

Charge re-organization timescales and loss tangents at semiconductor surfaces measured by nc-AFM

Megan Cowie

Department of Physics
McGill University
Montréal, Québec, Canada

A thesis submitted to McGill University
in partial fulfillment of the requirements of the degree of

Doctor of Philosophy

July 2023

© 2023 Megan Cowie

Abstract

The charge density at semiconductor surfaces dynamically re-organizes in response to time-varying electric fields. This charge re-organization has a finite timescale which is associated with energy loss. This Thesis presents a methodology to quantify energy loss at semiconductor surfaces with nanometer spatial resolution and nanosecond temporal resolution. Charge re-organization timescales are shown to be bias-dependent and to increase significantly around defect sites. It is also shown that charge re-organization timescales can fluctuate at defect sites, which is a significant source of low-frequency $1/f$ noise in semiconductor devices.

Résumé

La densité de charge aux surfaces des semi-conducteurs se réorganise dynamiquement en réponse aux champs électriques variant dans le temps. Cette réorganisation de charge a une échelle de temps finie qui est associée à une perte d'énergie. Cette thèse présente une méthodologie permettant de quantifier la perte d'énergie aux surfaces des semi-conducteurs avec une résolution spatiale de l'ordre du nanomètre et une résolution temporelle de l'ordre de la nanoseconde. On montre que les échelles de temps de réorganisation de charge dépendent du biais et augmentent significativement autour des sites de défauts. Il est également démontré que les échelles de temps de réorganisation de charge peuvent fluctuer au niveau des sites de défauts, ce qui constitue une source importante de bruit $1/f$ à basse fréquence dans les dispositifs à semi-conducteurs.

Acknowledgements

First, thank you to Peter for your mentorship, encouragement, and kindness, as well as for interesting scientific and philosophical discussions. Forever fondly etched into my memory will be your pronunciation of *data*. In particular, the word for me is associated with your enthusiasm for exciting experimental results.

Thank you to the Grütter group of past and present for group lunches, interesting discussions, and conference memories. Thank you to Robert, Richard, John, Janney, Juan, and Phil for your technical support.

Thank you to friends made along the way: To Rikke, my lab partner of several years, for being a smiling, shining light to work with; to Hannah and Lia, for your warm-heartedness and friendship; to the *Cinque à set* volleyball team – Aidan, Dave, Soud, Tony, David, Cat, Chris, Brendon, Khady, and Ada – for fun and post-game strategizing at Benelux; to Matteo, who was also on the volleyball team but who was foremost my office-mate, for cheerful greetings in the mornings and your perpetual positive attitude; and to Clara, my “sister friend forever”, Guilherme, Dan, Siggie, Alice, Josephine, José, Simon, Adam, Daniel, Kevin, Alessandro, and others, for happy memories.

Finally, to my family: You bore witness to many of the hours I spent on this project – often with rabbits on my lap – and you empowered me the whole way. Allyson, thank you for your empathy and thoughtfulness; I have always been grateful to have you in my corner. Mom, and Dad, from the bottom of my heart, thank you for your love and support. This thesis is dedicated to you.

Statement of Originality

The following aspects of this thesis represent original knowledge advancement:

- In Chapter 4, a methodology to measure surface charge reorganization timescales and dielectric loss with fm-AFM is presented.
- In Chapter 5, spatially resolved dielectric loss was measured at an individual defect site at a semiconductor surface: Megan Cowie, Taylor J.Z. Stock, Procopios C. Constantinou, Neil Curson, and Grütter, P. Spatially resolved dielectric loss at the Si/SiO₂ interface. *ArXiv* [Submitted, under review.] **2023**. DOI: 10.48550/arXiv.2306.13648
- In Chapter 6, correlated number and mobility random telegraph fluctuations were measured at an individual defect site at a Si/SiO₂ surface: Megan Cowie, Taylor J.Z. Stock, Procopios C. Constantinou, Neil Curson, and Grütter, P. Spatially resolved random telegraph fluctuations of a single trap at the Si/SiO₂ interface. [In progress.] **2023**.
- In Chapter 7, random telegraph fluctuations were measured at an MoSe₂ surface: Megan Cowie, Rikke Plougmann, Zeno Schumacher, and Peter Grütter, P. Single-dopant band bending fluctuations in MoSe₂ measured with electrostatic force microscopy. *Physical Review Materials*. **2021**, 6(10). DOI: 10.1103/physrevmaterials.6.104002
- In Chapter 7, random telegraph fluctuations were measured in pentacene on KBr.

Contributions of Authors

This thesis is based on the three articles listed above, with co-authors Megan Cowie (M.C.), Peter Grütter (P.G.), Rikke Plougmann (R.P.), Zeno Schumacher (Z.S.), Taylor Stock (T.S.), Procopios Constantinou (P.C.), and Neil Curson (N.C.). Additional contributions were made by lab members Josephine Spiegelberg (J.S.), Adam Prus-Czarnecki (A.P-C.), and Simon Briesenick (S.B.), and collaborator Philipp Nagler (P.N.).

- For Chapters 3-6, the Si/SiO₂ sample was fabricated by T.S., P.C., and N.C. Additionally, J.S., A.P-C., and S.B. assisted in setting up Si/SiO₂ sample for experiments: Specifically, they helped locate a particular region of interest on the sample surface.
- For Chapter 4, Z.S. was involved in early discussions regarding data interpretation for a similar system.
- For Chapter 6, T.S., and N.C. were involved in discussions regarding the data analysis.
- For Chapter 7, P.N. prepared the MoSe₂ sample. R.P. assisted in measurements of the MoSe₂ sample.
- P.G. was involved in discussions for all chapters regarding the experimental design and interpretation.
- M.C. performed all of the measurements, modelling, and analysis shown in this thesis.

Contents

1	Introduction	1
2	The electronic properties of a semiconductor surface	5
2.1	Band structure	6
2.2	Bulk carrier statistics	8
2.2.1	Intrinsic semiconductors	8
2.2.2	Extrinsic semiconductors	9
2.3	The MIM capacitor	11
2.3.1	Contact potential difference	12
2.3.2	MIM capacitor force	14
2.4	The MIS capacitor	15
2.4.1	Band bending	16
2.4.2	Bias regimes	17
2.4.3	MIS capacitor force	21
3	Frequency-modulated atomic force microscopy	24
3.1	fm-AFM operating principle	26
3.1.1	Optical detection and feedback controllers	30
3.2	Electrostatic force microscopy	32
3.2.1	fm-EFM operating principle	33
3.2.2	fm-ac-EFM operating principle	34
3.2.3	fm-ac-KPFM operating principle	35
3.3	The MIS fm-AFM model	37

3.3.1	MIS frequency shift	37
3.3.2	MIS excitation	38
3.3.3	Geometrical considerations	40
4	The time-dependent interfacial polarization	44
4.1	Measured lag time of Si/SiO ₂	46
4.1.1	Bias dependence	46
4.1.2	Dopant density dependence	49
4.2	fm-AFM measurement of the interfacial polarization	52
4.2.1	Temporal dispersion	54
4.2.2	Spatial anisotropy and non-homogeneity	57
4.2.3	fm-AFM MIS expression	59
4.3	Dielectric loss at the Si/SiO ₂ interface	60
4.3.1	Loss tangent	61
4.3.2	Relation to mobility	62
5	Defect scattering at Si/SiO₂ trap states	68
5.1	Spatial heterogeneity of Si/SiO ₂	69
5.1.1	Loss peaks at an interfacial trap	71
5.1.2	Dispersion rings	74
5.1.3	Peak shape dependencies	75
5.2	Dopant density-dependent defect scattering	77
6	Interfacial noise	83
6.1	Defect-localized noise	85
6.2	Quantifying RTSs	86
6.2.1	Fluctuation amplitude	87
6.2.2	Exponential wait times	87
6.2.3	Lorentzian power spectral density	88
6.2.4	Correlated number and mobility fluctuations	90
6.3	Multi-state reaction pathways	91

