convolution model is pre-defined for N RTN components, and the density of the signal is fit to the proper model constrained to the number of expected RTN components. Fitting is accomplished using a non-linear least-squares minimization algorithm implemented in the *lmfit* Python library, and redefines the signal density in terms of the GMM components for each individual RTN component. Decomposition was performed on the signal from Figure 3.6a, and the starting distribution and resulting distributions are shown in Figure 3.9.

Digitization and decomposition yields the transition amplitude ΔRTN and stationary distributions (Equation 2.35 and Equation 2.36) for any RTN subcomponents in the signal. The next chapter will focus on RTN in MOSFET devices and discuss the statistical description and parameters of RTN in the context of device physics.

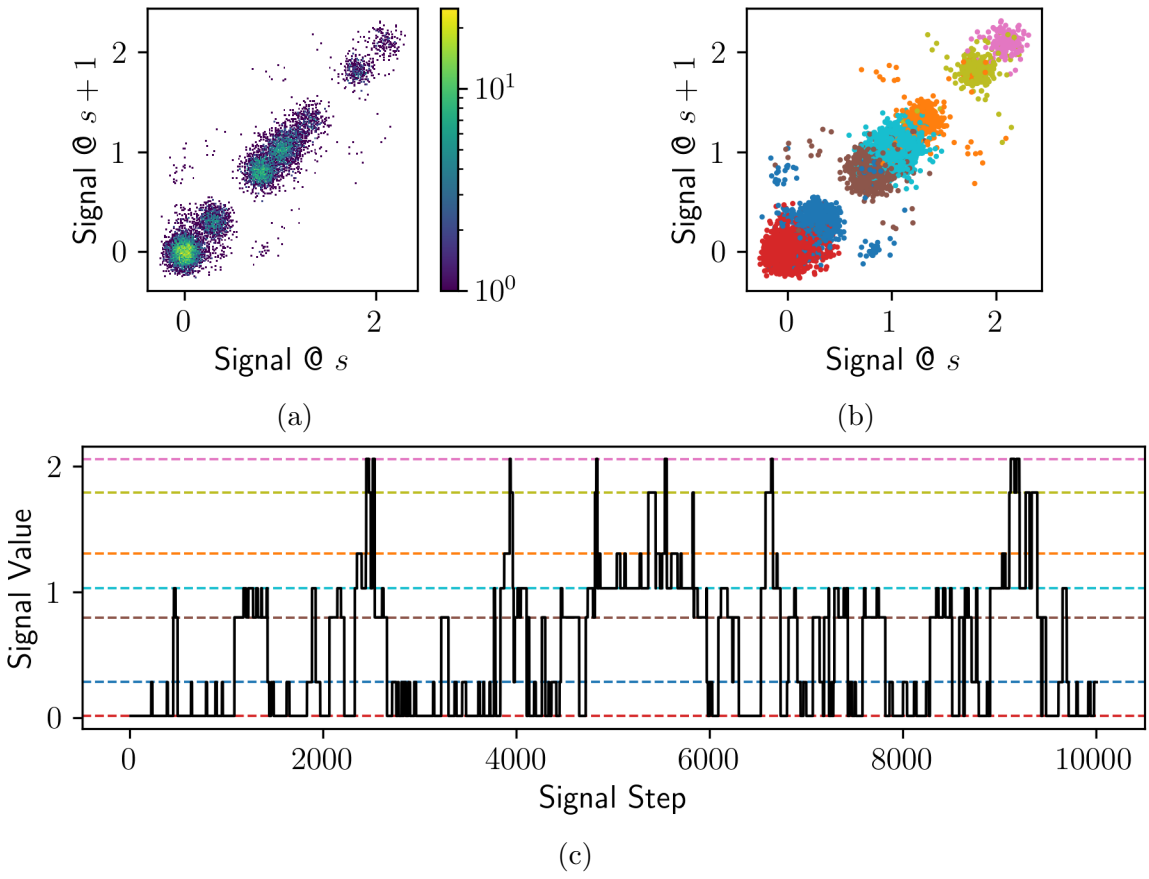


Figure 3.7: (a) Two-dimensional histogram of time-lag data from Figure 3.6, with a logarithmic colour-scale. White represents regions with zero counts and hence is undefined on the colour-scale. (b) Flat (i.e. no density component) time-lag data with colours denoting clusters. (c) Digitized signal (black) achieved by mapping the cluster means to each index point in the original signal. The dashed coloured lines represent the mean values for each associated cluster in (b).

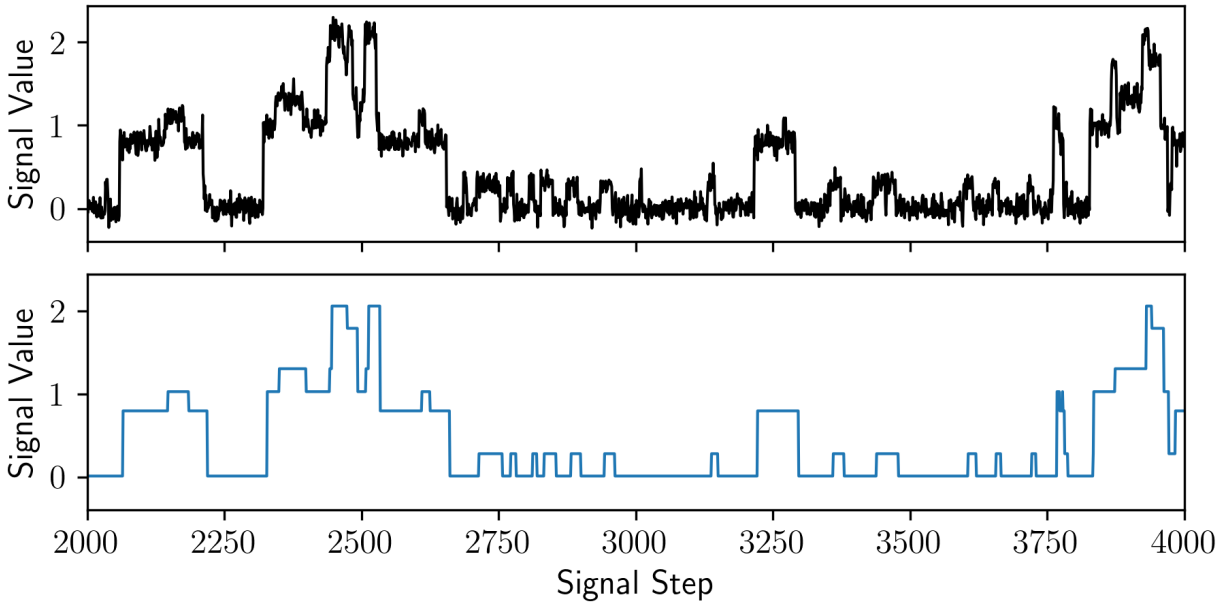


Figure 3.8: Comparing the original signal (top) against digitized signal (bottom) in a narrower range of data points.

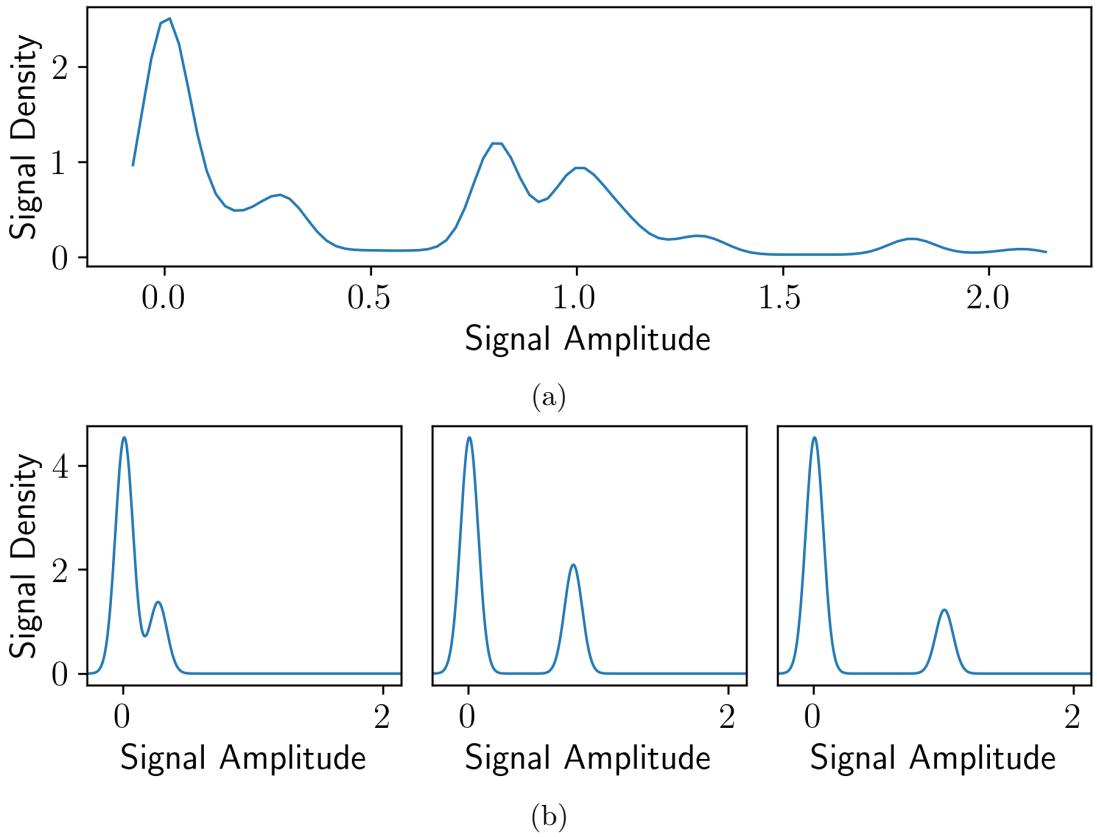


Figure 3.9: (a) Signal density distribution for the RTN signal to be decomposed. (b) Distributions for the three RTN subcomponents.

Chapter 4

Random Telegraph Noise in Devices

RTN is a fundamental noise process which is caused by localized defects in the structure of an electronic device. In this chapter, I aim to describe how RTN is a consequence of device physics and its physical origin. I will be describing it specifically in the context of MOSFETs. Although it affects more devices, MOSFETs are well-understood and relatively common. In the process, I will also describe the origins of other noise sources which are common in electronic devices.

4.1 MOSFETs

A MOSFET is a switching transistor with three terminals, the gate, source, and drain.[1] Two fundamental types of MOSFETs exist: the NMOS and the PMOS, defined by the majority charge carrier when the MOSFET is turned on[2]. The structure for both devices is identical, except for doping types:

- A moderately doped body, p-type for NMOS and n-type for PMOS.
- Heavily doped drain and source wells implanted into the body:
  - n^+-type for NMOS, and
  - p^+-type for PMOS.

[1] Technically the MOSFET body is a fourth terminal, but is often shorted to the source in practice.

[2] The majority charge carrier for an NMOS device is the electron, while for a PMOS device it is the hole.

- A vertically stacked gate structure on the substrate surface consisting of a dielectric barrier and conducting material. Traditionally, the dielectric is an oxide like SiO_2 and the conductor was a metal[3], hence the term MOS.

For both NMOS and PMOS the application of a gate voltage, denoted V_G or V_{GS}, will create a conducting channel between the drain and source wells, as shown in Figure 4.1 for an NMOS. The current amplitude is determined by the applied drain voltage V_{DS}.

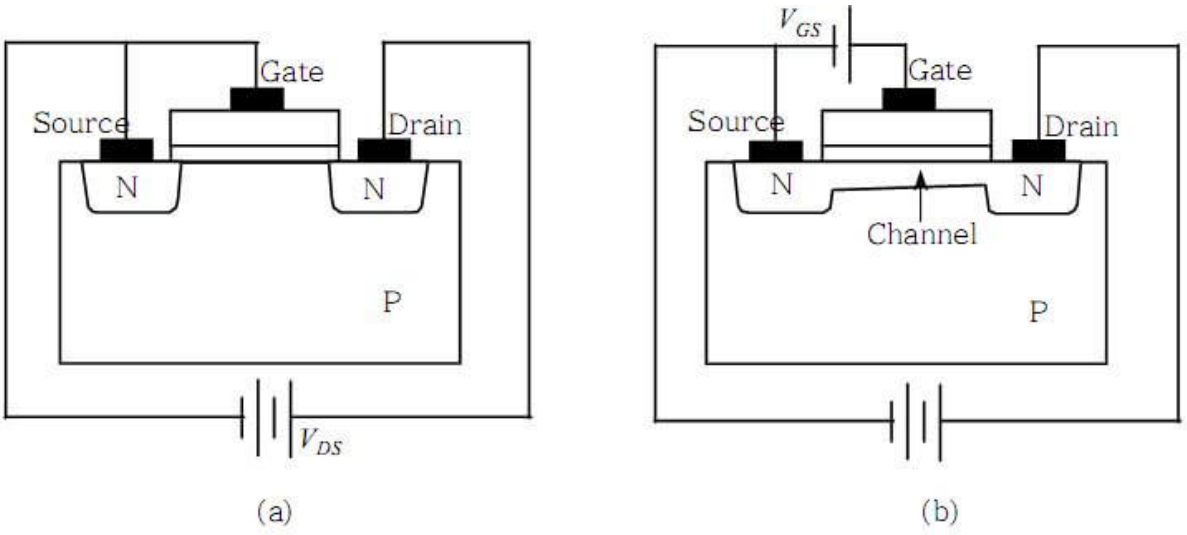


Figure 4.1: Diagram of an NMOS device. In (a) the device is switched off. In (b) an applied gate voltage V_{GS} forms a channel between the drain and source. At this point, an applied voltage V_{DS} will cause current to flow through the channel. [28].

More accurately, a channel is only formed when V_{GS} is beyond a certain threshold value V_{th} — when this is achieved, the MOSFET is said to be in inversion mode. If this threshold value is not reached the MOSFET may be in either accumulation or depletion mode, depending on whether the MOS structure is forward or reverse biased. Exact conditions for each mode are given in Table 4.1, with charge densities visualized for an NMOS device are shown in Figure 4.2. Modes are described as follows:

Accumulation When the applied gate voltage is the wrong polarity to turn the MOSFET on. This causes an accumulation of charge carriers of the type defined by the body dopant.

[3]Modern devices use a polysilicon material.

Depletion When the applied gate voltage begins to cause free carriers in the substrate to be repelled from region near the body-oxide interface.

Inversion When the amplitude of the applied voltage attracts minority carriers to the body-oxide interface, inverting charge in a localized region and forming a conducting channel between the drain and source.

Table 4.1: Zones of operation for NMOS and PMOS devices.

Zone	NMOS	PMOS
Accumulation	$V_{GS} \leq 0$	$0 \leq V_{GS}$
Depletion	$0 < V_{GS} \leq V_{th}$	$V_{th} \leq V_{GS} < 0$
Inversion	$0 < V_{th} < V_{GS}$	$V_{GS} < V_{th} < 0$

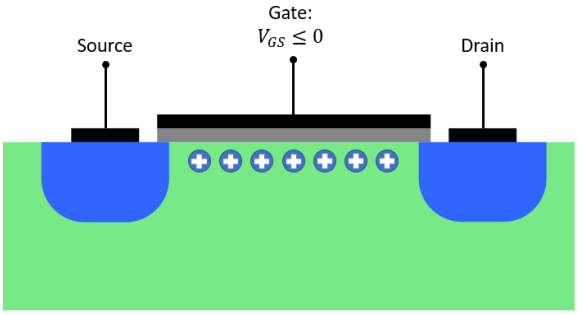
4.2 RTN in MOSFETs

As we have established a statistical model for describing and analyzing RTN, we must establish a physical description of RTN in devices. It has been shown that defects in the oxide layer near the substrate interface are the source of RTN in MOSFETs[44, 4]. The defects act as charge traps, which channel carriers can tunnel to and occupy for a characteristic time, as shown in Figure 4.3. In addition it is possible for multiple independent traps to exist in the channel. Interactions may occur, but this requires trap separation which is less than the Fermi screening length of the traps [32]. Otherwise, the traps may independently contribute their own RTN signature to the overall signal, and the result will appear to have multiple levels, as shown in Figure 4.4.

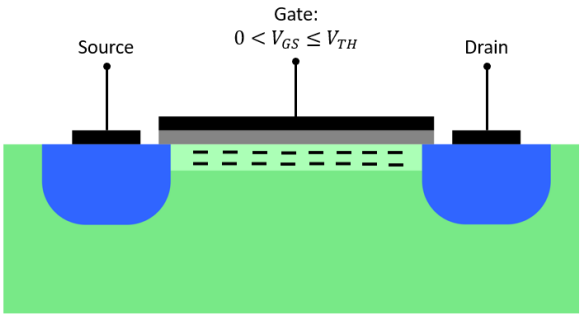
Following [25], we can describe the defect statistics using the grand partition function, as the defect can exchange energy and particles with the channel reservoir:

$$Z_G = \sum_{i,N} \exp\left(-\frac{E_i - NE_F}{k_B T}\right) \tag{4.1}$$

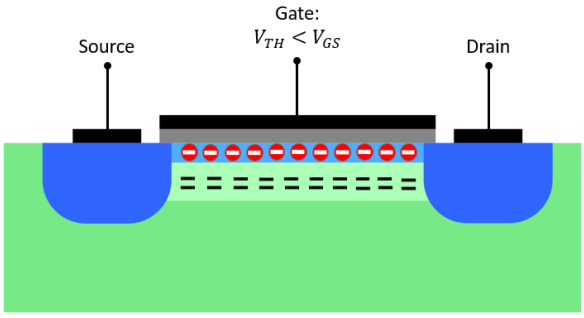
where the sum over i, N represents all combinations of states and particle numbers, E_i is the energy for each state, and E_F is the Fermi energy equivalent to the chemical potential



(a) NMOS charge distribution in accumulation.