6.3.1	Hydrogen passivation / activation of Pb centers	92
6.3.2	Structural meta-stabilities of oxide traps	94
6.4	Bias dependencies	94
6.4.1	Reaction rate	95
6.4.2	Fluctuation amplitude	99
7	Comparison systems	105
7.1	Other semiconductors	106
7.1.1	MoSe ₂	106
7.1.2	Pentacene	107
7.2	Bias spectroscopy	109
7.2.1	Sample thickness dependence	112
7.3	Defects	114
7.3.1	MoSe ₂ rings	114
7.3.2	Pentacene edges	116
7.4	Random telegraph noise	119
7.4.1	MoSe ₂ noise	119
7.4.2	Pentacene noise	121
8	Conclusions	128
Appendix A	Experimental details	135
A.1	UHV System	135
A.2	KPFM imaging	137
A.3	Height measurements with AFM	138
A.4	Samples	142
Appendix B	Model & fitting details	144
B.1	Distance dependencies	144
B.2	Si/SiO ₂ fit assessment	146
B.3	MoSe ₂ fit assessment	148
B.4	Pentacene fit assessment	149

B.5	RTS timescales	150
Appendix C Derivations		151
C.1	The classical limit of Fermi-Dirac	151
C.2	Bulk carrier densities	152
C.3	Bulk intrinsic energy and density	153
C.4	Bulk extrinsic carrier densities	154
C.5	The MIM capacitor force	155
C.6	The MIS capacitor electric field	156
C.7	The MIS capacitor energy	157
C.8	The MIS capacitor force	158
C.9	The damped, driven free harmonic oscillator	159
C.10	fm-ac-EFM	160
C.11	Exponential probability distribution	161

Chapter 1

Introduction

The Stone, Bronze, and Iron Ages are periods of human history that are named for the preeminent technological materials of the time. This materials-based timeline is based, pragmatically, on archaeological findings,* but it carries a certain poeticism from the perspective of a materials scientist: It highlights that materials and human history are intimately interwoven; that it is by studying material properties and understanding how to engineer them that humankind builds technologies which revolutionize civilizations. Our modern era is centered around information and computation which has, as its heart, semiconducting materials. Our current place in the timeline of humanity is therefore named for the most important semiconductor that exists today. We belong to the Silicon Age[1].

Silicon is the foundational building block of digital electronics. The reason for this is that the most fundamental ingredients of logic – “not”, “and”, “or”, etc. – can be expressed using simple combinations of wires and transistors, and silicon has proven to be an excellent transistor material. The near-unimaginable complexity of digital computation arises from endless streams of ones and zeros navigating labyrinths of silicon-based logic gates. Over time, as demands for computational complexity have increased, transistors have gotten smaller and integrated circuits exponentially denser (following Moore’s Law). Computers

* This timeline was originally developed as a categorization methodology for pre-historic archaeological findings[2]. It is now considered by many archaeologists to be overly simplistic, since it is Eurocentric and ignores many regions of the world which had an altogether different technological trajectory.

have transformed from roaring, room-sized calculators into laptops, smartphones, and the infrastructure of the internet. Today, microfabrication capabilities for silicon-based integrated circuits are wildly unparalleled for any other material. A modern integrated circuit facility costs more than 10 billion dollars to build, and more than 5 million dollars per day to operate[3]. Within these factories, atomically flat silicon wafers with diameters of almost half a meter are patterned with nanoscale precision; this difference in scale is an astounding 9 orders of magnitude.

In addition to transistors, countless other technologies of today are built out of semiconductors. For instance, diodes (including light-emitting diodes, laser diodes, photodiodes, Schottky diodes, pn junction diodes, etc.) are used for light generation, light sensing, and as circuit components in digital electronics. Diodes are predominantly made out of silicon, germanium, or gallium arsenide. Solar panels, which have begun to transform global energy production, are another semiconductor-based technology. Today's commercially available solar panels are predominantly silicon-based, but there are vast ongoing research efforts bent toward developing solar panels based on other photovoltaic materials, including perovskites, quantum dots, and organic thin films and liquid dyes.

The next revolutionary semiconductor technology may be quantum computers. The field is still in its infancy, but future quantum computers will exponentially outpace classical computers in certain applications of optimization and cryptography. Consequentially, quantum computing could revolutionize many foundational aspects of modern society, such as currency storage and exchange, personalized medicine, and database searching[4]. One emerging quantum bit (qubit) contender is the electron spin qubit. Silicon is a particularly promising material to host spin qubits, in part due to the sophistication of existing silicon micro- and nanofabrication technology[5]. In one architecture, individual atoms are buried a few nanometers beneath a silicon surface, allowing for single spins to be electronically accessed using an applied gate voltage[5–8]. In another design, spins are confined to gate-defined quantum dots, which are again located a few nanometers beneath the surface[5, 9]. The next step for this technology is to increase the computational complexity of these devices; that is, to increase the number of silicon-based qubits which can be fabricated and accessed in a reliable way.

In the context of having led to such revolutionary and widespread real-world technologies, semiconductor physics is arguably (in competition with medicine) the most successful scientific field of all time. And still, this story is far from its conclusion. The basis of many of the most exciting materials science discoveries of recent decades is that the electronic properties of a material depend on its size. Graphene, for example, which was the first two-dimensional material to be isolated in high quality, became an early cornerstone of nanotechnology because its electronic structure departs so steeply from the bulk[10, 11]. Transistors are shrinking toward atomic dimensions, atomically thin materials have become promising photovoltaics, and the most advanced silicon-based quantum processor to date contains 12 isolated spin qubits, and was announced in June of this year[12].

With all of these progressions, semiconductor devices are shrinking in at least one dimension. Proportionally, they have less bulk and more surface; less ideal infinite crystallinity and more atomic-scale variability. And so, it is increasingly important to consider how the semiconductor’s nanoscale structure affects its global electronic properties. In this thesis, the most fundamental semiconductor device – a metal-insulator-semiconductor capacitor – is characterized at the nanoscale with atomic force microscopy. Chapters 2 and 3 introduce the semiconductor physics and measurement methodology; Chapters 4-6 show bias- and dopant density-dependent nanoscale spatial heterogeneities in charge re-organization timescales, dielectric dispersion, and noise of a silicon surface; and in Chapter 7, two widely different semiconducting materials – MoSe₂ and pentacene – are similarly explored. The introduction to each chapter is intended for the lay reader, with the aim (for the most part) of relating some of the concepts explored in this thesis to common experiences of everyday life.

References

- [1] Ross, I. M. “The foundation of the Silicon Age”. *Physics Today* 50.12 (**1997**). DOI: 10.1063/1.881628.
- [2] John Hunter-Duvar. *The stone, bronze, and iron ages: A popular treatise on early archaeology*. S. Sonnenschein, **1892**.

- [3] Howard Tigelaar. *How transistor area shrank by 1 million fold*. Springer, **2020**. Chap. ISBN: 9783030400217. DOI: 10.1007/978-3-030-40021-7.
- [4] Brooks, M. “Quantum computers: what are they good for?” *Nature* 617.7962 (**2023**). DOI: 10.1038/d41586-023-01692-9.
- [5] Morton, J. J. L., McCamey, D. R., Eriksson, M. A., and Lyon, S. A. “Embracing the quantum limit in silicon computing”. *Nature* 479.7373 (**2011**). DOI: 10.1038/nature10681.
- [6] Wellard, C. J., Hollenberg, L. C. L., Parisoli, F., Kettle, L. M., Goan, H.-S., McIntosh, J. A. L., and Jamieson, D. N. “Electron exchange coupling for single-donor solid-state spin qubits”. *Physical Review B* 68.19 (**2003**). DOI: 10.1103/physrevb.68.195209.
- [7] Zhang, X., Li, H.-O., Cao, G., Xiao, M., Guo, G.-C., and Guo, G.-P. “Semiconductor quantum computation”. *National Science Review* 6.1 (**2018**). DOI: 10.1093/nsr/nwy153.
- [8] Gonzalez-Zalba, M. F., Franceschi, S. de, Charbon, E., Meunier, T., Vinet, M., and Dzurak, A. S. “Scaling silicon-based quantum computing using CMOS technology”. *Nature Electronics* 4.12 (**2021**). DOI: 10.1038/s41928-021-00681-y.
- [9] Zwerver, A. M. J. et al. “Qubits made by advanced semiconductor manufacturing”. *Nature Electronics* 5.3 (**2022**). DOI: 10.1038/s41928-022-00727-9.
- [10] Geim, A. K. “Graphene: Status and prospects”. *Science* 324.5934 (**2009**). DOI: 10.1126/science.1158877.
- [11] Novoselov, K. S., Geim, A. K., Morozov, S. V., Jiang, D., Zhang, Y., Dubonos, S. V., Grigorieva, I. V., and Firsov, A. A. “Electric field effect in atomically thin carbon films”. *Science* 306.5696 (**2004**). DOI: 10.1126/science.1102896.
- [12] *Intel’s New Chip to Advance Silicon Spin Qubit Research for Quantum Computing*. <https://www.intel.com/content/www/us/en/newsroom/news/quantum-computing-chip-to-advance-research.html>. Accessed: 2023-07-08.

Chapter 2

The electronic properties of a semiconductor surface

Every semiconductor-based electronic technology involves, as its most fundamental process, the application of an electric field to the semiconductor. The simplest system to apply an electric field to a semiconductor, and the device at the core of this thesis, is the metal-insulator-semiconductor (MIS) capacitor. The importance of the MIS capacitor is stated in *Physics of Semiconductor Devices* which, initially published in 1969 and now in its third edition, has become perhaps the world's most widely cited materials science reference[1]:

“The metal-insulator-semiconductor (MIS) capacitor is the most useful device in the study of semiconductor surfaces. Since most practical problems in the reliability and stability of all semiconductor devices are intimately related to their surface conditions, an understanding of the surface physics with the help of MIS capacitors is of great importance to device operations.”

– S. M. Sze and K. K. Ng, *Physics of Semiconductor Devices*, (2007)

Materials science occurs at the threshold between microscopic and macroscopic. By consequence, materials tend to be described using an assortment of quantum mechanical and classical models. The MIS capacitor is an example of such a semi-classical system. A

basic understanding of the MIS capacitor begins with a simple quantum mechanical model of an electron in a crystalline solid (Section 2.1) and a semi-classical description of the global organization of charge in the semiconductor bulk (Section 2.2) and, ultimately, at the semiconductor surface (Section 2.4).

2.1 Band structure

A simple model of crystalline solids considers a single electron in a periodic potential $V(\vec{r})$. (This derivation is shown in e.g. [2].) If the electron is free – that is, if $V(\vec{r}) = 0$ – the energy eigenvalues are:

$$E(\vec{k}) = \frac{\hbar^2 k^2}{2m_o} \quad (2.1)$$

This parabolic $E - \vec{k}$ relationship is illustrated in Figure 2.1a. If $V(\vec{r})$ is nonzero and small, the degenerate free electron levels shift away from each other, as illustrated in Figure 2.1b. The defining feature of semiconducting materials is that the Fermi level E_f^* falls within a band gap E_g . The conduction band energy E_c is the minimum energy of unoccupied states

* Throughout this text E_f is written instead of μ , the chemical potential. Strictly, $E_f = \mu(T = 0)$, but up to room temperature (the temperature of all measurements presented in this work), μ and E_f agree to a high degree of precision, so E_f was used to match the typical notation in semiconductor texts.

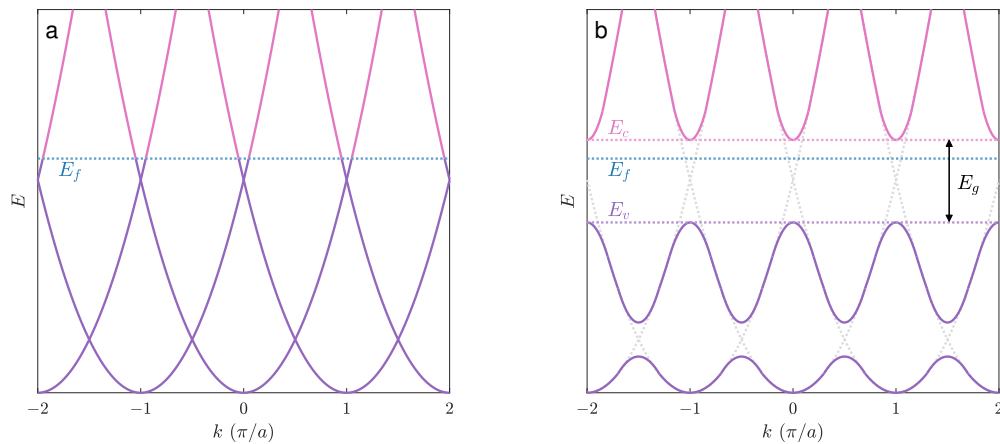


Figure 2.1: **Free and nearly free electron band structure.** $E - \vec{k}$ relationship for a (a) free and (b) nearly free electron in a 1D lattice with periodicity a . States below the Fermi energy E_f are occupied, and states above E_f are unoccupied. In the nearly free electron model E_f falls within a band gap E_g , and the conduction band energy E_c is the lowest unoccupied energy and the valence band energy E_v is the highest occupied energy at $T = 0$.

and the valence band energy E_v is the maximum energy of occupied states when $T = 0$. The energy eigenvalues at the conduction and valence band edges are then:

$$E(\vec{k}) = \begin{cases} \frac{\hbar^2 k^2}{2m_e} + E_c \\ \frac{\hbar^2 k^2}{2m_h} + E_v \end{cases} \quad (2.2)$$

where m_e and m_h are the effective masses of the electron and hole, respectively, given by:

$$m_{e,h} \equiv \hbar^2 \left[\frac{\partial^2 E}{\partial k^2} \right]^{-1} \quad (2.3)$$

i.e. the effective mass is defined by the band curvature.

In the free electron model (Equation 2.1), Equation 2.3 gives a constant effective mass for all E and all $\vec{k}$, which is equal to the electron rest mass, m_o . In the nearly free electron model (Equation 2.2), the effective mass depends on the $E - \vec{k}$ landscape, which varies from band to band. Real materials generally have an anisotropic band edge, but all of the carrier density derivations to follow are done with the approximation that the band is isotropically parabolic (i.e. m_e and m_h are constant). The validity of this approximation for the materials studied in later chapters of this thesis, and the degree to which the model depends on effective mass values, will be discussed in Chapter 7.