(b) NMOS charge distribution in depletion.



(c) NMOS charge distribution in inversion.

Figure 4.2: Charge distributions for an NMOS device in (a) accumulation, (b) depletion, and (c) inversion.

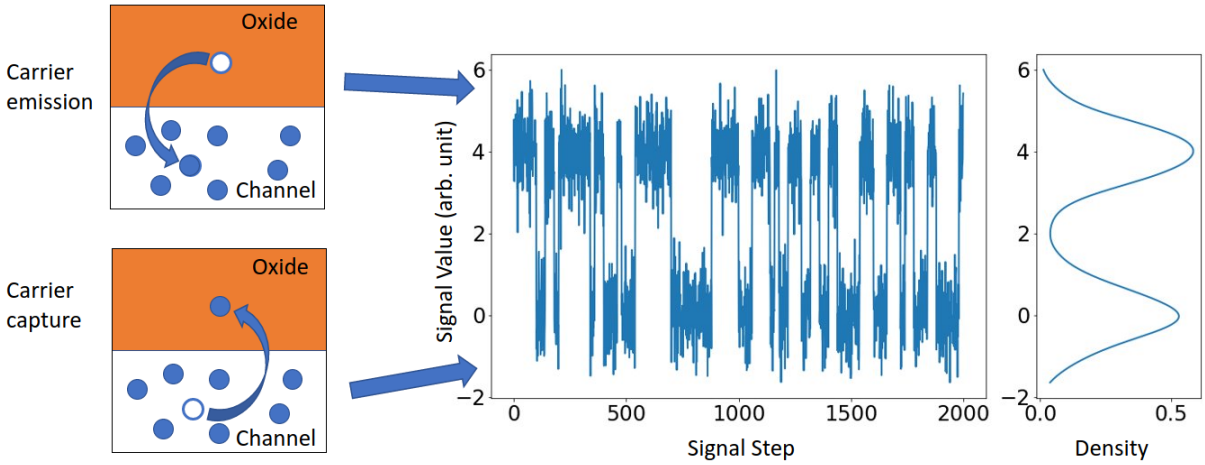


Figure 4.3: A single defect in the oxide trigger discrete fluctations in current though the capture and emission of carriers. The distribution of current will reflect the average value of each current state.

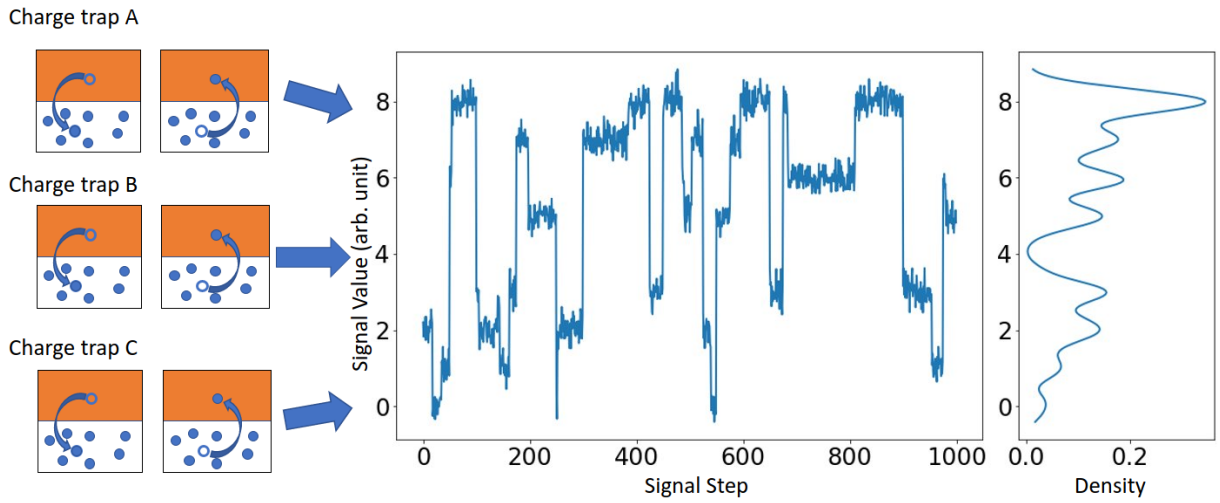


Figure 4.4: Simulated multi-level RTN as a result of three independent charge traps.

of the material. If we consider a defect that can be either unoccupied ($N = 0$) or occupied ($N = 1$), then the unoccupied state has energy $E_0 = 0$ and E_T represents the energy level where the defect enters the occupied state. Then Equation 4.1 becomes

$$Z_G = \gamma_0 \exp\left(-\frac{E_0 - N_0 E_F}{kT}\right) + \gamma_T \exp\left(-\frac{E_T - N_T E_F}{kT}\right) \tag{4.2}$$

$$= \gamma_0 + \gamma_T \exp\left(-\frac{E_T - E_F}{kT}\right) \tag{4.3}$$

where γ represents the degeneracy of the state. Furthermore, the probability that the defect will be observed in state i is

$$p(i) = \frac{\gamma_i \exp[-(E_i - N_i E_F)/kT]}{Z_G}. \tag{4.4}$$

From Equation 4.3 and Equation 4.4, the probability the defect is occupied is

$$p(T) = \frac{\gamma_T \exp[-(E_T - E_F)/kT]}{\gamma_0 + \gamma_T \exp[-(E_T - E_F)/kT]} \tag{4.5}$$

$$= \frac{1}{1 + \gamma_0/(\gamma_T \exp[-(E_T - E_F)/kT])}$$

$$= \left[1 + g \exp\left(\frac{E_T - E_F}{kT}\right)\right]^{-1} \tag{4.6}$$

where $g = \gamma_0/\gamma_T$. The result derived using the grand partition function can be related to the RTN time constants through the stationary distribution Equation 2.36, giving

$$p(T) = \frac{\tau_c}{\tau_c + \tau_e} = \left[1 + g \exp\left(\frac{E_T - E_F}{kT}\right)\right]^{-1}, \tag{4.7}$$

where τ_c and τ_e are the characteristic times for the occupied state and unoccupied states respectively.[4] If we let $g = 1$ to show orbital degeneracy is a non-factor in state probability, we can rewrite Equation 4.7 as

$$\ln \frac{\tau_c}{\tau_e} = -\frac{E_T - E_F}{kT} \tag{4.8}$$

[4]The result of Equation 4.7 is slightly different from the equivalent in Kirton & M. Uren [25]: in that work, and others, the stationary distribution for the occupied state is defined as $\tau_e/(\tau_c + \tau_e)$. However, this is contrary to their earlier definition of the time constants and of the stationary distribution given in other works, such as Machlup [35].

Using the MOSFET energy levels at flat-band and under a bias, we can express $E_T - E_F$

$$E_T - E_F = (E_{C_{ox}} - E_T) - (E_C - E_F) - \phi_0 + q\psi_s + q\frac{x_T}{T_{ox}}(V_{GS} - V_{FB} - \psi_s), \tag{4.9}$$

where $E_{C_{ox}}$ is the conduction band energy of the oxide, E_C is the conduction band energy of silicon, ϕ_0 is the conduction band offset between the oxide and silicon, $q\psi_s$ is the degree of band bending in silicon, and $q\frac{x_T}{T_{ox}}(V_{GS} - V_{FB} - \psi_s)$ is the oxide band offset of the gate voltage, flat-band voltage, and band bending[5] respectively. The last term is scaled by the depth x_T of the defect into the oxide with thickness T_{ox}. The defects location is given in the energy diagram, shown in Figure 4.5.

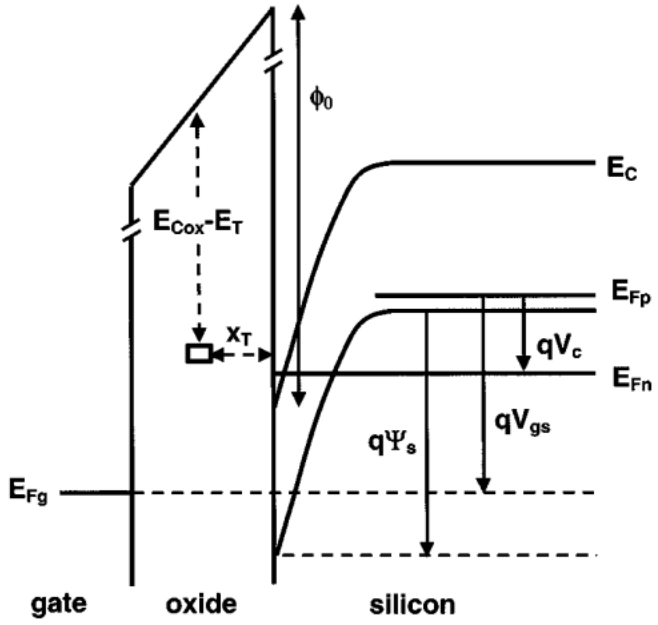


Figure 4.5: Band bending in MOSFET inversion mode causes the Fermi level to approach that of the trap which increases the likelihood of a carrier tunneling through the x_T oxide distance [6].

Equation 4.8 can be differentiated with respect to V_{GS} to investigate the effect of gate

[5]Some bending may occur due to trapped charge

voltage on trap time constants

$$\frac{d\ln(\tau_c/\tau_e)}{dV_{GS}} = -\frac{E_T - E_F}{kT}, \tag{4.10}$$

$$\frac{d\ln(\tau_c/\tau_e)}{dV_{GS}} = -\frac{q}{kT}\frac{x_T}{T_{ox}}, \tag{4.11}$$

which provides a technique for estimating trap depth.

One limitation of this derivation is that it assumes constant energy levels along the drain-source direction. Celik-Butler *et al.* [6] derived an extended model which accounts for differences in band structure at either end of the channel region (i.e. near the source and near the drain). This requires both forward and reverse drain-source biasing until a maximum τ_c/τ_e is found in both directions,

$$\ln\left(\frac{(\tau_c/\tau_e)_{f_{max}}}{(\tau_c/\tau_e)_{r_{max}}}\right) = \frac{q}{k_B T}\frac{x_T}{T_{ox}}\left[\frac{y_T}{L}V_{DS_{f_{max}}} - \left(1 - \frac{y_T}{L}\right)V_{DS_{r_{max}}}\right] \tag{4.12}$$

where y_T is the trap distance from the source, L is the channel length, $(\tau_c/\tau_e)_{f_{max}}$, $(\tau_c/\tau_e)_{r_{max}}$ are the maximum τ_c/τ_e ratios in both directions, and $V_{DS_{f_{max}}}$, $V_{DS_{r_{max}}}$ are the corresponding drain-source voltages. A complete diagram of the defects location within the MOSFET is given in Figure 4.6.

It was shown in subsection 2.3.1 that the PSD of an RTN signal is a Lorentzian, with amplitude and cut-off defined by the RTN parameters:

$$S_x(f) = \frac{1}{\pi}\frac{(\Delta I)^2\tau_0\tau_1}{(\tau_0+\tau_1)^2}\frac{1}{1+\left(\frac{2\pi f\tau_0\tau_1}{\tau_0+\tau_1}\right)^2} + S_x(0) \tag{4.13}$$

where $\tau_a = (1/\tau_c + 1/\tau_e)^{-1}$.

The RTN is inversely proportional to f^2 so if other noise sources are low enough it will be visible at low frequencies, as shown in Figure 4.7.

4.3 Other Noise Sources in MOSFETs

4.3.1 White Noise

White noise is the common name for noise sources which have no frequency dependence - the noise power is distributed evenly across the entire spectrum of frequencies. In electronic

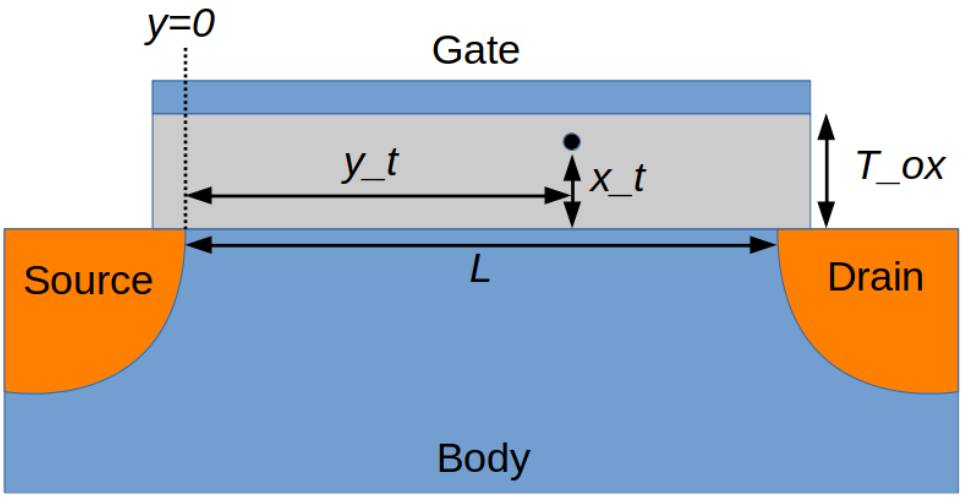


Figure 4.6: Given a MOSFET with gate oxide thickness T_{ox} and channel length L, we set the spatial origin $(x, y) = (0, 0)$ to be at the intersection of the source, body, and oxide, with y in the direction of current flow and x into the oxide. The charge trap is located a distance of x_t into the oxide, and a distance y_t towards the drain.

devices, white noise can be caused by either Johnson-Nyquist noise (also called thermal noise) or shot noise.

Johnson-Nyquist noise is a result of the Brownian motion of the electrons in a crystal from thermal energy in an equilibrium state.[6] Johnson-Nyquist noise has no dependence on actual device bias and is always present except at 0 K. As a temperature of 0 K is impossible to reach, it represents a minimum bound on device noise. The PSD of Johnson-Nyquist noise is

$$S_x(f) = 4k_B T R \tag{4.14}$$

where k_B is Boltzmann's constant, T is absolute temperature, and R is the impedance of the device.

In comparison, shot noise is a result of the discretization of current as charge carriers. Although the average current through a point may remain macroscopically constant, in reality there are fluctuations in the total number of carriers passing through that point over time. A larger average current means a wider ranger of fluctuations, and if there is no net current flow than these fluctuations cannot occur. Shot noise only depends on device

[6] At equilibrium the average temperature of the solid is constant. Brownian motion may cause localized fluctuations, but these are random and average out.

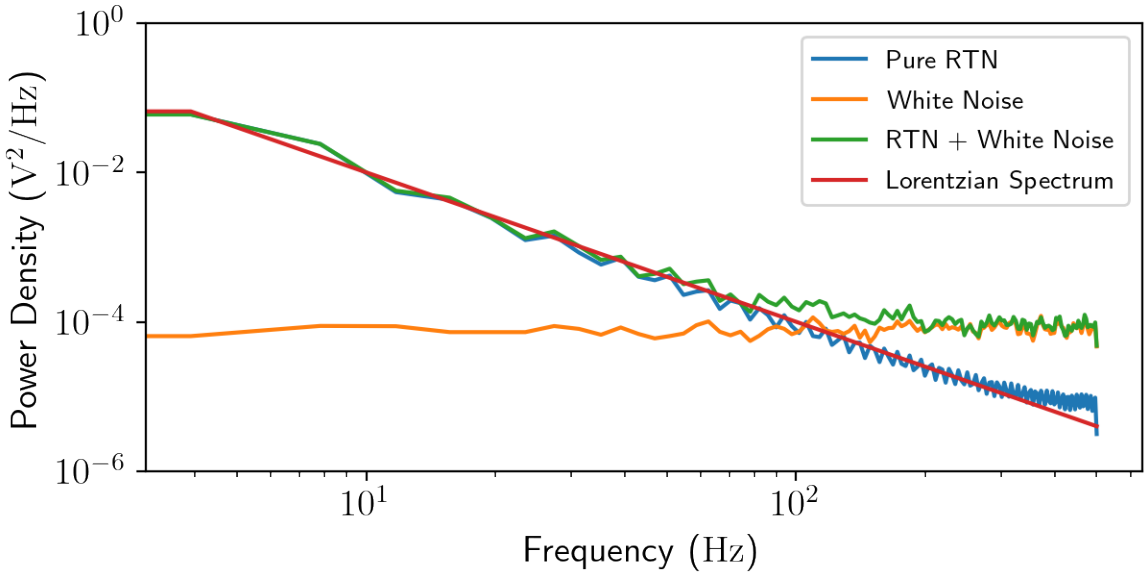


Figure 4.7: A pure RTN signal (blue) will follow a Lorentzian power spectrum (red), as in Equation 4.13. However, in measurements there are other noise sources present in the signal. At low frequencies the RTN may dominate the power spectrum but as it drops off with f^2 other noise sources will overtake it, as shown here. The total signal spectrum (green) follows the white noise spectrum (orange) at higher frequencies.