Following directly from Equation 2.2, the density of conduction and valence band states are:

$$g_c(E) = \frac{1}{2\pi^2} \left(\frac{2m_n}{\hbar^2} \right)^{3/2} \sqrt{E - E_c} \quad (2.4a)$$

$$g_v(E) = \frac{1}{2\pi^2} \left(\frac{2m_p}{\hbar^2} \right)^{3/2} \sqrt{E_v - E} \quad (2.4b)$$

where $g_c(E)$ is the density of electron states in the conduction band and $g_v(E)$ is the density of hole states in the valence band. Ultimately, many concepts which originate from the nearly free electron model form the foundation of the MIS capacitor model. These have all been explicitly stated above (E_g , E_f , E_c , E_v , m_n , m_p , $g_c(E)$, $g_v(E)$).

2.2 Bulk carrier statistics

The occupation of the conduction and valence bands is determined by $g_c(E)$, $g_v(E)$, and the probability of occupying those energy states. (This derivation is shown in e.g. [1, 3].) Explicitly, the populations of free electrons (n) and holes (p) in a semiconductor are:

$$n = \int_{E_C}^{\infty} g_c(E) f_{FD}(E) \partial E \quad (2.5a)$$

$$p = \int_{-\infty}^{E_V} g_v(E) [1 - f_{FD}(E)] \partial E \quad (2.5b)$$

where $f_{FD}(E) = \left[1 + \exp\left(\frac{E - E_f}{k_B T}\right)\right]^{-1}$ is the Fermi-Dirac distribution. At typical energy scales, if the Fermi level is sufficiently far from the conduction and valence band edges, the Fermi-Dirac distribution approaches the Maxwell-Boltzmann distribution, which is given by $f_{MB}(E) = \exp\left(\frac{E_f - E}{k_B T}\right)$ (see Appendix C.1). Therefore, the carrier densities are approximately:

$$n = N_c \exp\left(\frac{E_f - E_c}{k_B T}\right) \quad (2.6a)$$

$$p = N_v \exp\left(\frac{E_v - E_f}{k_B T}\right) \quad (2.6b)$$

where N_c and N_v are the effective number of conduction and valence band states, respectively. N_c , N_v , and the derivation of Equation 2.6 are provided in Appendix C.2.

2.2.1 Intrinsic semiconductors

If a semiconductor is intrinsic – that is, completely free of dopants and defects – n and p are necessarily equal due to charge conservation. $n = p = n_i$ is the intrinsic charge density, and the Fermi level of the intrinsic semiconductor, $E_f = E_i$, is:

$$E_i = \frac{E_c - E_v}{2} + \frac{k_B T}{2} \ln\left(\frac{m_p}{m_n}\right)^{3/2} \quad (2.7)$$

which is exactly mid-gap if $m_n = m_p$. The intrinsic carrier density is:

$$n_i = \sqrt{N_c N_v} \exp\left(\frac{-E_g/2}{k_B T}\right) \quad (2.8)$$

Equations 2.7 and 2.8 are derived in Appendix C.3.

2.2.2 Extrinsic semiconductors

An intrinsic semiconductor that has been doped is called an extrinsic semiconductor. Dopants have a bond mismatch with the native lattice, meaning that they introduce states that can fall within the band gap (see Figure 2.2). Dopants with excess valence electrons introduce extra electron states (i.e. they are “donors” of electrons), and n-type (electron) conduction occurs in the conduction band. For example, phosphorous, which has seven valence electrons, is a common n-type dopant for silicon, which has six. Dopants that lack electrons as compared to the intrinsic lattice have extra hole states (i.e. they are “acceptors” of electrons), so p-type (hole) conduction occurs in the valence band. For example, boron, which has five valence electrons, is a common p-type dopant for silicon. At low dopant concentrations, the interaction between dopants is negligible, and the lattice remains approximately crystalline.

Donors have an energy ΔE_d below E_c , such that E_d is the ionization energy of the dopant state. If the electron in this state is thermally excited into the conduction band, due to charge conservation, the total density of electrons in the conduction band is $n = N_d^+ + p$,

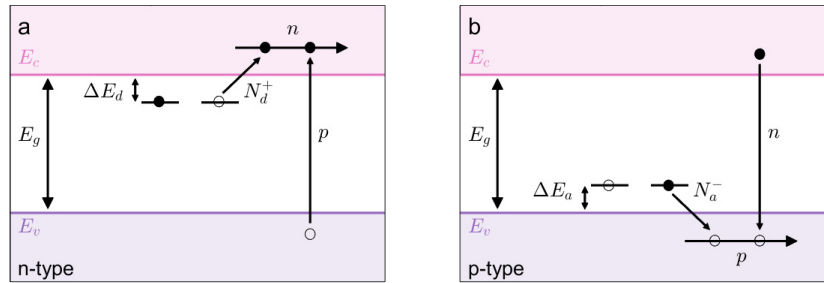


Figure 2.2: **Carrier behaviour of extrinsic semiconductors.** (a) n-type and (b) p-type semiconductor bulk with band gap E_g . Schematic donor and acceptor dopant states ΔE_d and ΔE_a below the conduction band and above the valence band, respectively. Electron conduction occurs in the n-type semiconductor, where $n = p + N_d^+$ and hole conduction occurs in the p-type semiconductor, where $p = n + N_a^-$.

where N_d^+ is the density of thermally ionized donors. Acceptors have an energy ΔE_a above E_v , such that E_a is the electron affinity of the dopant state. If an electron in the valence band is excited into this state, the total density of holes in the valence band is $p = N_a^- + n$. N_d^+ and N_a^- are temperature dependent but, in general, dopants are fully ionized at room temperature (that is, $N_d^+ \approx N_d$ and $N_a^- \approx N_a$). Charge neutrality demands that $n = p$, and therefore:

$$0 = p - n + N_d - N_a \quad (2.9)$$

Given Equation 2.6, the effect of dopants/defects in a semiconductor is to shift the Fermi energy. In an n-doped semiconductor, E_f moves up toward the conduction band, and in a p-doped semiconductor, E_f moves down toward the valence band. This is shown in Figure 2.3. Throughout this thesis and in the MIS capacitor model, E_f is found as a numerical solution

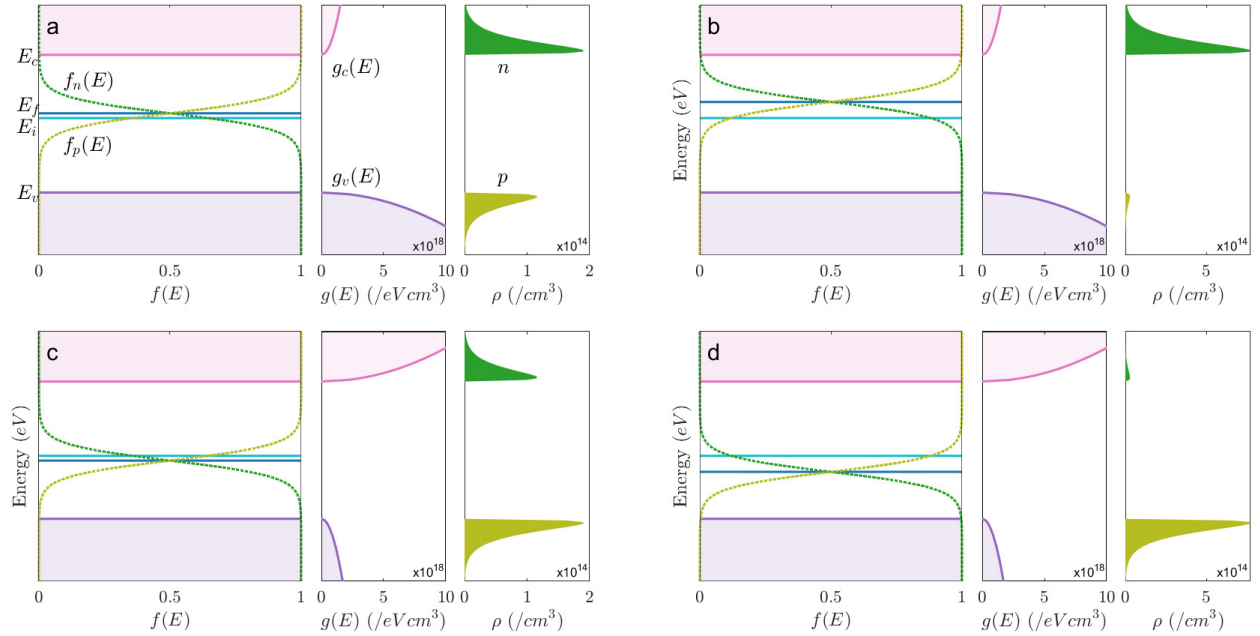


Figure 2.3: Bulk semiconductor carrier statistics. Illustration of the integrals written in Equation 2.5 for extrinsic semiconductors, showing the electron and hole probability distribution functions ($f_n(E) = f_{FD}(E)$ and $f_p(E) = 1 - f_{FD}(E)$), the density of conduction and valence band states ($g_c(E)$ and $g_v(E)$), and the electron and hole densities (n and p). (a-b) Show n-type semiconductors with $m_n = 1, m_p = 2$, (c-d) show p-type conductors with $m_p = 1, m_n = 2$. (a,c) show lower dopant densities ($1e16/cm^3$), (b,d) show higher dopant densities ($1e17/cm^3$). $E_g = 1$ eV for all, and 750 K. (This very high temperature is shown so that $\rho \gg 0$, and the carrier densities can be seen on these axes.)

to the combination of Equations 2.6 and 2.9.

The total electron and hole populations in the bulk for an n-type and p-type semiconductor are therefore:

$$n_n = \frac{N_d - N_a}{2} + \sqrt{\left(\frac{N_d - N_a}{2}\right)^2 + n_i^2}; \quad p_n = \frac{n_i^2}{n_n} \quad (2.10a)$$

$$p_p = \frac{N_a - N_d}{2} + \sqrt{\left(\frac{N_a - N_d}{2}\right)^2 + n_i^2}; \quad n_p = \frac{n_i^2}{p_p} \quad (2.10b)$$

Equation 2.10 is derived in Appendix C.4.

2.3 The MIM capacitor

This section is a detour from semiconductor physics, but is included so that the properties of the simplest capacitor – the metal-insulator-metal (MIM) capacitor – can be directly compared to the MIS capacitor (to be discussed in Section 2.4). A central theme of this thesis involves accounting for deviations between the MIM and MIS capacitors.

The MIM capacitor consists of two neutral perfect conductors (meaning that they have an infinite density of states at the Fermi energy, and infinite conductivity) separated by a neutral perfect insulator, where there is no tunneling between the metals. The electrostatic characteristics of the metals and insulator comprising the MIM capacitor are determined by Gauss' Law and Poisson's Equation: $\nabla^2 V(r) = \vec{\nabla} \cdot \vec{\mathcal{E}}(r) = \rho(r)/\epsilon$. When there is a potential difference V_{AB} between the capacitor plates (A and B), an electric field is established between them. Under the influence of the electric field, electrons in A move toward / away from the metal surface and the electrons in B move away from / toward the metal surface, until each plate carries an equal and opposite charge $Q = Q_A = -Q_B$. Because each metal plate has an effectively infinite density of states, Q assembles over an extremely short depth at the metal surfaces, such that the electric field $\vec{\mathcal{E}}$ is completely screened, and is equal to zero everywhere within the metals (and, equivalently, the potential V is spatially uniform). Within the insulator, $Q = 0$, meaning that $\vec{\mathcal{E}}$ is constant and V varies linearly.

2.3.1 Contact potential difference

V_{AB} can be more precisely understood as the difference in chemical potential between the MIM capacitor plates (where, as stated previously, the “chemical potential” is denoted throughout this thesis as E_f). V_{AB} , therefore, depends on V_g , the applied (gate) voltage across the plates, which raises or lowers the chemical potential of one plate $E_{f,A}$ with respect to the other $E_{f,B}$. In an “ideal” capacitor, it is assumed that under the floating condition (where the vacuum levels are aligned i.e. $E_{vac,A}^F = E_{vac,B}^F$) the chemical potentials are equal (i.e. $E_{f,A}^F = E_{f,B}^F$). In this case, $V_{AB} = V_g$. However, this situation is usually untrue: Generally, $E_{f,A}^F \neq E_{f,B}^F$ (for example, if the two plates are different metals). This inequality is called the contact potential difference, V_{CPD} , which is equal to $E_{f,A}^F - E_{f,B}^F$. Equivalently, V_{CPD} corresponds to the difference in work function ϕ (i.e. the energy required to move an electron from the material bulk to the vacuum level) between the two materials:

$$V_{CPD} = \frac{E_{f,A}^F - E_{f,B}^F}{-e} = \frac{(E_{f,A} - E_{vac,A}) - (E_{f,B} - E_{vac,B})}{-e} = \frac{\phi_A - \phi_B}{-e} \quad (2.11)$$

Given the requirement that the potential across the MIM capacitor is continuous, it is necessarily true that:

$$0 = V_g - V_{CPD} - V_{AB} \quad (2.12)$$

Consequently, the condition where the capacitor is uncharged ($Q = 0$), which necessarily corresponds to $V_{AB} = 0$, occurs when $V_g = V_{CPD}$, not when $V_g = 0$. This equipotential condition is illustrated in Figure 2.4.

In the case of a perfect MIM capacitor, $V_{AB} = V_{ins}$, the potential drop across the insulator (where $V_{ins} = -eQ/C_{ins}$). This is necessarily true because, as stated previously, in a perfect MIM capacitor the plates are neutral and the potential drop across a metal equals zero.* In practice there are many “non-perfect” cases in which there is a charge density localized at the surface of one or both capacitor plates, which leads to an extra potential contribution in the capacitor in the form of a surface dipole, V_{SD} . For example, the surface charge density varies depending on the crystal face structure (e.g. a [111]-oriented cubic structure has a larger

* To clarify a potential source of confusion: A “perfect” MIM capacitor, which consists of perfect metal plates and a perfect insulator, is not the same as an “ideal” capacitor, in which $V_{CPD} = 0$.