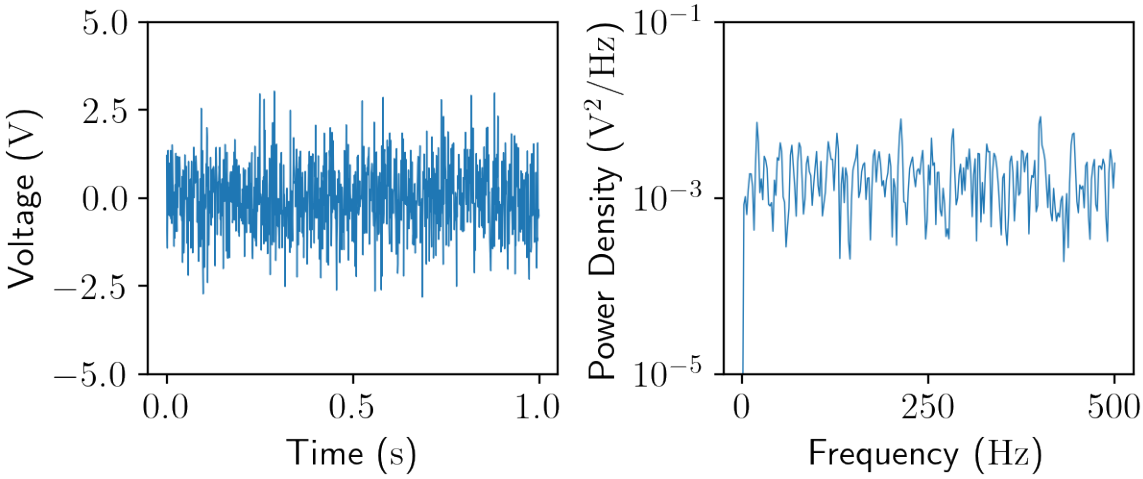


Figure 4.8: Left: A time domain white noise signal generated with a 1 kHz sampling frequency. Right: The corresponding PSD. As there are no DC components to the signal, the PSD falls off very rapidly towards $f = 0$ Hz, but elsewhere the spectrum frequency independent.

operation, expressed in the PSD formula:

$$S_x(f) = 2qI \tag{4.15}$$

where q is the elementary charge and I is the average current.

4.3.2 Flicker Noise

Flicker noise, sometimes referred to as pink noise, is the dominant low-frequency noise source. Its PSD varies with $1/f^\alpha$, where α is very close to 1. For this reason, it is often referred to as $1/f$ noise in academic literature.

Flicker, or $1/f$, noise was originally described empirically, with no adequate physical model proposed. In the 1950s and 1960s however, two models gained prominence [34]. The first, proposed by McWhorter in 1957, suggests that $1/f$ is caused by trapping of significant numbers of carriers, where the time constants τ of each trapping event are themselves distributed following $1/\tau$. This model suggests that fluctuations in the number of carriers is the source of $1/f$ noise.

The second, proposed by Hooge in 1969, is similar to McWhorter's in that it suggests the trapping of carriers by oxide defects leads to $1/f$. However, rather than the fluctu-

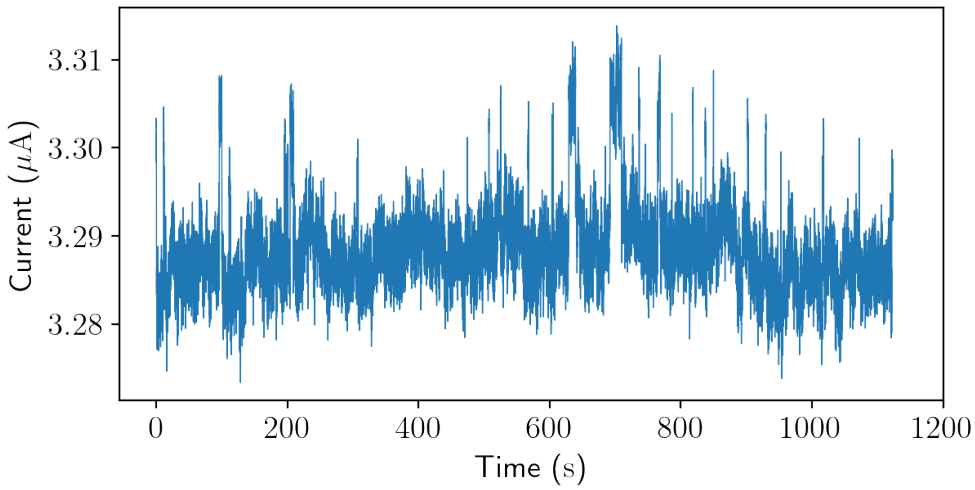


Figure 4.9: $1/f$ noise can be difficult to identify in the time domain. Current data from measurement of carbon nanotube devices by HeeBong Yang.

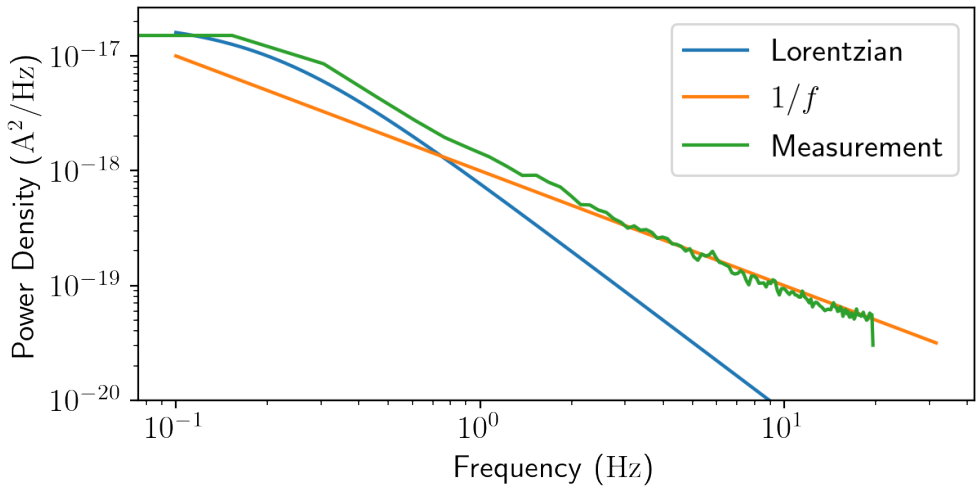


Figure 4.10: In the frequency domain, $1/f$ is much easier to identify. The Lorentzian spectrum of RTN falls off much quicker, while white noise does not exhibit the same frequency dependence as $1/f$. In this measurement, RTN components are visible at the lowest frequencies.

1

ation of carrier numbers being the source, it suggests that trapping causes variations in carrier mobilities. The model posits that it is mobility fluctuations, rather than number fluctuations, that causes $1/f$.

In either case, both models suggest that $1/f$ is the summation of many individual RTN processes, a suggestion that has been corroborated theoretically and empirically by others such as Howard [17].

Chapter 5

Device Measurement and Analysis

This research was motivated by a desire to characterize CMOS devices operating at cryogenic temperatures, with the goal of assessing RTN dependence on operational temperature or device bias levels. Measurements were carried out on 28 nm PMOS devices using an ARS 4 K dry cryostat and an HP 4156A parameter analyzer — as such, measurements were limited to 10,001 points. To achieve a high density of test coverage the parameter analyzer was automated using a Python script, which sets the optimal measurement resolution, programmed the sampling biases, and pulled and saved the measurement data. I then used the algorithm to analyze these data.

The device used in testing was a low threshold PMOS device integrated in a TSMC 28 nm CMOS chip, with width $W = 1$ μm and a length of $L = 35$ nm.[1] There were many individual devices on the chip, but many exhibited failure from the start or during the cooldown process. Individual devices were first functionally tested using an Everbeing probe station – if the device was working, it was wirebonded to a customized chip carrier using 30 μm gold wire. The device was tested once again after mounting in the cryostat chamber. If still functional, the cryostat was sealed and lowered to its base temperature - over the course of these measurements that was approximately 5.5 K. In addition to the 5.5 K base temperature, higher temperatures were set using a metallic heater and a Lakeshore 335 temperature controller. A combination of factors such as device heating and measurement duration (spanning over a week of constant measuring) resulted in slight temperature fluctuations over time - as such, the temperatures presented are the *set* temperatures rather than measured. Sampling intervals were set at 5 ms and 20 ms, as previous experiment results showed resolution artifacts at shorter sampling intervals and

[1] It is unknown whether these are effective dimensions or actual dimensions.

fewer RTN samples above that. Finally, every combination of measurement parameters was repeated for three measurements to increase statistical accuracy.

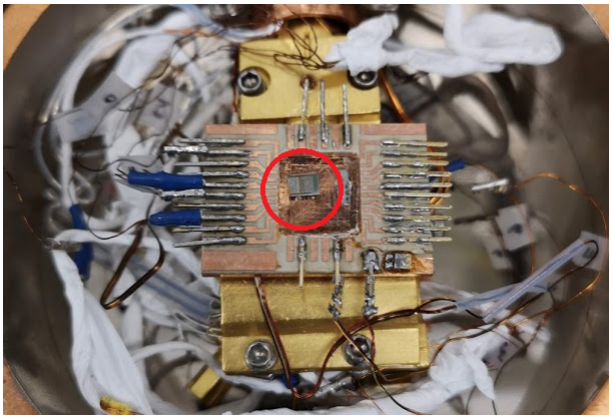


Figure 5.1: CMOS chip (circled in red) wirebonded to custom chip carrier, which in turn is mounted on the cryostat cold finger.

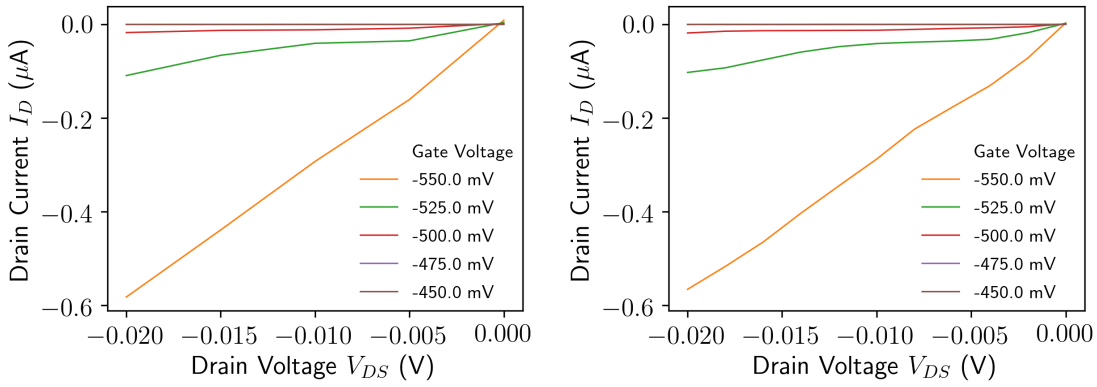
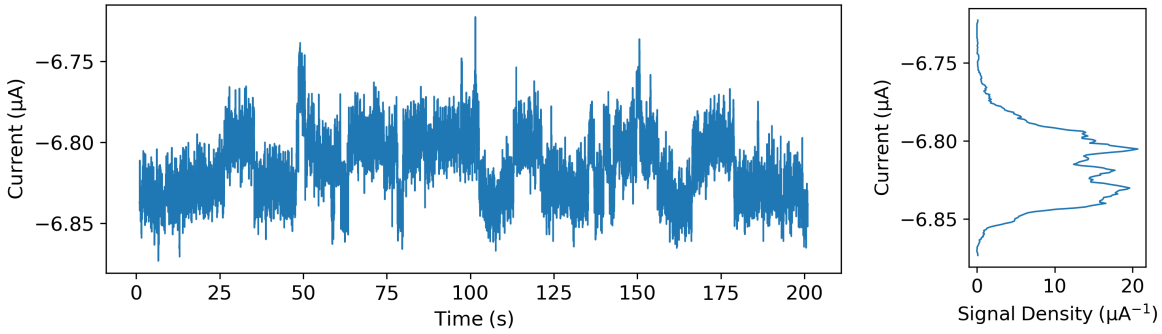


Figure 5.2: (a) MOSFET current-voltage (I-V) curve at room temperature, operating in linear mode. (b) IV curve at 5.5 K. Despite change in temperature, MOSFET linear region is still active.

5.1 Example Analysis

In this section I will present a step-by-step visualization of application of the algorithm to measurement data. The presented data were measured at 5.5 K at a gate voltage of −565 mV and drain voltage of −25 mV. In addition, the sampling time was set at 20 ms. First, the measurement itself and current distribution is shown in Figure 5.3a and Figure 5.3b respectively.



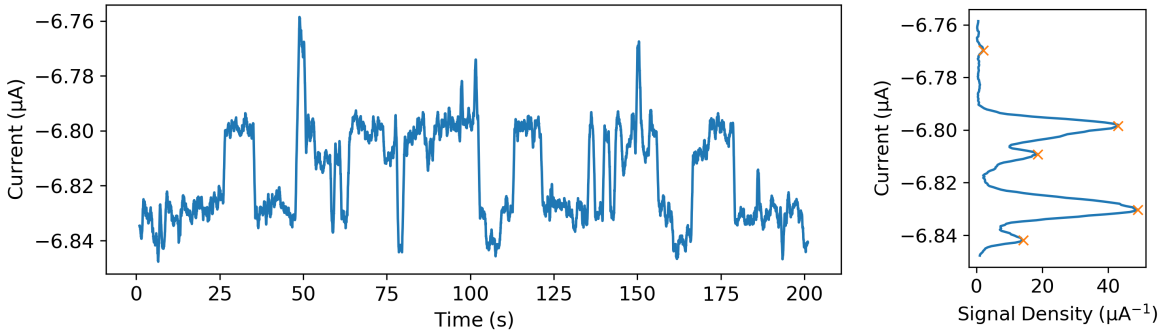
(a) (b)

Figure 5.3: (a) Unprocessed measurement signal for device measured at 5.5 K. (b) Estimated density distribution.

The filtered data and estimated density estimation resulting from prior analysis are shown in Figure 5.4a and Figure 5.4b respectively as well. The peaks representing RTN states are marked with orange x-marks.

The output of clustering and digitization is overlaid over the filtered data in Figure 5.5. The estimated distribution for the corresponding convolution of mixture models is shown in Figure 5.6a, and the decomposed distributions are shown in Figure 5.6b and Figure 5.6c.

Finally, the estimated power spectral density for the digitized signal in Figure 5.5 is shown in Figure 5.7, along with the calculated Lorentzian spectrums corresponding to each of the RTN components.



(a) (b)

Figure 5.4: (a) Measurement signal after application of rolling average filter to minimize normally-distributed fluctuations. (b) Estimated density distribution of filtered data, with peaks corresponding to identified states marked with orange checks.

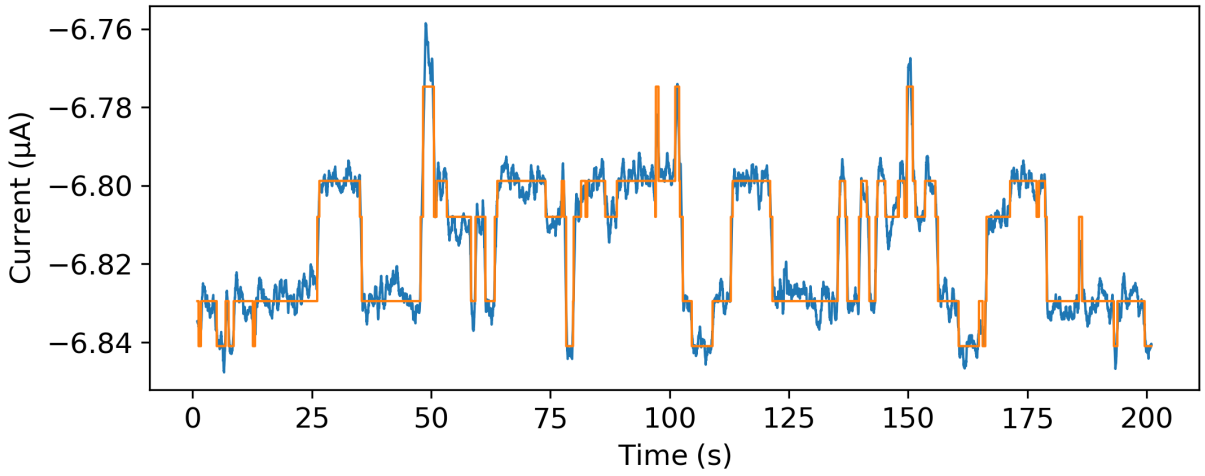
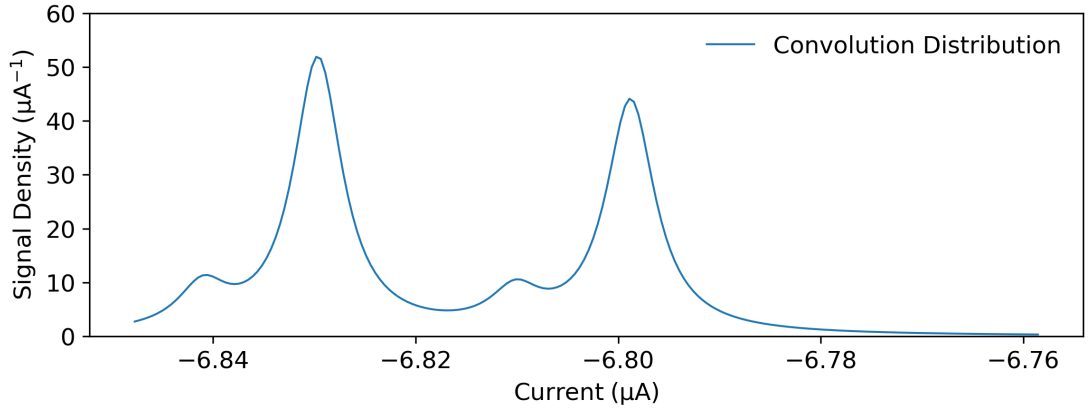


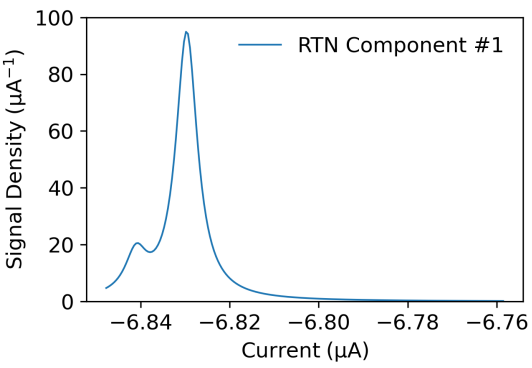
Figure 5.5: Digitized signal (orange) overlaid filter measurement signal (blue).

Table 5.1: Extracted trap amplitude and time constant parameters for each component in the signal.

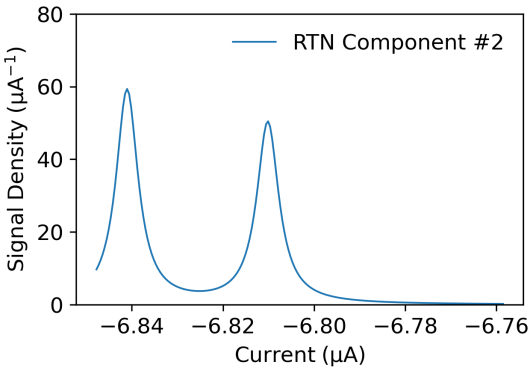
	ΔI(nA)	τ_c(ms)	τ_e(ms)
Component #1	11.3	149.4	964.8
Component #2	30.8	526.7	446.6



(a)



(b)



(c)

Figure 5.6: (a) Estimated density distribution formed by the convolution of two Gaussian mixture model components. (b) Density distribution of the first subcomponent. (c) Density distribution of the second subcomponent.

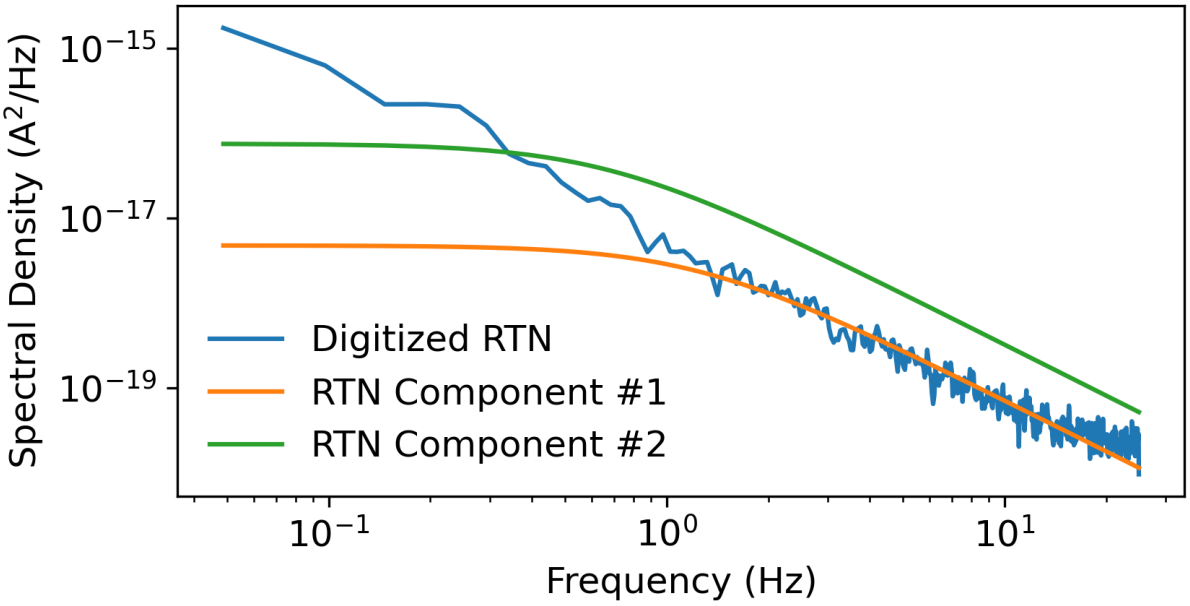


Figure 5.7: Comparing estimated power spectral density of digitized signal (estimated using Welch's method) against the subcomponent spectrums calculated using a Lorentzian spectrum and extracted parameters.

5.2 Gate Dependence Sorted by Temperature

As gate dependence of RTN was shown in Equation 4.8, this was the focus of measurement analysis. RTN dependence on temperature was explored at the same time, by analyzing each measurement set independently — the sets of measurement conditions are given in Table 5.2. For each temperature, the first step is exploring the number of traps. Next, the RTN amplitude for each trap is characterized along with its relation to the average signal current. Finally, the trap depth x_T/T_{ox} is investigated based on Equation 4.8.

Table 5.2: Parameters used to measure drain current through the transistors channel.

Test Set	Temp (K)	$\lvert V_G \rvert$ (mV)	$\lvert V_D \rvert$ (mV)
A	~ 5.5	480 - 620	5 - 30
B	~ 14	480 - 600	5 - 30
C	~ 100	480 - 600	5 - 25

5.2.1 5.5 Kelvin Data

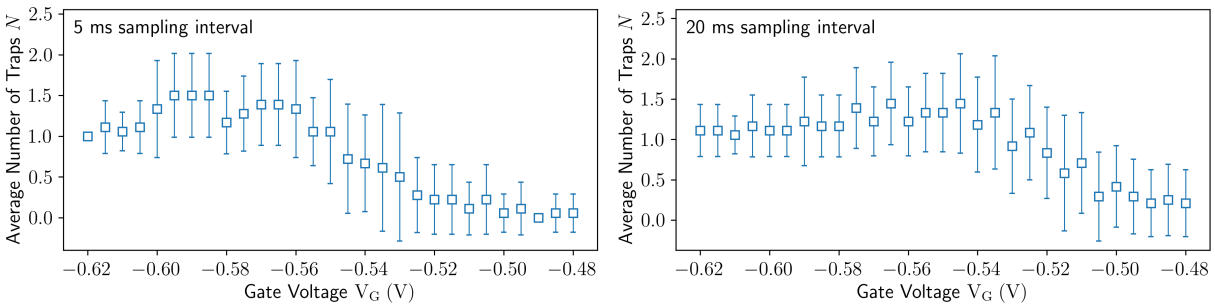


Figure 5.8: Average number of traps at 5.5 K. As the gate voltage is increased, more traps become active until a saturation point is reached.

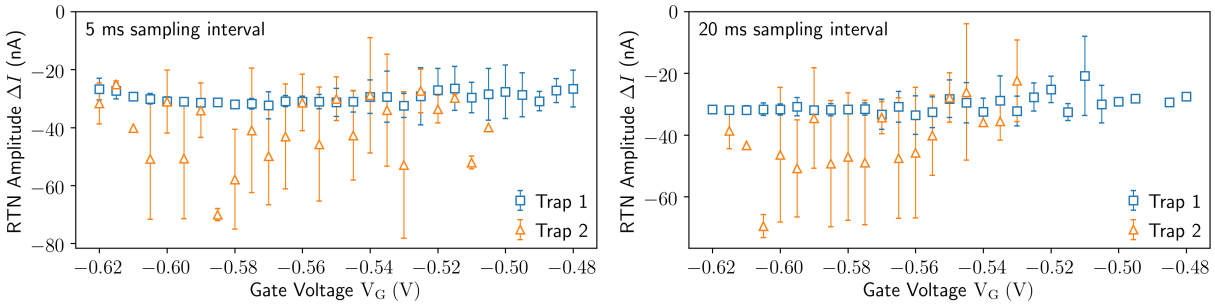


Figure 5.9: Average amplitude ΔI of the observed traps. The amplitude of the primary trap remains relatively constant across gate voltage, but there are large fluctuations in amplitude of the secondary traps.

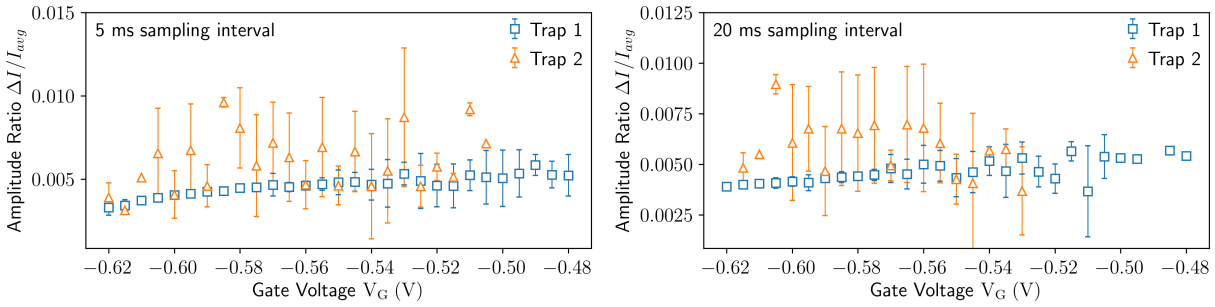


Figure 5.10: Average amplitude of the observed trap relative to the mean signal value $\Delta I/I_{avg}$. The primary trap ratio slightly decreases with increasing gate voltage, which corresponds to increasing I_{avg}. The secondary trap ratio does not appear to have the same relationship.

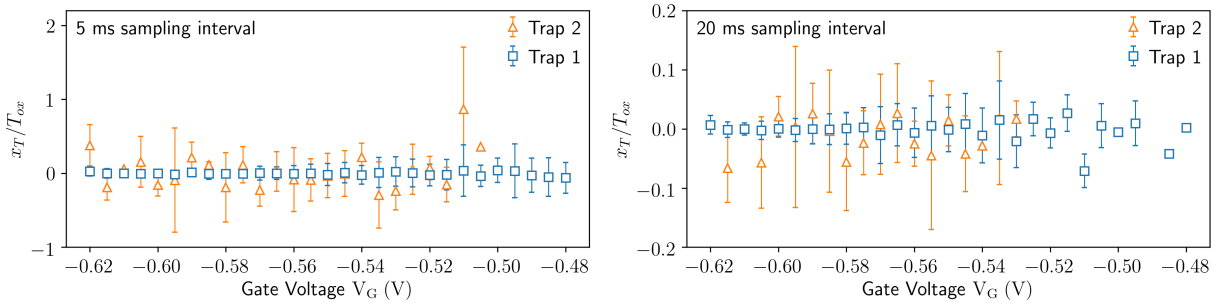


Figure 5.11: Trap distance x_T into the oxide (thickness T_{ox}) as a percentage. From the percentages, the trap depth x_T appears quite shallow – the depth is negative at some gate voltages however, which suggests these traps are not in the oxide, but rather in the channel itself. An alternate explanation is that these traps have the opposite polarity. This possibility will be discussed later in this chapter.

5.2.2 14 Kelvin Data

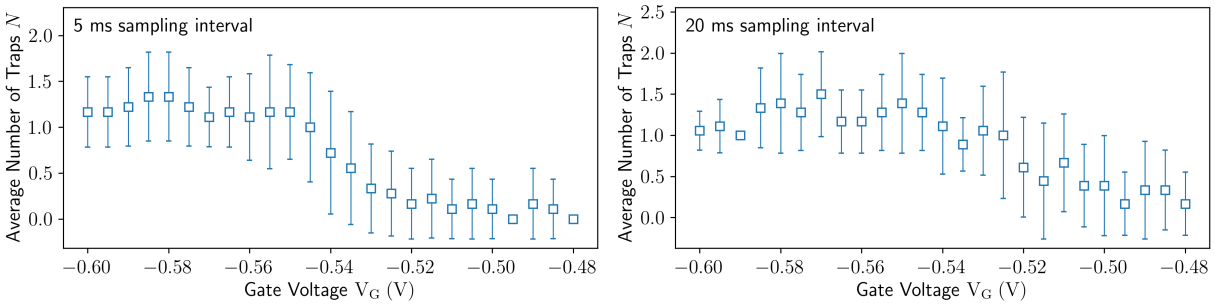


Figure 5.12: Average number of traps at 14 K. The behaviour is quite similar to Figure 5.8, where increasing gate voltage leads to more activated traps up to a saturation level.

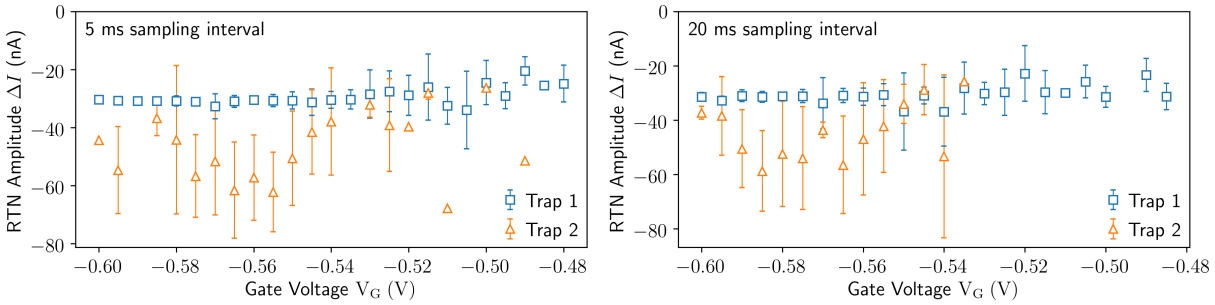


Figure 5.13: Average amplitude ΔI of the observed traps. The primary trap once again behaves similarly to the 5.5 K data — a strong gate voltage dependence is seen on the secondary trap however.

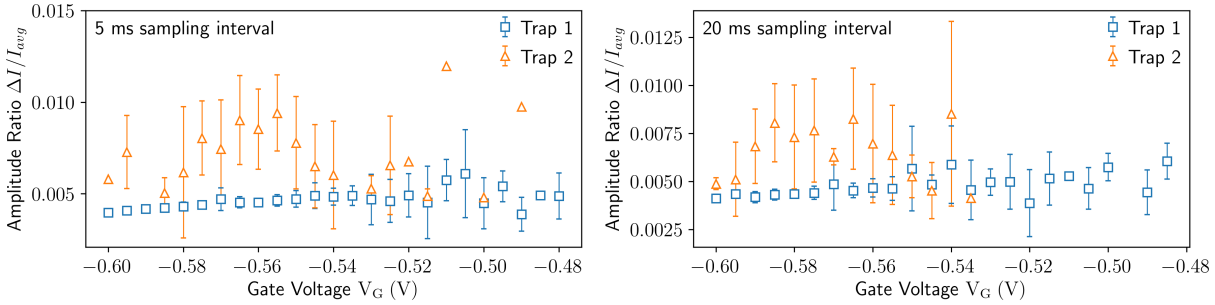


Figure 5.14: Average amplitude of the observed trap relative to the mean signal value $\Delta I/I_{avg}$. The behaviour does not deviate unexpectedly when taking it as a proportion average current.

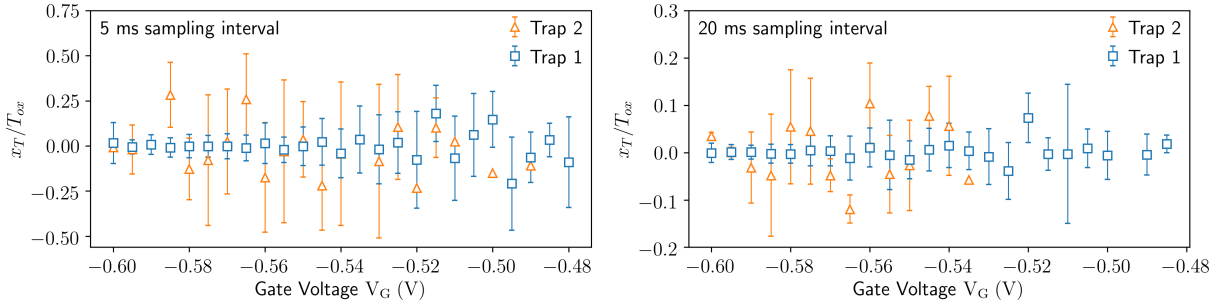


Figure 5.15: Trap distance x_T into the oxide (thickness T_{ox}) as a percentage. Trap behaviour once again does not substantially change from the 5.5 K data.