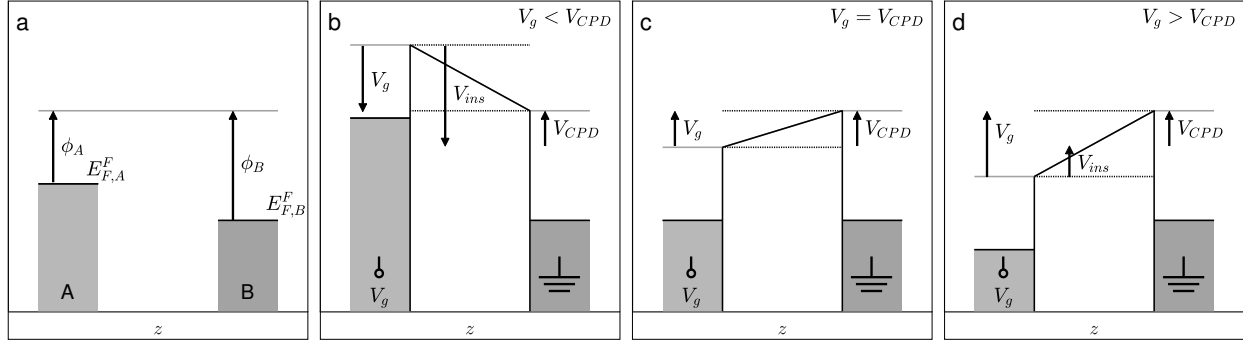


Figure 2.4: **Potential continuity of the MIM capacitor.** Band diagrams of an MIM capacitor with plates A and B of work functions ϕ_A and ϕ_B , where $\phi_A < \phi_B$ and therefore (according to Equation 2.11) $V_{CPD} > 0$. The MIM capacitor is shown (a) floating; (b-d) with an applied bias V_g (with respect to ground, applied to plate A). (b) $V_g < V_{CPD}$, (c) $V_g = V_{CPD}$, and (d) $V_g > V_{CPD}$. The capacitor charge is not shown explicitly, but in (b) $-Q_B = Q_A < 0$, (c) $Q_B = Q_A = 0$, and (d) $-Q_B = Q_A > 0$.

density of surface states than a [100]-oriented structure); crystalline step edges, adatoms, and vacancies all similarly have variable charge density; as do, of course, adsorbates and other defects.

V_{SD} can be thought of as a modification to V_{CPD} : Given that any electron moving from bulk to vacuum necessarily passes through a surface, the work function is necessarily a surface concept, so a variable surface charge density could be thought of as a surface-dependent ϕ . Then, whether a V_{SD} is a positive or negative additive contribution V_{CPD} depends on the sign of the surface dipole. Then, given that E_f of the bulk is essentially unaffected by the surface dipole, $V_{AB} = V_{SD} + V_{ins}$ so the potential continuity equation in Equation 2.13 becomes:

$$0 = V_g - V_{CPD} - V_{SD} - V_{ins} \quad (2.13)$$

The behaviour of the MIM capacitor, then, depends highly on the nature of V_{SD} . For instance, V_{SD} might be spatially non-homogeneous, so the charge in the capacitor would be dependent on the area of the capacitor plates, and on their positions with respect to one another. Or, perhaps V_{SD} could be a non-constant term; that is, it might depend on the insulator thickness z_{ins} or V_g . The implications of both of these examples for the MIM capacitor will be discussed in greater detail in Section 3.3.

2.3.2 MIM capacitor force

The energy of a capacitor with plates A and B is generically:

$$U_{AB} = \frac{1}{2} \left(\int_A \rho_A V_A \partial v_A + \int_B \rho_B V_B \partial v_B \right) \quad (2.14)$$

where ρ is the volume charge density, V is the potential, and v is the volume. For a MIM capacitor, the charge density and potential are spatially constant, so the above expression reduces to:

$$U_{AB} = \frac{1}{2} C V_{AB}^2 \quad (2.15)$$

and, given Equation 2.12, if $V_{SD} = 0$, the force acting between the capacitor plates is:

$$F_{AB} = -\frac{1}{2} \frac{\partial C}{\partial z_{ins}} (V_{AB})^2 = -\frac{1}{2} \frac{\partial C}{\partial z_{ins}} (V_g - V_{CPD})^2 \quad (2.16)$$

where C is the capacitance and z_{ins} is the insulator thickness. Derivations of Equations 2.15 and 2.16 are shown in Appendix C.5. Equation 2.16 shows that $V_{AB}(V_g)$ varies linearly (given that V_{CPD} is constant), and $F_{AB}(V_g)$ varies parabolically. The curvature of the $F_{AB}(V_g)$ parabola (called the Kelvin parabola) is proportional to the capacitive gradient. The peak bias corresponds to V_{CPD} , so if $V_g = V_{CPD}$, $F_{AB} = 0$.

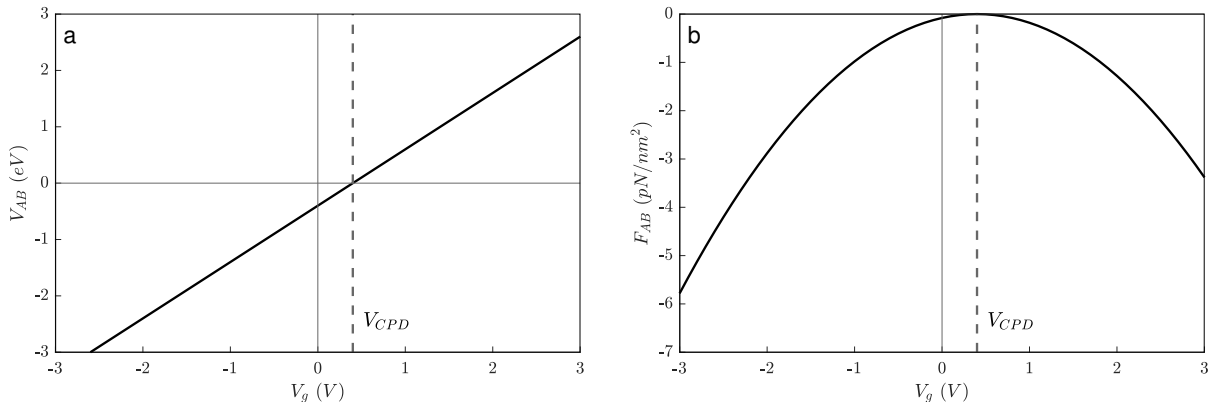


Figure 2.5: **The MIM capacitor potential and force.** Modelled curves for an MIM capacitor with $\partial C/\partial z = 0.1$ pF/m, $V_{CPD} = 0.4$ V, and $V_{SD} = 0$. (a) MIM potential $V_{AB} = V_g - V_{CPD}$ and (b) force F_{AB} , given in Equation 2.16 (the force and capacitive gradient are given per unit area of the capacitor). When $V_g = V_{CPD}$, $V_{AB} = 0$ and $F_{AB} = 0$.

If V_{SD} is nonzero with zero penetration of the surface charge density into the metallic bulk, U_{AB} can be found following the same procedure as in the derivation in Appendix C.5, where $V_{SD} = 0$. Accounting for the additional surface charge density, $U_{AB} = \frac{1}{2}C(V_{AB} - V_{SD})^2$. Assuming that V_{SD} does not depend on z_{ins} , $F_{AB} = -\frac{1}{2}\frac{\partial C}{\partial z_{ins}}(V_g - V_{CPD} - V_{SD})^2$, and the Kelvin parabola peak corresponds to $V_{CPD} + V_{SD}$. If the spatially localized surface charge density is non-linearly bias dependent, $F_{AB}(V_g)$ exhibits some non-parabolicity.