5.2.3 100 Kelvin Data

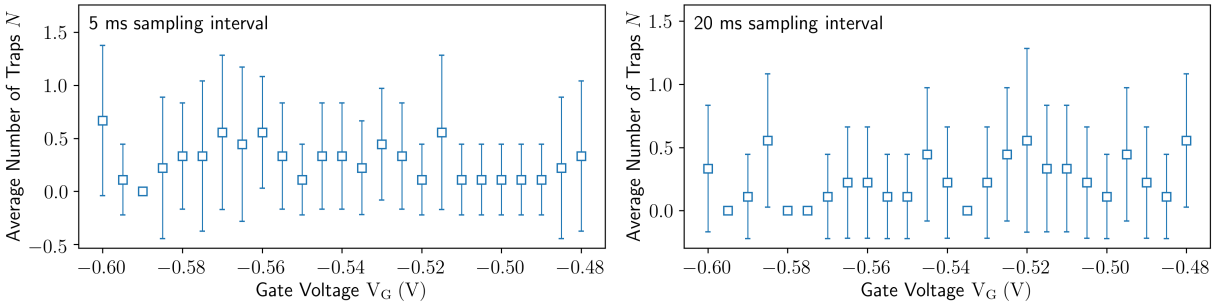


Figure 5.16: Average number of traps at 100 K. Far fewer traps are activated or observed when increasing the gate voltage then at the lower temperatures, which may suggest thermal noise is masking them.

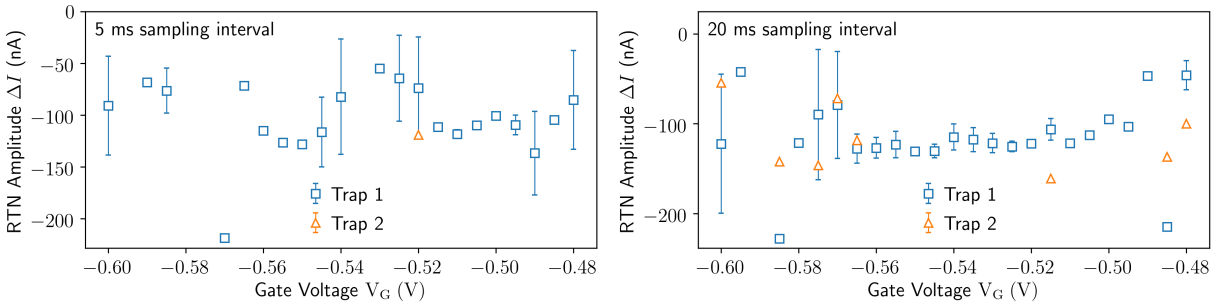


Figure 5.17: Average amplitude ΔI of the observed traps. The trap amplitude has much greater fluctuations at 100 K and has a larger overall amplitude than the 5.5 K and 14 K results.

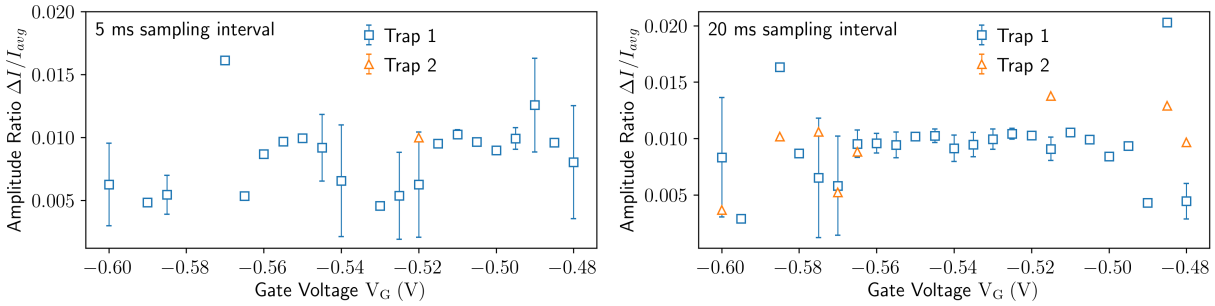


Figure 5.18: Average amplitude of the observed trap relative to the mean signal value $\Delta I/I_{avg}$. The trap amplitude does not appear to have increased or decreased relative to average current when compared to the 5.5 K and 14 K data.

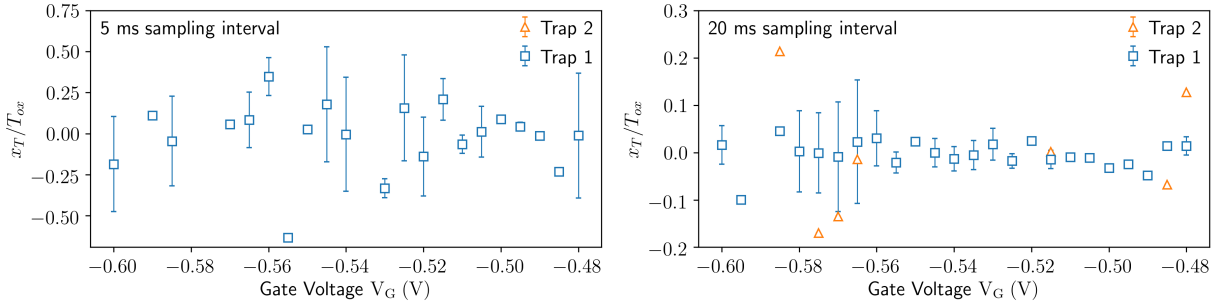


Figure 5.19: Trap distance x_T into the oxide (thickness T_{ox}) as a percentage. Percent depth of traps again fluctuates around zero at shallow levels.

5.3 Discussion

These results show there is a strong temperature dependence of some form on the observation of RTN in device measurements. The 5.5 K and 14 K measurements yielded very similar results for the number of traps, trap amplitude, trap depth, and gate voltage dependence. A marked difference was seen in the results of the 100 K measurements, with less consistent behaviour along gate bias, and only a handful of multi-level RTN signals observed. It is possible that this is caused by an increase in thermal noise, which masks any RTN. However thermal noise should be independent of gate voltage, and RTN is actually observed less at higher bias levels — it should be consistent if thermal noise is the cause. This suggests that there is another thermal effect at play, possibly on charge mobility or carrier density. This would be the next area to explore in device characterization using RTN.

Observations of the trap depth x_T/T_{ox} show both positive and negative values for trap depth. Defects causing RTN in MOSFETs have been shown to exist in the oxide[25], which would correspond to a positive depth ratio — a negative value implies a defect located in the channel. This may be a possibility, although the derivation for the x_T/T_{ox} relation requires a trap energy inside the defect. Another possibility, suggested by Ralls *et al.* [44], is for defects of opposite polarity to exist. One defect might be neutral to start and become charged following carrier capture — this is the model used to define the RTN parameters in terms of current. Another type may be charged to start and be neutralized following carrier capture — the same capture mechanics would apply, however the definition of τ_c and τ_e would apply for changes in current of the opposite polarity.

For both 5.5 K and 14 K, relatively consistent behaviour is observed across the gate voltage sweep in terms of RTN amplitude ΔI and trap depth x_T/T_{ox} — at least for the primary trap (labelled "Trap 1"). There is a larger variation in analysis results for secondary traps, both in terms of mean value and the standard deviations of results — this is clearest in the trap amplitude plots for both temperatures. This suggests the possibility that the primary trap is the same physical defect while the second trap may be caused by one of a few defects, activated at random by fluctuations in bias, carrier density, temperature, etc. Another stage of this research may focus on discerning any consistency or repetition in the secondary trap data, in an attempt to identify another unique trap.

Chapter 6

Conclusion

We demonstrated the origin of RTN in semiconducting devices as caused by crystal defects. These defects introduce an energy state that a charge carrier is able to occupy, which causes modulation in channel parameters. When observed, these modulations cause discrete transitions in channel current, voltage, or resistance. The lifetime of each discrete transition is a random variable which is either geometrically or exponentially distributed, depending on whether it is in discrete or continuous time. We can derive this result by modelling the signal as a Markov process, which assumes the signal is a memoryless process whose state can transition at any point. By calculating the autocorrelation function and then applying the Wiener-Khinchin theorem we can likewise calculate the power spectrum of the RTN signal. This can be verified experimentally, which offers validation of the statistical model.

While this type of frequency domain analysis is well established, time domain analysis has typically be restricted to signals with only two-discrete levels. As multi-level RTN is shown to be common especially at cryogenic temperatures, we developed an algorithmic technique for characterizing multi-level RTN. This algorithm uses a two-dimensional clustering technique using Gaussian mixture models to identify the highly auto-correlated RTN levels and digitize it. Following digitization, a constrained curve fitting technique decomposes the signal into the individual RTN components, which allows parameters corresponding to each trap to be identified.

More rigorous examination of the trap properties could not be accomplished due to multiple factors. However, these factors are external limitations and not a result of a weakness in the algorithm. The algorithm provides the necessary information to identify the depth of multiple traps into the oxide and also the distance of the traps along the channel length. Finally, the amplitude corresponding to each trap is extracted and can

be further used for analysis. The demonstrated research has not engaged in modelling or explaining the effects of operating parameters on traps, but rather enabled more rigorous examination than existing techniques allow. The next steps in this research would delve deeper into the trap models to explain the observed dependence on temperature and gate voltage.

Bibliography

1. Amir, G. *Continuous Time Markov Chains* http://u.math.biu.ac.il/~amirgi/CTMCnotes.pdf (2019).
2. Awano, H., Tsutsui, H., Ochi, H. & Sato, T. *Multi-trap RTN parameter extraction based on Bayesian inference* in *International Symposium on Quality Electronic Design (ISQED)* (IEEE, Mar. 2013).
3. Basnet, P. *METAL OXIDE PHOTOCATALYTIC NANOSTRUCTURES FABRICATED BY DYNAMIC SHADOWING GROWTH* PhD thesis (Apr. 2015).
4. Bollu, M., Koch, F., Madenach, A. & Scholz, J. Electrical switching and noise spectrum of Si-SiO2 interface defects generated by hot electrons. *Applied Surface Science* **30,** 142–147 (Oct. 1987).
5. Bury, K. V. *Statistical Models in Applied Science* ISBN: 9780471125907 (Wiley, 1975).
6. Celik-Butler, Z., Vasina, P. & Vibhavie Amarasinghe, N. A method for locating the position of oxide traps responsible for random telegraph signals in submicron MOSFETs. eng. *IEEE transactions on electron devices* **47,** 646–648. ISSN: 0018-9383 (2000).
7. Chau, R. *et al.* Application of high-κ gate dielectrics and metal gate electrodes to enable silicon and non-silicon logic nanotechnology. *Microelectronic Engineering* **80,** 1–6 (June 2005).
8. Cleve, R., Ekert, A., Macchiavello, C. & Mosca, M. Quantum algorithms revisited. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* **454,** 339–354 (Jan. 1998).
9. Colombeau, B. *et al.* Advanced CMOS devices: Challenges and implant solutions. *physica status solidi (a)* **211,** 101–108 (Dec. 2013).
10. Couch, L. W. *Digital and Analog Communication Systems* 8th ed. ISBN: 9780132915380 (Pearson, 2013).

11. Dongaonkar, S., Giles, M. D., Kornfeld, A., Grossnickle, B. & Yoon, J. *Random telegraph noise (RTN) in 14nm logic technology: High volume data extraction and analysis* in *2016 IEEE Symposium on VLSI Technology* (IEEE, June 2016).

12. Eroms, J. *et al.* Low-frequency noise in Josephson junctions for superconducting qubits. *Applied Physics Letters* **89,** 122516 (Sept. 2006).

13. Ghahramani, Z. in *Series in Machine Perception and Artificial Intelligence* 9–41 (World Scientific, June 2001).

14. Green, M. L., Gusev, E. P., Degraeve, R. & Garfunkel, E. L. Ultrathin (< 4 nm) SiO2 and Si–O–N gate dielectric layers for silicon microelectronics: Understanding the processing, structure, and physical and electrical limits. *Journal of Applied Physics* **90,** 2057–2121 (Sept. 2001).

15. Guo, S., Wang, R., Mao, D., Wang, Y. & Huang, R. Anomalous random telegraph noise in nanoscale transistors as direct evidence of two metastable states of oxide traps. *Scientific Reports* **7** (July 2017).

16. Harlingen, D. J. V. *et al.* Decoherence in Josephson-junction qubits due to critical-current fluctuations. *Physical Review B* **70** (Aug. 2004).

17. Howard, R. M. *A modified random telegraph signal with a 1/f PSD* in *2017 International Conference on Noise and Fluctuations (ICNF)* (IEEE, June 2017).

18. Howard, R. M. *Principles of Random Signal Analysis and Low Noise Design* (John Wiley & Sons, Inc., July 2002).

19. Iosifescu, M. *Stochastic processes and applications in biology and medicine* ISBN: 354006270X (Editura AcademieiSpringer-Verlag, Bucureşti,New York, 1973).

20. Irwin, J. O. The Frequency Distribution of the Difference between Two Independent Variates following the same Poisson Distribution. *Journal of the Royal Statistical Society* **100,** 415. ISSN: 09528385. https://www.jstor.org/stable/10.2307/2980526?origin=crossref (2019) (1937).

21. James, D. *Intel Ivy Bridge unveiled - The first commercial tri-gate, high-k, metal-gate CPU* in *Proceedings of the IEEE 2012 Custom Integrated Circuits Conference* (IEEE, Sept. 2012).

22. Jayaraman, R. & Sodini, C. A 1/f noise technique to extract the oxide trap density near the conduction band edge of silicon. *IEEE Transactions on Electron Devices* **36,** 1773–1782 (Sept. 1989).

23. Kandiah, K., Deighton, M. O. & Whiting, F. B. A physical model for random telegraph signal currents in semiconductor devices. *Journal of Applied Physics* **66,** 937–948 (July 1989).

24. Kasap, S. *Principles of Electronic Materials and Devices* 3rd ed. ISBN: 0073104647 (McGraw-Hill, Inc., USA, 2005).

25. Kirton, M. & Uren, M. Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise. *Advances in Physics* **38,** 367–468 (Jan. 1989).

26. Kittel, C. *Introduction to Solid State Physics* 8th ed. ISBN: 9780471415268 (Wiley, 2004).

27. Knill, E. Quantum computing. *Nature* **463,** 441–443 (Jan. 2010).

28. Kumar, A. *MOSFET Transistor: Easiest Basics, Operation and Types* Feb. 2016. https://analyseameter.com/2016/02/mosfet-operation-types-symbols.html.

29. Li, M., Wang, R., Zou, J. & Huang, R. Characterization of Random Telegraph Noise in Scaled High- /Metal-Gate MOSFETs with SiO2/HfO2 Gate Dielectrics. *ECS Transactions* **52,** 941–946 (Mar. 2013).

30. Li, Z. *et al.* Random telegraph noise from resonant tunnelling at low temperatures. *Scientific Reports* **8** (Jan. 2018).

31. Liu, B. *et al.* Semiconductor Solid-Solution Nanostructures: Synthesis, Property Tailoring, and Applications. *Small* **13,** 1701998 (Sept. 2017).

32. Liu, F. & Wang, K. L. Correlated Random Telegraph Signal and Low-Frequency Noise in Carbon Nanotube Transistors. *Nano Letters* **8,** 147–151 (Jan. 2008).

33. Liu, F. *et al.* Giant random telegraph signals in the carbon nanotubes as a single defect probe. *Applied Physics Letters* **86,** 163102 (Apr. 2005).

34. Lundberg, K. H. *Noise Sources in Bulk CMOS* 2002.

35. Machlup, S. Noise in Semiconductors: Spectrum of a Two-Parameter Random Signal. *Journal of Applied Physics* **25,** 341–343. ISSN: 0021-8979, 1089-7550. http://aip.scitation.org/doi/10.1063/1.1721637 (2019) (Mar. 1954).

36. Mahato, N., Ansari, M. & Cho, M. H. in (July 2015).

37. Mistry, K. *et al. A 45nm Logic Technology with High-k+Metal Gate Transistors, Strained Silicon, 9 Cu Interconnect Layers, 193nm Dry Patterning, and 100% Pb-free Packaging* in *2007 IEEE International Electron Devices Meeting* (IEEE, Dec. 2007).

38. Montanaro, A. Quantum algorithms: an overview. *npj Quantum Information* **2** (Jan. 2016).

39. Moore, G. E. Cramming more components onto integrated circuits. *Electronics* (1965).

40. Nagumo, T., Takeuchi, K., Hase, T. & Hayashi, Y. *Statistical characterization of trap position, energy, amplitude and time constants by RTN measurement of multiple individual traps* in *2010 International Electron Devices Meeting* (IEEE, Dec. 2010).

41. Parzen, E. *Modern Probability Theory and Its Applications* (John Wiley & Sons, Inc., 1967).

42. Patra, B. *et al.* Cryo-CMOS Circuits and Systems for Quantum Computing Applications. *IEEE Journal of Solid-State Circuits* **53,** 309–321 (Jan. 2018).

43. Prati, E., Fanciulli, M., Calderoni, A., Ferrari, G. & Sampietro, M. Microwave irradiation effects on random telegraph signal in a MOSFET. *Physics Letters A* **370,** 491–493 (Oct. 2007).

44. Ralls, K. S. *et al.* Discrete Resistance Switching in Submicrometer Silicon Inversion Layers: Individual Interface Traps and Low-Frequency (1f?) Noise. *Physical Review Letters* **52,** 228–231 (Jan. 1984).

45. Realov, S. & Shepard, K. L. Analysis of Random Telegraph Noise in 45-nm CMOS Using On-Chip Characterization System. *IEEE Transactions on Electron Devices* **60,** 1716–1722. ISSN: 0018-9383, 1557-9646. http://ieeexplore.ieee.org/document/6502225/ (2018) (May 2013).

46. Realov, S. & Shepard, K. L. Analysis of Random Telegraph Noise in 45-nm CMOS Using On-Chip Characterization System. *IEEE Transactions on Electron Devices* **60,** 1716–1722 (May 2013).

47. Realov, S. & Shepard, K. L. *Random telegraph noise in 45-nm CMOS: Analysis using an on-chip test and measurement system* in *2010 International Electron Devices Meeting* (IEEE, Dec. 2010).

48. Rogers, C. T. & Buhrman, R. A. Composition of 1f Noise in Metal-Insulator-Metal Tunnel Junctions. *Physical Review Letters* **53,** 1272–1275 (Sept. 1984).

49. Rozanov, Y. & Silverman, R. A. *Probability theory: a concise course* Rev. English ed. 148 pp. ISBN: 978-0-486-63544-6 (Dover Publications, New York, 1977).

50. Sedra, A. S. & Smith, K. C. *Microelectronic circuits* Seventh edition. ISBN: 9780199339136 (Oxford University Press, 2015).

51. Simoen, E. *et al.* Random telegraph noise: The key to single defect studies in nano-devices. *Thin Solid Films* **613,** 2–5 (Aug. 2016).

52. Simoen, E. & Claeys, C. *Random Telegraph Signals in Semiconductor Devices* (IOP Publishing, 2017).

53. *Skellam distribution* Wikipedia. https://en.wikipedia.org/w/index.php?title=Skellam_distribution (2019).

54. Smith, S. W. *The scientist and engineer's guide to digital signal processing* ISBN: 9780966017670 (California Technical Pub., 1999).

55. Tega, N. *et al. Increasing threshold voltage variation due to random telegraph noise in FETs as gate lengths scale to 20 nm* in *2009 Symposium on VLSI Technology* (2009), 50–51.

56. Troudi, M., Bergaoui, Y., Bondavalli, P. & Sghaier, N. Time domain analysis of traps generated random telegraph signal in (SWCNT) based sensors. *Sensors and Actuators A: Physical* **252,** 185–189 (Dec. 2016).

57. Uren, M. J., Day, D. J. & Kirton, M. J. 1/fand random telegraph noise in silicon metal-oxide-semiconductor field-effect transistors. *Applied Physics Letters* **47,** 1195–1197 (Dec. 1985).

58. Viswanathan, C. *Low-temperature characterization of CMOS devices* in *International Symposium on VLSI Technology, Systems and Applications* (IEEE).

59. Weisstein, E. W. *Modified Bessel Function of the First Kind* http://mathworld.wolfram.com/ModifiedBesselFunctionoftheFirstKind.html (2019).

60. Yuzhelevski, Y., Yuzhelevski, M. & Jung, G. Random telegraph noise analysis in time domain. *Review of Scientific Instruments* **71,** 1681–1688 (Apr. 2000).

Appendix A

Device Physics

A.1 Conductivity in Solids

Electrons are fermions, which means they obey the Pauli exclusion principle - fermions must have unique quantum numbers and cannot occupy the same quantum state. A single atom has well defined quantum states that an electron can occupy, determined by the electrons energy (n), orbital angular momentum (l), angular momentum vector or projection (m_l), and electron spin (s). The quantum numbers n determines the electrons shell, l determines the subshell, m_l the *orbital* an electron occupies, and finally s is a parameterization to obey the exclusion principal. These numbers are sufficient to describe the quantum states of a single atom, but when two like atoms are brought together to the point that their orbitals overlap, there will not be enough quantum states for each electron. To allow for this overlap to occur, the overlapping orbitals for each shell will split into two *sub-orbitals* - one with slightly larger energy and one with slightly lower energy. The energy difference between these sub-orbitals is also much smaller than the energy difference between shells and subshells. Orbital splitting preserves the average energy of each electron but introduces the necessary quantum states to describe them. This is the case when two atoms are brought together, but the orbital splitting continues as more and more atoms are brought together. As the number increases to the point of forming a macroscopic solid, the orbital splitting can be approximated as a continuous band of quantum states - the energy difference between states is so minute that electrons can move freely between them. This process is demonstrated in Figure A.1.

An atom's behaviour and interactions are largely governed by its valence electrons, and this continues in band structure. The energy bands of interest in a solid are the conduction

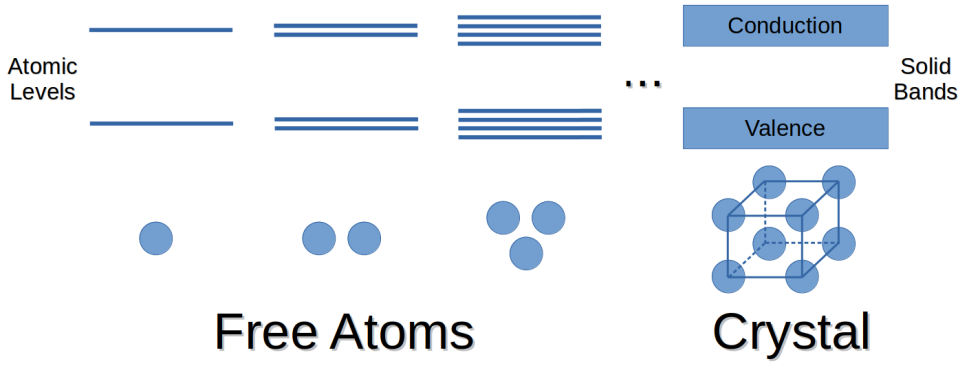


Figure A.1: As atoms are brought together, splitting occurs in the atomic levels to obey the Pauli exclusion principle. At the point of forming a macroscopic crystal, the splitting of states is so fine that the density of states form an almost continuous band of states, where individual states cannot be distinguished.

and valence bands, along with the band gap that separates them. In solids band structure, the conduction band is first *unfilled* band and the valence band is the final *filled* band. More accurately, the conduction band and the valence band are the two bands closest to the *Fermi level*. As fermions, electrons are described using Fermi-Dirac statistics, and the Fermi level is equal to the energy state that has a 50% probability of being occupied at any time. This is calculated using the Fermi-Dirac distribution, shown in Figure A.2 and written as

$$f(\epsilon) = \frac{1}{e^{(\epsilon-\mu)/k_B T} + 1} \tag{A.1}$$

where ϵ is the electron energy, μ is the solids *chemical potential*, k_B is the Boltzmann constant, and T is absolute temperature.

As the Fermi level is defined as the 50% probability of occupation, then $\epsilon = \mu$. The band gap, when it exists, is a region with no occupiable states and is a consequence of the crystal structure, containing energy levels without a corresponding solution to Schrodinger's equation. Solids are classified as metals, insulators, or semiconductors based on a combination of the band-gap and how filled their valence or conduction bands are, as seen in Figure A.3. When the conduction band is only partially filled then electrons have accessible states to move and the solid is called a metal.[1] If the valence band is full and the band-gap is very large, the solid is an insulator as electrons have no free states to

[1]When there is small overlap between the conduction and valence bands, it is called a semimetal instead.

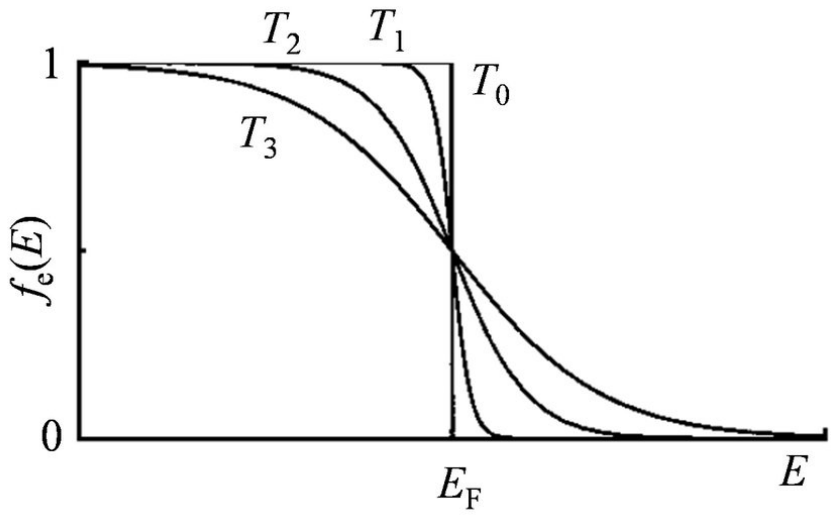


Figure A.2: The Fermi-Diract distribution, for $T_3 > T_2 > T_1 > T_0 = 0$. The distribution describes the distribution of electron occupation of energy states E. The Fermi value E_F corresponds to the point where the state is 50% likely to be occupied. The number of electrons is calculated by multiplying the distribution at an energy by the actual density of states [3].

move to. If the band-gap is small (ie. within the range of thermal excitation energy) then the material is called a semiconductor, as its conductivity is directly proportional to the number of electrons that can be excited to the conduction band.

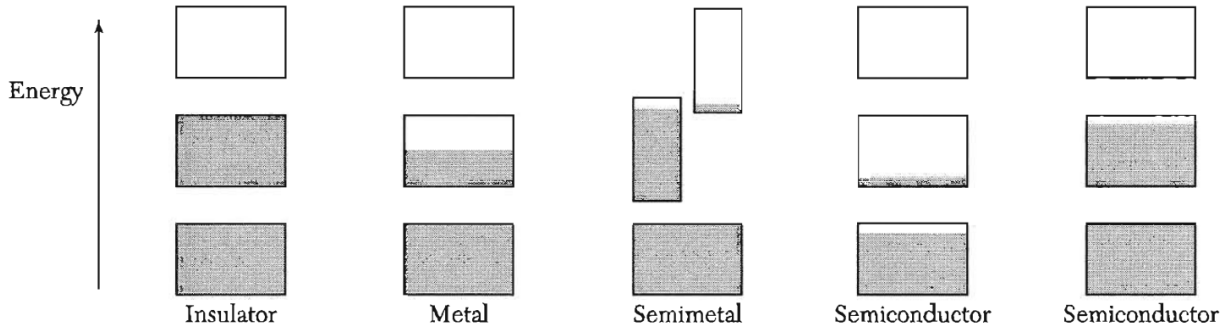


Figure A.3: The electrical difference between "types" of solids is a matter of how full and the separation of the materials conduction and valence bands. The conduction and valence bands represent energy regions of non-zero density of states, while inside the bandgap region the density of states is zero [26].

A.1.1 Semiconductors

Semiconductors formed strictly from Group IV elements such as silicon and germanium or from a combination of Group III and Group V elements (shorthand III-V), such as gallium-arsenide, are called intrinsic semiconductors.

In an intrinsic semiconductor, there is always the same concentration of electrons (n) as concentration of holes (p): $n = p$. This quantity is called the *intrinsic carrier concentration* n_i, with

$$n_i = p = n \tag{A.2}$$

The number of free carriers in a unit volume is given by

$$n_i = BT^{3/2}e^{-E_g/2k_BT} \tag{A.3}$$

where B is a material dependent parameter, T is the absolute temperature in K, E_g is the bandgap of the material, and k_B is Boltzmann's constant. For intrinsic silicon, $B = 7.3 \times 10^{15}\text{cm}^{-3}\text{K}^{-3/2}$ and $E_g = 1.12\text{eV}$. The conductivity σ of an intrinsic semiconductor is calculated using

$$\sigma = qn_i(\mu_n + \mu_p) \tag{A.4}$$

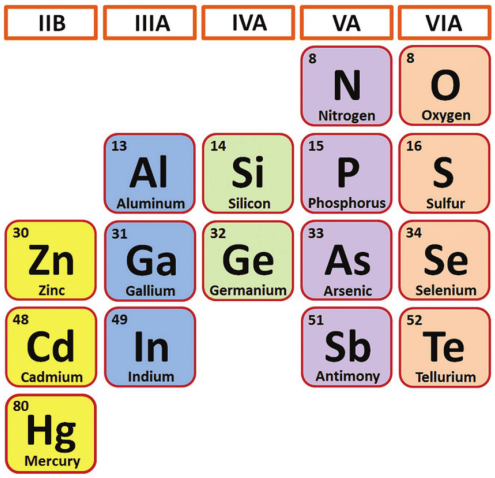


Figure A.4: The groups and associated elements that can be used in various combinations to make semiconductors [31].

where μ_n and μ_p are the electron and hole mobility respectively. One can see the conductivity of an intrinsic semiconductor is therefore a function of the material and temperature only. From Equation A.3, we see the density of free electrons in the semiconductor, and therefore its conductivity, increases with temperature. On the other hand, at $T = 0K$ all intrinsic semiconductors would become insulators unless otherwised biased. This relationship has been plotting in Figure A.5.

When an electron is excited from the valence band to the conduction band, it leaves a vacant quantum state behind that other electrons can then move to. This vacancy is referred to as a *hole* and can be modelled as a particle with positive charge q. Holes can be treated nearly the same as electrons, as fermions which can be the dominant charge carrier in semiconducting devices. They differ in that their *movement* is actually the result of a sequence of electrons moving to occupy the vacancy, which is mostly indistinguishable from movement of the vacancy itself. However, it means that a hole has lower mobility and higher effective mass than an electron.

To modify conductivity the carrier concentrations are changed through the use of *dopants*, a process most commonly performed on silicon. Dopants are atoms with one more or one less valence electron than silicon, but which displace and replace atoms in the crystal when bombarded at high energies. When injected in low concentrations they do not have a major effect on the crystal structure or behaviour, but each dopant introduces a free carrier. When the dopant has 3 valence electrons (ie. it introduces a free hole) it is called an acceptor, and a dopant with 5 valence electrons (ie. it introduces a free electron) is

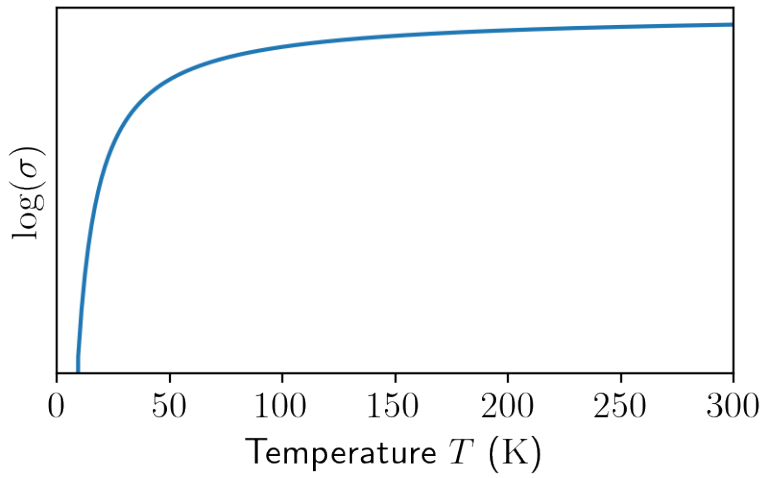


Figure A.5: The logarithmic behaviour of conductivity σ as a function of temperature T for intrinsic silicon.

called a donor. Furthermore, because a dopant either adds or removes an electron, the number of occupied states is different in a doped semiconductor compared to an intrinsic semiconductor. This causes a shift in the Fermi level - towards the conduction band in n-type semiconductors, or towards the valence band in p-type semiconductors. Figure A.6 shows a visualization of how a dopant replaces an original atom and how it causes the Fermi level to shift.

$$\sigma = q(p\mu_p + n\mu_n) \tag{A.5}$$

In sufficient quantities, semiconductor doping will lead to a significant change in the electrical conductivity of the material - highly doped (or *degenerate*) semiconductors will have conductivity approaching that of metal, whereas intrinsic or minimally doped (*non-degenerate*) silicon has quite poor conductivity in comparison. The carrier concentrations are expressed using the following ratio:

$$n_0 p_0 = n_i^2 \tag{A.6}$$

where n_0 is the conducting electron concentration, p_0 the conducting hole concentration, and n_i is the intrinsic concentration of the base material. When a semiconductor is electron dominant it is called an n-type semiconductor, and p-type when hole dominant. Semiconductor doping enables fine tuned control of the semiconductors electrical properties, and is what enables the fabrication of complex devices from semiconductors.