2.4 The MIS capacitor

The metal-oxide-semiconductor (MOS) capacitor, specifically a MOS capacitor comprised of metallically doped Si, a SiO_2 insulating gap, and doped Si, is the most widespread microelectronic structure in existence today. This is because the MOS capacitor is a central component of the MOS field effect transistor (MOSFET), which numbers in the billions in modern microelectronic circuits. While the description “MOS” capacitor indicates the specific case in which the insulator is an oxide, its more general form is the MIS capacitor, in which the insulating layer is comprised of any neutral medium which prohibits charge transport between the metallic and semiconducting plates. In this thesis (to be discussed in Chapter 3), the insulator always consists at least in part of a vacuum gap, so the generic title “MIS” capacitor is used throughout.

The electrostatic characteristics in each of the three materials comprising the MIS capacitor are completely distinct. The nature of the potential, electric field, and charge in metals and insulators were discussed in Section 2.3. In a semiconductor, the density of mobile carriers is nonzero but much smaller than in the metal (due to the finite $g(E)$ and the width of $F_{FD}(E)$, see Figure 2.3), meaning that the field penetrates some distance into the semiconductor before it is completely screened. This leads to a spatially varying potential (V_z), electric field ($\mathcal{E}_z$), and charge (Q_z), which is an effect called band bending. (This derivation is shown in e.g. [1, 4].)

2.4.1 Band bending

In an MIS capacitor, the number of electrons and holes near the semiconductor surface ($n_z(z)$ and $p_z(z)$) is z -dependent due to the z -dependent electric field, such that the total charge density $\rho_z(z)$ is:

$$\rho_z(z) = |e| (p_z(z) - n_z(z) - N_a + N_d) \quad (2.17)$$

where

$$n_z(z) = n \exp\left(\frac{|e|V_z(z)}{k_B T}\right) \quad (2.18a)$$

$$p_z(z) = p \exp\left(\frac{-|e|V_z(z)}{k_B T}\right) \quad (2.18b)$$

where $V_z(z)$ is the spatially variable potential and n and p are the bulk carrier densities given by Equation 2.10. Substituting Equations 2.17 and 2.18 into the one-dimensional Poisson equation ($-\frac{\partial^2 V(z)}{\partial z^2} = \frac{\rho_z(z)}{\epsilon}$) gives the spatially variable electric field:

$$\mathcal{E}_z^2(z) = \frac{k_B T}{\epsilon} \left[p \left(\exp\left(\frac{-|e|V_z(z)}{k_B T}\right) + \frac{|e|V_z(z)}{k_B T} - 1 \right) + n \left(\exp\left(\frac{|e|V_z(z)}{k_B T}\right) - \frac{|e|V_z(z)}{k_B T} - 1 \right) \right] \quad (2.19)$$

Equation 2.19 is derived in Appendix C.6. The spatially variable charge per unit area $Q_z(z)$, according to Gauss' Law (given the symmetry of the capacitor), is:

$$Q_z(z) = -\epsilon \mathcal{E}_z(z) \quad (2.20)$$

The total potential drop across the semiconductor surface is $V_z(z=0) = V_S$. (Correspondingly, the electric field and charge at the surface are $\mathcal{E}_S = \mathcal{E}_z(z=0)$ and $Q_S = Q_z(z=0)$). Therefore, the potential continuity equation for the MIS capacitor is:

$$0 = V_g - V_{CPD} - V_S - V_{ins} \quad (2.21)$$

where $V_{ins} = -eQ_S/C_{ins}$ is the potential drop across the insulator (if the insulator is vacuum, such that $\epsilon_{ins} = \epsilon_o$, then $C_{ins} = \epsilon_o/z_{ins}$) and V_{CPD} is the difference in the chemical potential of the metal and semiconductor when their vacuum levels are aligned. The work function of

a semiconductor, like that of a metal, is the energy required to remove one electron from the surface to vacuum (i.e $\phi_S = E_{vac} - E_f$). So, for an n-type and p-type MIS capacitor, V_{CPD} can be written as:

$$V_{CPD,n} = \frac{\phi_M - \phi_S}{-e} = \frac{\phi_M - [\chi + (E_c^B - E_f)]}{-e} \quad (2.22a)$$

$$V_{CPD,p} = \frac{\phi_M - \phi_S}{-e} = \frac{\phi_M - [\chi + E_g/2 - (E_v^B - E_f)]}{-e} \quad (2.22b)$$

where χ is the semiconductor electron affinity ($E_{vac} - E_v$) and E_c^B and E_v^B are the bulk conduction and valence band energies. (Equations 2.22a and 2.22b are actually equivalent, but are written this way according to the norm.)

The generic solution to V_S is found numerically by solving Equation 2.21, given Equations 2.19 and 2.20 at $z = 0$ (and given the bulk n , p , and E_f defined in Section 2.2). The potential and position inside the semiconductor are related by $\mathcal{E} = -\frac{\partial V}{\partial z}$, such that for any given potential V'_z , the corresponding position z' is:

$$z' = \int_{V_S}^{V'_z} \frac{1}{\mathcal{E}(V)} \partial V \quad (2.23)$$

Then, at this z' , $\mathcal{E}'_z$, and Q'_z are known according to Equations 2.19 and 2.20. So, once V_S is known, the full spatially dependent band diagram, $V_z(z)$, $\mathcal{E}_z(z)$, $Q_z(z)$ can be calculated.

2.4.2 Bias regimes

Figure 2.6 shows band diagrams and $\mathcal{E}_z(z)$ and $Q_z(z)$ of an n-type and p-type capacitor at various V_g , with the potentials of Equation 2.21 labelled. In the MIS capacitor, the chemical potential difference between the capacitor plates is $V_{AB} = V_S + V_{ins}$. The value of V_g for which there is no charge on the capacitor plates (corresponding to $V_{AB} = 0$) is called the flatband potential V_{FB} , so named because at $V_g = V_{FB}$ there is no band bending ($V_S = 0$). In the ideal MIS capacitor (where $V_{CPD} = 0$), $V_{FB} = 0$. In the more general case where $V_{CPD} \neq 0$, $V_{FB} = V_{CPD}$. The flatband condition is illustrated in Figure 2.6(a,e) for an n-type and p-type capacitor, respectively.