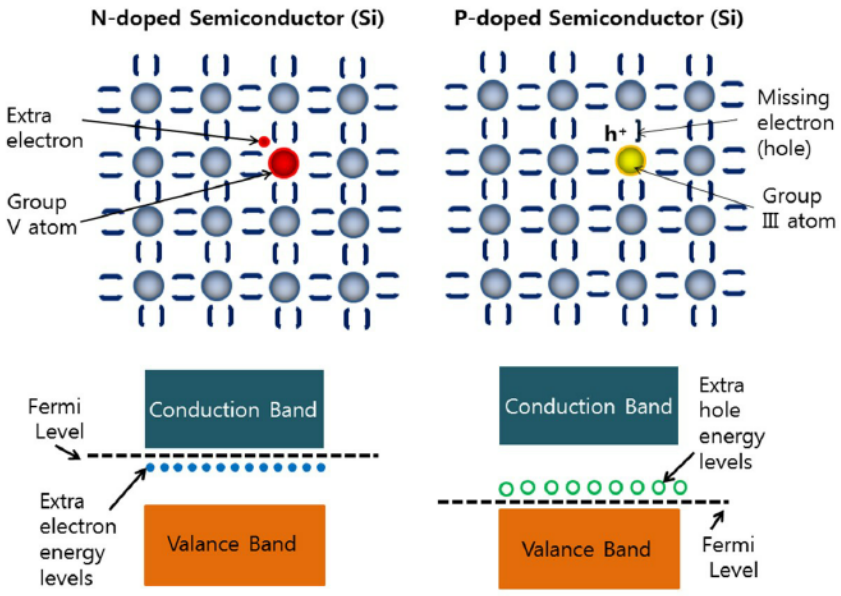


Figure A.6: When dopants are added to intrinsic silicon, the carrier density in the conduction band increases which leads to an increase Fermi level. Likewise, acceptors increase the holel density in the valence band, which in turn decreases the Fermi level [36].

A.2 Devices

A.2.1 Semiconducting P-N Junction

The simplest semiconductor device is a PN-junction, which is created when an interface is formed between an n-type and p-type semiconductor. Although the Fermi level of each semiconductor will be shifted due to concentration differences, when the junction is formed and the two sides reach an equilibrium, by its very definition the Fermi level must be constant across the junction, demonstrated in Figure A.7. The equaliization of the Fermi level across the junction shifts the relative energies of the conduction and valence bands on each side of the junction as well. It becomes energetically favourable for a certain number carriers to move across the junction, resulting in the recombination of the free carriers introduced by doping. The consequence of this recombination is that there is region around the junction lacking in free charge carriers but which is still charged due to the ionized nuclei of the dopants. This is called the depletion region, and a built-in voltage is generated across it; there is zero net current across the junction.

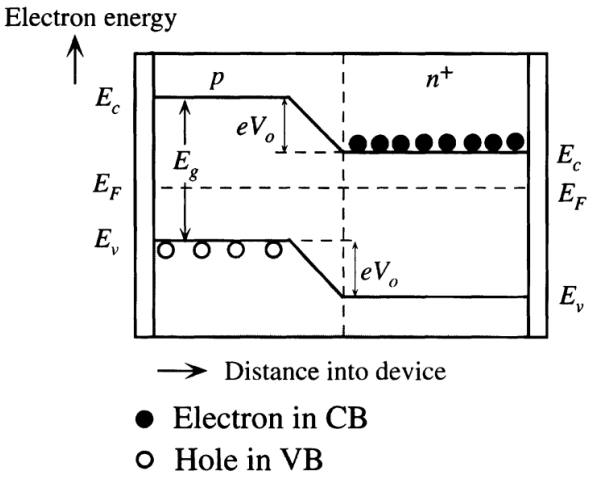


Figure A.7: When a junction between p-type and n-type semiconductors are formed, band bending occurs because the Fermi level E_F equalizes across the junction [24].

There are two ways to generate further current across the junction:

1. A voltage can be applied which reduces the depletion region width and which cancels the built-in voltage. When the depletion region reaches zero width, current will flow.

2. A voltage is applied in the opposite direction, increasing the depletion region width. The crystal will support up to a critical voltage level, after which breakdown will occur in the crystal and current will flow.

A diode and LED are practical examples of a p-n junction, and p-n junctions are the fundamental building blocks of most other semiconducting devices, though others exist such as the p-i-n junction. BJTs and MOSFETs are examples of devices that are formed using p-n junctions.

Appendix B

Markov Chain Proofs

B.1 Theory - Continuous Time Model

B.1.1 Autocorrelation

$$\phi(t) = \langle X(t_0)X(t_0+t)\rangle$$

$$\phi(t) = \left\langle \left(x_0\frac{\tau_0}{\tau_0+\tau_1} + x_1\frac{\tau_1}{\tau_0+\tau_1}\right)\left(x_1P_{11}(t) + x_1P_{01}(t) + x_0P_{10}(t) + x_0P_{00}(t)\right)\right\rangle$$

$$\begin{aligned}\phi(t) = \Bigg\langle &\left(x_0\frac{\tau_0}{\tau_0+\tau_1} + x_1\frac{\tau_1}{\tau_0+\tau_1}\right)x_1P_{11}(t) + \left(x_0\frac{\tau_0}{\tau_0+\tau_1} + x_1\frac{\tau_1}{\tau_0+\tau_1}\right)x_1P_{01}(t)\\ &+ \left(x_0\frac{\tau_0}{\tau_0+\tau_1} + x_1\frac{\tau_1}{\tau_0+\tau_1}\right)x_0P_{10}(t) + \left(x_0\frac{\tau_0}{\tau_0+\tau_1} + x_1\frac{\tau_1}{\tau_0+\tau_1}\right)x_0P_{00}(t)\Bigg\rangle\end{aligned}$$

$$\begin{aligned}\phi(t) = \Bigg\langle &\cancel{x_0\frac{\tau_0}{\tau_0+\tau_1}x_1P_{11}(t)} + x_1\frac{\tau_1}{\tau_0+\tau_1}x_1P_{11}(t) + \cancel{x_0\frac{\tau_0}{\tau_0+\tau_1}x_1P_{01}(t)} + \cancel{x_1\frac{\tau_1}{\tau_0+\tau_1}x_1P_{01}(t)}\\ &+ \cancel{x_0\frac{\tau_0}{\tau_0+\tau_1}x_0P_{10}(t)} + \cancel{x_1\frac{\tau_1}{\tau_0+\tau_1}x_0P_{10}(t)} + \cancel{x_0\frac{\tau_0}{\tau_0+\tau_1}x_0P_{00}(t)} + \cancel{x_1\frac{\tau_1}{\tau_0+\tau_1}x_0P_{00}(t)}\Bigg\rangle\end{aligned}$$

$$\phi(t) = \left\langle x_1^2\frac{\tau_1}{\tau_0+\tau_1}P_{11}(t)\right\rangle$$

B.1.2 Power Spectral Density and Wiener-Khinchin Theorem

$$S(\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty}\phi(t)e^{-i\omega t}dt \tag{B.1}$$

$$= \frac{1}{\pi}\int_{0}^{\infty}\phi(t)e^{-i\omega t}dt \tag{B.2}$$

$$= \frac{1}{\pi}\int_{0}^{\infty}\left(\frac{\Delta X\tau_1}{\tau_0+\tau_1}\right)^2 e^{-i\omega t}dt + \frac{1}{\pi}\int_{0}^{\infty}\frac{(\Delta X)^2\tau_0\tau_1}{(\tau_0+\tau_1)^2}e^{-\left(\frac{1}{\tau_0}+\frac{1}{\tau_1}\right)t}e^{-i\omega t}dt \tag{B.3}$$

$$= S_A + S_B \tag{B.4}$$

The first term S_A is a constant in time, hence

$$S_A(\omega) = \left(\frac{\Delta X\tau_1}{\tau_0+\tau_1}\right)^2\frac{\delta(\omega)}{\pi}. \tag{B.5}$$

The second term S_B is

$$S_B(\omega) = \frac{1}{\pi}\int_{0}^{\infty}\frac{(\Delta X)^2\tau_0\tau_1}{(\tau_0+\tau_1)^2}e^{-\left(\frac{1}{\tau_0}+\frac{1}{\tau_1}\right)t}e^{-i\omega t}dt \tag{B.6}$$

$$= \frac{1}{\pi}\int_{0}^{\infty}\frac{(\Delta X)^2\tau_0\tau_1}{(\tau_0+\tau_1)^2}e^{-\left(\frac{1}{\tau_0}+\frac{1}{\tau_1}\right)t}\cos(\omega t)dt \tag{B.7}$$

$$= \frac{1}{\pi}\frac{(\Delta X)^2\tau_0\tau_1}{(\tau_0+\tau_1)^2}\left[1+\left(\frac{\omega\tau_0\tau_1}{\tau_0+\tau_1}\right)^2\right]^{-1} \tag{B.8}$$

B.1.3 Power Spectral Density Scaling Factor

A scaling factor of 2π may or may not be included in the Wiener-Khinchin derived PSD. The presence of such a scaling factor depends on how the Fourier transform is defined (specifically if using ω instead of f.) An additional scaling factor of 2 may be present based on restrictions placed upon f. Consider that the signals power is found through integrating the PSD over all frequencies:

$$P = \int_{-\infty}^{\infty}S_x(f)df = 2\int_{0}^{\infty}S_x(f)df$$

Using the property that $S_x(f) = S_x(-f)$ if x is a real process. If we restrict our region of interest to $f > 0$, which is valid considering this is a real signal, then the extra scaling factor would need to be introduced to satisfy overall power.

Source: https://www.asc.ohio-state.edu/jayaprakash.1/846/noiseho.pdf

B.1.4 Alternative to Kolmogorov's Forward Equation

Kolmogorov's Forward Equation is used to calculate the probability of being in a state after some time t, but it can seem to be an arbitrary way of arriving there. It treats $P_{ij}(t)$ as an abstraction from the actual probabilities and the solution is found with a differential equation. However, we can use the exponential distribution of calculating the probability of specific life times or we can use the Poisson distribution to calculate the probability of n transitions in a time window. I attempted to use the Poisson probability to calculate $P_{11}(t)$ without use of the different equation:

Let N be Poisson-distributed with mean $1/\tau$. The probability of n transitions occurring in time t is given through the Poisson point process:

$$P(N(t) = n) = \frac{(t/\tau)^n e^{-t/\tau}}{n!}$$

Let X be Poisson-distributed with mean $1/\tau_u$, corresponding to transitions up. Let Y be Poisson-distributed with mean $1/\tau_d$, corresponding to transitions down. The probability $P_{11}(t)$ must equal the probability that Y is equal to X:

$$\begin{aligned} P_{11}(t) &= P(X = Y) \\ &= P(X - Y = 0) \\ &= e^{-(1/\tau_u + 1/\tau_d)} I_0(2\sqrt{1/(\tau_u \tau_d)}) \end{aligned}$$

Which follows the Skellam distribution, where $I_0(x)$ is the zero-order modified Bessel function of the first kind. This introduces two problems: the removal of any time dependence and the problem of analytically reducing a modified Bessel function whose input is variable. Despite other attempts to formulate the probability $P_{11}(t)$ in terms of explicit probabilities, many problems would always arise and a satisfactory method could not be found. Kolmogorov's Forward Equation is an effective way of bypassing or simplifying the complexity of trying to account for all possible transition sequences in an arbitrary timeframe.

B.2 Theory - Discrete Time

B.2.1 Transition Matrix Simplification

Lemma: *Let $A = (a_{ij})$ and $B = (b_{ij})$, both $N \times N$ matrices. The product matrix $A \times B = AB$ has elements:*

$$(AB)_{ij} = \sum_k a_{ik} b_{kj}$$

Using matrix multiplication, we can show how to simplify the transition matrix. First, solve for transition probability after $s = 2$ steps, and then after $s = 3$ steps:

$$\begin{aligned} p_{ij}(2) &= P\{X_2 = j \mid X_0 = i\} \\ &= \sum_k P\{X_2 = j \mid X_1 = k\} P\{X_1 = k \mid X_0 = i\} \\ &= \sum_k p_{kj} p_{ik} \\ &= \sum_k p_{ik} p_{kj} \\ &= P^2 \end{aligned}$$

$$\begin{aligned} p_{ij}(3) &= P\{X_3 = j \mid X_0 = i\} \\ &= \sum_k P\{X_3 = j \mid X_2 = k\} P\{X_2 = k \mid X_0 = i\} \\ &= \sum_k p_{kj} P^2 \\ &= P^3 \end{aligned}$$

This will extend to the general case of $p_{ij}(s) = P^s$.

Source: https://www.stat.auckland.ac.nz/~fewster/325/notes/325book.pdf

B.2.2 Further simplification

The single step transition probability matrix:

$$P = \begin{bmatrix} p & q \\ p & q \end{bmatrix}$$

for $p + q = 1$. The two step transition probability matrix:

$$P^2 = PP = \begin{bmatrix} p & q \\ p & q \end{bmatrix} \begin{bmatrix} p & q \\ p & q \end{bmatrix} = \begin{bmatrix} pp + qp & pq + qq \\ pp + qp & pq + qq \end{bmatrix} = \begin{bmatrix} p(p+q) & q(p+q) \\ p(p+q) & q(p+q) \end{bmatrix} = \begin{bmatrix} p & q \\ p & q \end{bmatrix}$$

The three step transition probability matrix:

$$P^3 = PP^2 = PP = P^2 = P$$

It follows that for n steps:

$$P^n = P$$

B.2.3 Eigenvector Calculation

We have:

$$P = \begin{bmatrix} p_0 & q_0 \\ q_1 & p_1 \end{bmatrix} \qquad \begin{aligned} \lambda_1 &= 1 \\ \lambda_2 &= p_0 + p_1 - 1 \end{aligned}$$

For the first eigenvector:

$$P\vec{u} = \lambda_1\vec{u}$$
$$(P - \lambda_1 I)\vec{u} = 0$$
$$\begin{bmatrix} p_0 - 1 & q_0 \\ q_1 & p_1 - 1 \end{bmatrix}\begin{bmatrix} u_0 \\ u_1 \end{bmatrix} = 0$$
$$\begin{bmatrix} -q_0 & q_0 \\ q_1 & -q_1 \end{bmatrix}\begin{bmatrix} u_0 \\ u_1 \end{bmatrix} = 0$$
$$\begin{bmatrix} -q_0u_0 + q_0u_1 \\ q_1u_0 - q_1u_1 \end{bmatrix} = 0$$

$$-q_0u_0 + q_0u_1 = 0 \qquad q_1u_0 - q_1u_1 = 0$$
$$u_1 = \frac{q_0u_0}{q_0} = u_0 \qquad u_1 = \frac{-q_1u_0}{-q_1} = u_0$$

$$\vec{u} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \tag{B.9}$$

To verify whether $\vec{u}$ is an eigenvector, check $P\vec{u} = c\vec{u}$ for c a constant:

$$\begin{bmatrix} p_0 & q_0 \\ q_1 & p_1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} p_0 + q_0 \\ q_1 + p_1 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Thus $\vec{u}$ is an eigenvector.

Likewise, for the second eigenvector:

$$P\vec{v} = \lambda_1\vec{v}$$
$$(P - \lambda_2 I)\vec{v} = 0$$
$$\begin{bmatrix} p_0 - (p_0 + p_1 - 1) & q_0 \\ q_1 & p_1 - (p_0 + p_1 - 1) \end{bmatrix}\begin{bmatrix} v_0 \\ v_1 \end{bmatrix} = 0$$
$$\begin{bmatrix} 1 - p_1 & q_0 \\ q_1 & 1 - p_0 \end{bmatrix}\begin{bmatrix} v_0 \\ v_1 \end{bmatrix} = 0$$
$$\begin{bmatrix} q_1v_0 + q_0v_1 \\ q_1v_0 + q_0v_1 \end{bmatrix} = 0$$

$$q_1v_0 + q_0v_1 = 0 \qquad q_1v_0 + q_0v_1 = 0$$
$$v_1 = \frac{-q_1v_0}{q_0} \qquad v_1 = \frac{-q_1v_0}{q_0}$$

$$\vec{v} = \begin{bmatrix} 1 \\ -q_1/q_0 \end{bmatrix} \tag{B.10}$$

To verify whether $\vec{v}$ is an eigenvector, check $P\vec{v} = c\vec{v}$ for c a constant:

$$\begin{bmatrix} p_0 & q_0 \\ q_1 & p_1 \end{bmatrix}\begin{bmatrix} 1 \\ -\dfrac{q_1}{q_0} \end{bmatrix} = \begin{bmatrix} p_0 - q_1 \\ q_1 - \dfrac{p_1q_1}{q_0} \end{bmatrix} = (p_0 - q_1)\begin{bmatrix} 1 \\ -\dfrac{q_1}{q_0} \end{bmatrix}$$

$$\begin{aligned} \frac{-q_1}{q_0}(p_0 - q_1) &= \frac{-q_1p_0}{q_0} + \frac{q_1^2}{q_0} \\ &= \frac{-q_1(1 - q_0)}{q_0} + \frac{(1 - p_1)q_1}{q_0} \\ &= \frac{-q_1(1 - q_0)}{q_0} + \frac{(1 - p_1)q_1}{q_0} \\ &= \frac{\cancel{-q_1}}{\cancel{q_0}} + q_1 + \frac{\cancel{q_1}}{\cancel{q_0}} - \frac{p_1q_1}{q_0} \\ &= q_1 - (p_1q_1)/q_0 \end{aligned}$$

Thus $\vec{v}$ is an eigenvector.

B.2.4 Expected value of X(s)

Using $x_0 = 0$, $x_1 = \Delta X$, and the following:

$$P^s = \begin{bmatrix} p_{00}(s) & p_{01}(s) \\ p_{10}(s) & p_{11}(s) \end{bmatrix} = \frac{q_0}{q_0+q_1} \begin{bmatrix} g(s)-\alpha & 1-g(s) \\ \alpha(g(s)-1) & 1-\alpha g(s) \end{bmatrix} \quad \gamma_0 = \frac{q_1}{q_0+q_1} \quad \gamma_1 = \frac{q_0}{q_0+q_1}$$

The expectation value $\mathrm{E}(X_s) = \pi_s^T \vec{x}$ is:

$$\begin{aligned}
\pi_s^T \vec{x} &= \pi_0^T P^s \vec{x} \\
&= \begin{bmatrix} \gamma_0 & \gamma_1 \end{bmatrix} \begin{bmatrix} p_{00}(s) & p_{01}(s) \\ p_{10}(s) & p_{11}(s) \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \\
&= \begin{bmatrix} \gamma_0 p_{00}(s) + \gamma_1 p_{10}(s) & \gamma_0 p_{01}(s) + \gamma_1 p_{11}(s) \end{bmatrix} \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \\
&= x_0(\gamma_0 p_{00}(s) + \gamma_1 p_{10}(s)) + x_1(\gamma_0 p_{01}(s) + \gamma_1 p_{11}(s)) \\
&= \Delta X(\gamma_0 p_{01}(s) + \gamma_1 p_{11}(s)) \\
&= \frac{\Delta X q_0}{q_0+q_1} (\gamma_0(1-g(s)) + \gamma_1(1-\alpha g(s))) \\
&= \frac{\Delta X q_0}{q_0+q_1} (\gamma_0 - \gamma_0 g(s)) + \gamma_1 - \gamma_1 \alpha g(s))) \\
&= \frac{\Delta X q_0}{q_0+q_1} (1 - g(s)(\gamma_0 + \gamma_1 \alpha)) \\
&= \frac{\Delta X q_0}{q_0+q_1} \left(1 - g(s) \left(\frac{q_1}{q_0+q_1} + \frac{q_0}{q_0+q_1} \frac{-q_1}{q_0} \right) \right) \\
&= \frac{\Delta X q_0}{q_0+q_1} \left(1 - g(s) \left(\frac{q_1}{q_0+q_1} - \frac{q_1}{q_0+q_1} \right) \right) \\
&= \frac{\Delta X q_0}{q_0+q_1}
\end{aligned}$$

Appendix C

Statistical Models

C.1 Normal Distribution

Also known as the Gaussian distribution, the normal distribution is one of the most common statistical models. It models a random variable x with an average value μ and symmetric variations about that average. The width of the distribution is given by the scaling factor σ. The PDF is

$$f_n(x;\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2, \qquad -\infty < x, \mu < \infty, 0 < \sigma \tag{C.1}$$

and is plotted, along with the CDF, in Figure C.1.

If we expected consistent pulse amplitudes in our dummy pulse train with variations described as white noise, the normal distribution could be used as a model for the average value. Although some amplitudes may be smaller and some may be higher, the average amplitude would still be μ and the probability of measuring other amplitudes would decrease as the distance from μ increased.

C.2 Poisson Distribution

When a discrete random variable N represents the number of pulses (or more generally events) in a continuous (most often in time) interval $(0, x)$, the Poisson distribution is used.

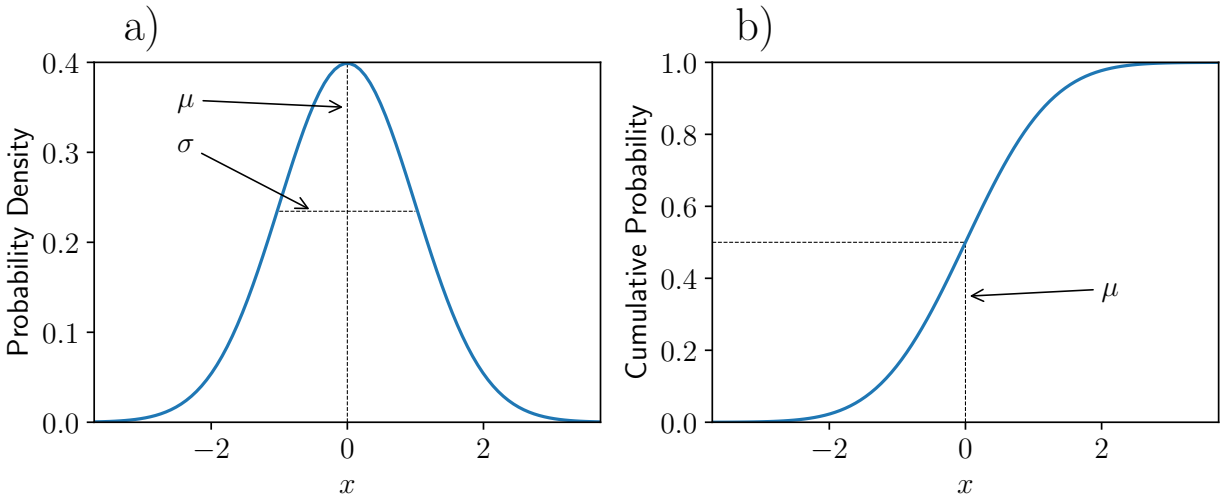


Figure C.1: a) PDF of the normal distribution with mean μ and variance σ marked. b) CDF of the normal distribution. The mean value corresponds to a cumulative probability of 50%.

The Poisson distribution is used very often in real-world applications to count random events over time. The processes where the Poisson distribution is used are collectively called Poisson point processes. Examples of Poisson point processes are the emission from radioactive decay or detector interactions from cosmic ray particles, such as muons. The Poisson distribution has a single parameters λ, the expected rate of events per unit time (assumed to be constant). Although the Poisson distribution is used for processes that occur in continuous space, as a counting process a Poisson random variable can only take discrete values. Thus it is described using a PMF, which is given as:

$$f_N(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!} \qquad 0 < \lambda,\ x \in \mathbb{Z}_+ \tag{C.2}$$

The PMF and CDF are plotted in Figure C.2.

The first and second moments of the Poisson distribution are equal:

$$\mu = \sigma = \lambda \tag{C.3}$$

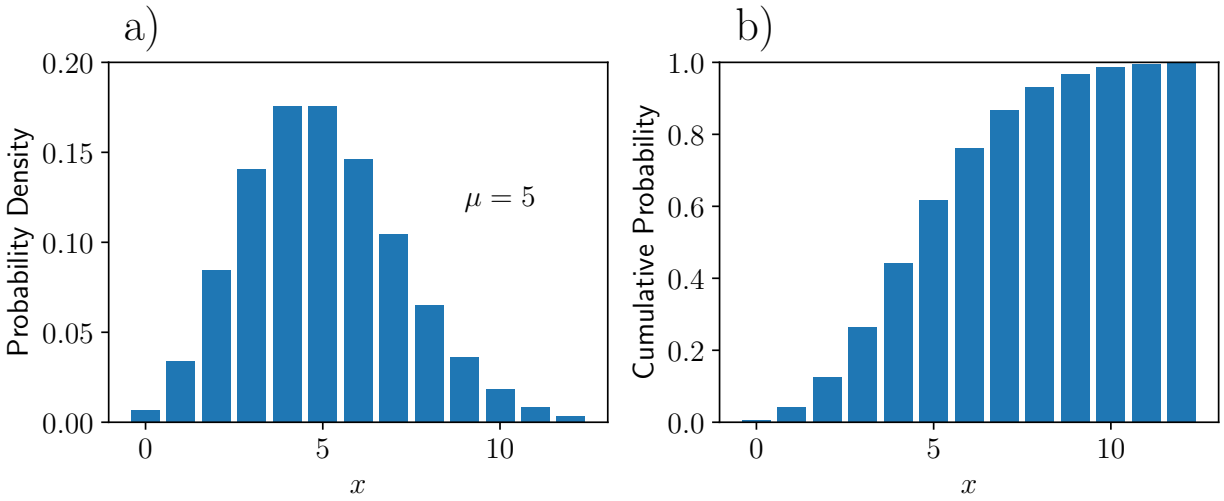


Figure C.2: a) PDF of the Poisson distribution for a process with an average rate λ of 5 events per interval. b) The corresponding CDF when $\lambda = 5$.

C.3 Exponential Distribution

When seeking to model the *time* between events in a Poisson point process rather than the number of events in some interval, the exponential distribution is used. Thus in contrast to the Poisson distribution, it is a continuous distribution over time or space. It can be derived from the Poisson distribution (see 2.3.1), and as a consequence is dependent on the same rate parameter λ. The model parameter is sometimes defined as $\tau = 1/\lambda$, which makes the units more intuitive.

The PDF of the exponential distribution is

$$f(x;\lambda) = \lambda e^{-\lambda x} \qquad 0 < \lambda, 0 \leq x \tag{C.4}$$

with CDF

$$F(x;\lambda) = 1 - e^{-\lambda x} \tag{C.5}$$

The exponential PDF and CDF are shown in C.3.

The exponential distribution has the interesting property of memorylessness:

$$P(T > s + t | T > s) = P(T > t), \tag{C.6}$$

where T is the random variable representing the time between events, and s and t are both intervals of time. What C.6 says is that if an event hasn't occured in an interval s after

the starting time, the probability it will occur in another interval t is the same as if t was counted from the starting time itself - the probability of duration T is the same after 30 seconds as it is after 0 seconds.

This is easily shown using conditional CDF to calculate the probability that the interval will be longer than $s + t$ is s has already elapsed:

$$P(T > s + t | T > s) = \frac{P(T > s + t \cap T > s)}{P(T > s)} \tag{C.7}$$

$$= \frac{P(T > s + t)}{P(T > s)} \tag{C.8}$$

$$= \frac{1 - F(s + t)}{1 - F(s)} \tag{C.9}$$

$$= \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} \tag{C.10}$$

$$= e^{-\lambda t} \tag{C.11}$$

$$= 1 - F(t) \tag{C.12}$$

$$= P(T > t) \tag{C.13}$$

Thus, the probability that the interval will be longer than $s + t$ given s has already elapsed is the same as if s had not occurred. In other words, its future value does not depend on what already happened.

C.4 Negative Binomial and Geometric Distribution

The negative binomial and geometric distributions are analogs of the Poisson and exponential distributions when the events occur in discrete space rather than continuous. The negative binomial distribution is used to model the number of failures (k) from a sequence of n independent Bernoulli trials before r successes occur.[1] For a given probability of success p, the PMF is

$$f(k; r, p) = \binom{k + r - 1}{k} p^r (1 - p)^k. \tag{C.14}$$

[1]The terms *failure* and *success* are not necessarily meant literally, but are placeholders for the two possible outcomes of a binary process.

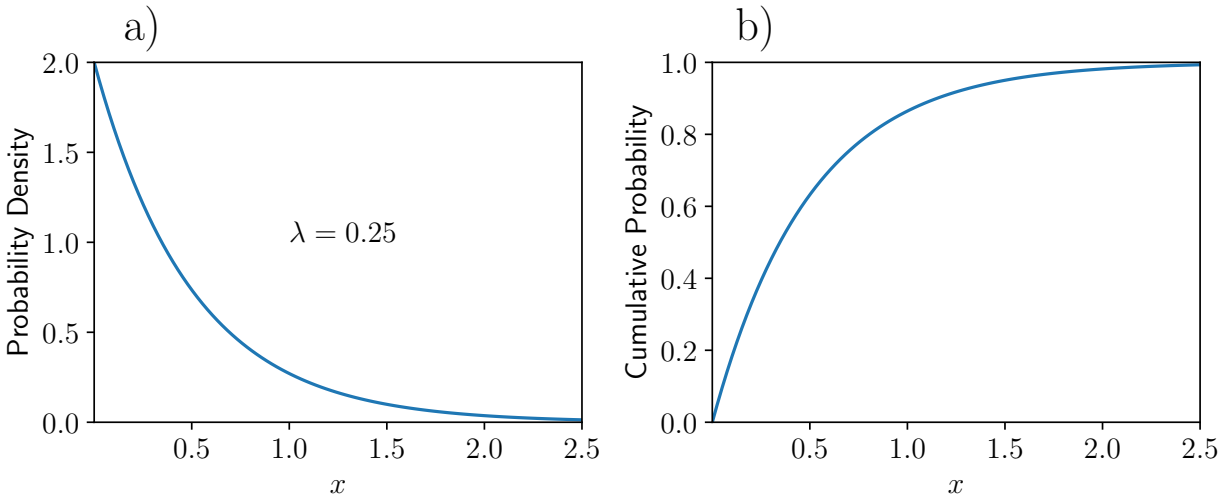


Figure C.3: PDF and CDF for an exponential distribution with rate parameter $\lambda = 0.25$, or alternatively $\tau = 4$.

The geometric distribution is a special case of the negative binomial distribution, which counts the number of failures before the first success ($r = 1$). It has PMF

$$f(k; r = 1, p) = \binom{k + 1 - 1}{k} p^1 (1 - p)^k \tag{C.15}$$

$$= p(1 - p)^k. \tag{C.16}$$

and like its continuous analog, the geometric distribution is memoryless. For a geometrically distributed random variable S,

$$P(S > s + t | S > s) = P(S > t) \qquad s, t \in \mathbb{Z}_+ \tag{C.17}$$

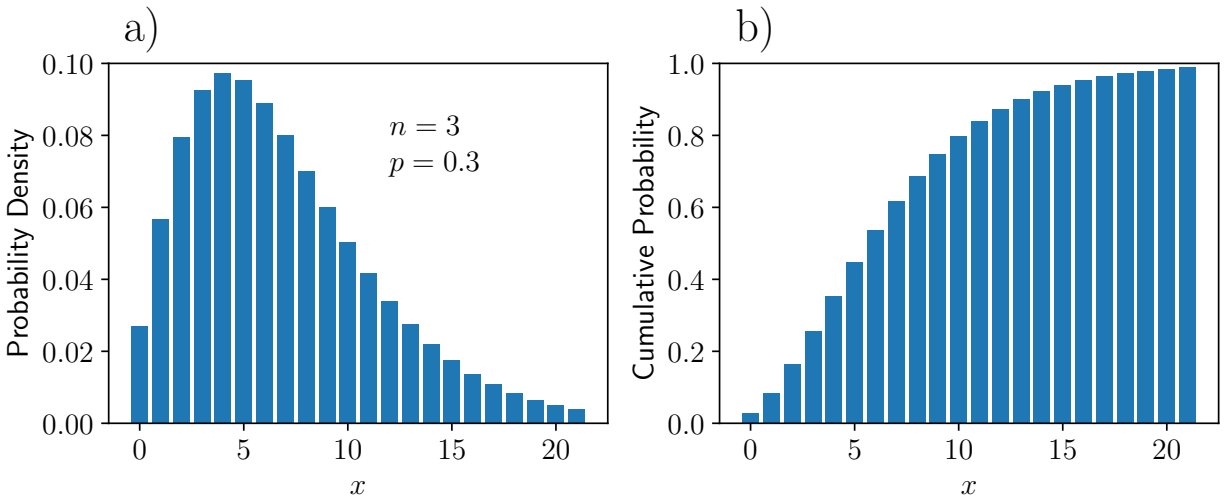


Figure C.4: PDF and CDF of a negative binomial distribution used to count the number of failures until $n = 3$ successes, which occur with a probability of $p = 0.3$.

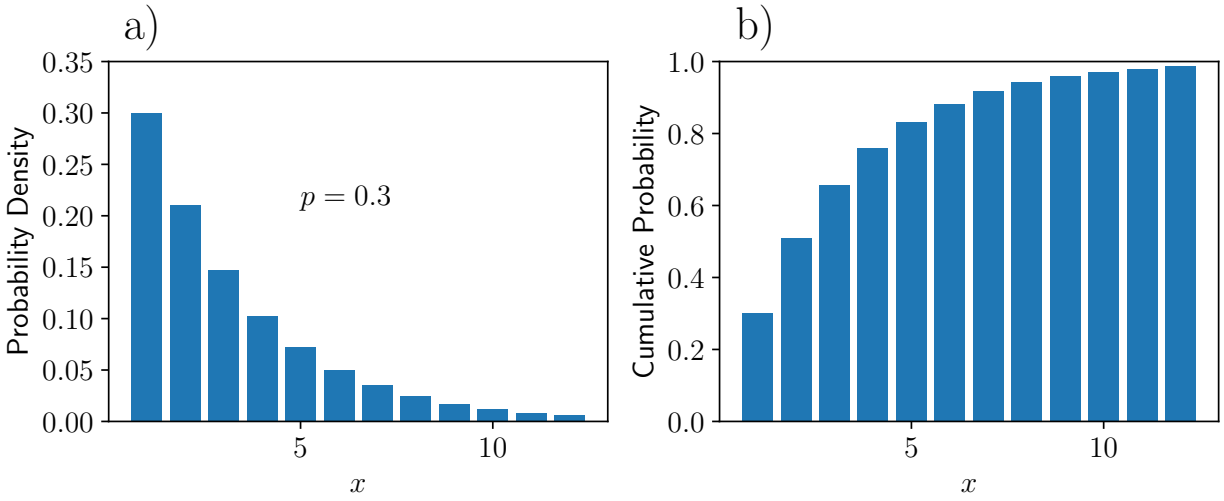


Figure C.5: When the number of failures $n = 1$, the negative binomial distribution is referred to as a geometric distribution. a,b) show the geometric PDF and CDF for the same probability $p = 0.3$ as Figure C.4

.