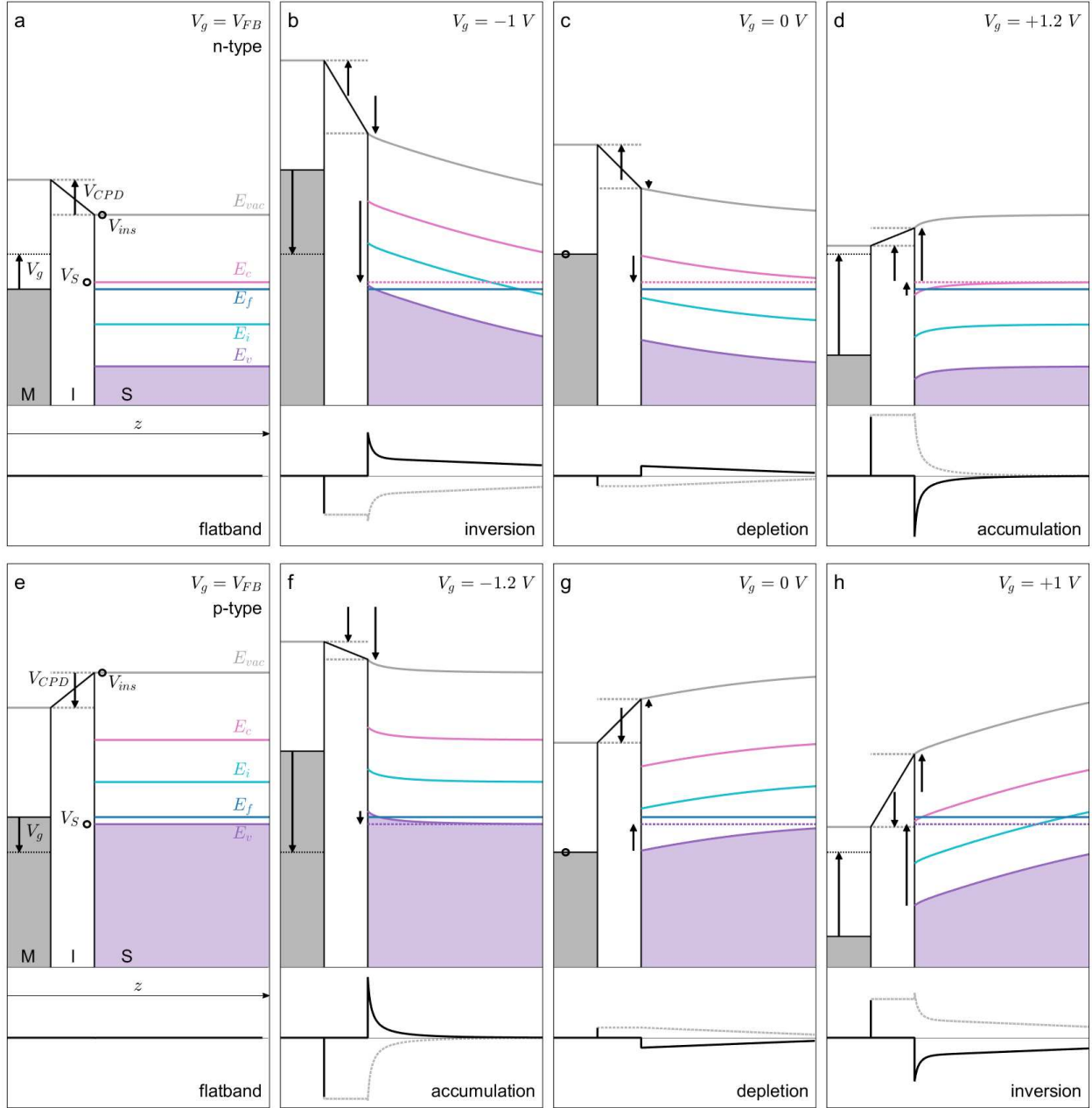


Figure 2.6: **Potential continuity of the MIS capacitor.** Modelled band diagrams (top) and spatially variable electric field $\mathcal{E}_z$ (bottom, dashed) and charge Q_z (bottom, solid) for an n-type (a-d) and p-type (e-h) MIS capacitor at various V_g . The potentials of Equation 2.21 are indicated with black arrows (or, if the potential equals zero, a black circle). In each band diagram, clockwise from the gate they are: V_g , V_{CPD} , V_{ins} , and V_S . The n- and p-type models correspond to: $z_{ins} = 1 \text{ nm}$, $E_g = 1 \text{ eV}$, $\epsilon = 1$, $\phi_M = 1.3 \text{ eV}$, $\chi = 0.8 \text{ eV}$, $m_e = 1$, $m_h = 1$, $T = 300 \text{ K}$, $N_{d,a} = 10^{18}/\text{cm}^3$, and $N_{a,d} = 0/\text{cm}^3$. The y-scales are the same for all band diagrams ($[0 : 4.8] \text{ eV}$), E_z ($[-0.7 : 0.7] \text{ V/nm}$), and Q_z ($[-0.04 : 0.04] \text{ e/nm}^2$), and the x-scale for all z is ($[-2 : 4] \text{ nm}$).

In an extrinsic semiconductor, as was discussed in Section 2.2.2, the Fermi level is located either near the conduction band edge (for an n-type semiconductor) or the valence band edge (for a p-type semiconductor). Consequently, there is an inherent asymmetry in the carrier populations in the conduction and valence bands. Due to this asymmetry, there is also an inherent asymmetry in the carrier organization at the semiconductor surface of an MIS capacitor depending on whether the bands bend up or down. An intuitive explanation of these different characteristic V_z , $\mathcal{E}_z$, and Q_z spatial dependencies on V_g are provided below for an n-type MIS capacitor.

At biases where $V_g > V_{FB}$ (Figure 2.6d), $V_S < 0$ and the bands bend downward. There is a large increase in the electron population at the surface, since the conduction band edge at the surface (E_c^S) quickly passes into the nonlinear regime of $f_{FD}(E)$. This results in a large accumulation of electrons in the conduction band, and consequently, the potential V_S drops off over a very small distance near the semiconductor surface. This is called the accumulation regime.

At biases where $V_g < V_{FB}$ but the band bending is minimal ($E_i^S < E_f$, Figure 2.6c), $V_S < 0$ and the bands bend upward. However, in this case the valence band edge at the surface E_v^S does not quickly pass into the nonlinear regime of $f_{FD}(E)$, since E_f is far from E_v . Consequently, there is a reduced capability for holes to accumulate at the valence band edge, and so the potential V_S drops off over a large distance near the semiconductor surface. This is called the depletion regime (since there is a region near the surface which is depleted of electrons), and the depth of this depletion layer is called the depletion width w_d . As the negative bias increases, w_d increases.

When $V_g < V_{FB}$ and upward band bending ($V_S < 0$) is appreciable (Figure 2.6b), the difference between the Fermi level and the conduction band edge at the surface increases, meaning that the population of electrons at the surface is smaller than the population of holes at the surface. The layer of holes (i.e. minority carriers) at the surface represents an inversion in the dominant carrier type as compared to the bulk, so this case is called the inversion regime. In the intermediate case where $E_i^S > E_f$ but $E_v^S < E_f$, this population of surface holes is minimal, and w_d continues to increase with negative bias. This is the weak inversion regime. When there is significant band bending such that $E_v^S > E_f$, the population of surface

holes increases rapidly, and w_d is approximately constant as the negative bias increases. This is the strong inversion regime.

As mentioned previously, the descriptions above of the MIS bias regimes (accumulation, depletion, weak inversion, and strong inversion) are for an n-type semiconductor, in which the majority carrier is electrons. In a p-type semiconductor, the Fermi level is near the valence band edge and the majority carrier is holes. So, for a p-type MIS capacitor, the accumulation regime occurs when the bands bend upward ($V_S < 0$ and $V_g < V_{FB}$, Figure 2.6f) and there is accumulation of holes at the surface; depletion occurs when $V_g > V_{FB}$ but $E_i^S > E_f$ (Figure 2.6g), $V_S > 0$ and the bands bend downward, such that there is a region near the surface of depth w_d which is depleted of holes; in the weak inversion regime $E_i^S < E_f$ but $E_c^S > E_f$ and there is a small population of electrons at the surface; and in the strong inversion regime (Figure 2.6h), $E_c^S < E_f$ and there is a large population of electrons at the surface.

The bias regimes of the MIS capacitor can also be understood given the characteristic capacitance curves shown in Figure 2.7. In the accumulation and strong inversion regimes, there is an exponential relationship between $|Q_S|$ and V_S . This is due to the exponential relationship between charge and potential in Equation 2.18: Given that w_d is approximately constant in these regimes, most of the surface charge organization due to the potential drop

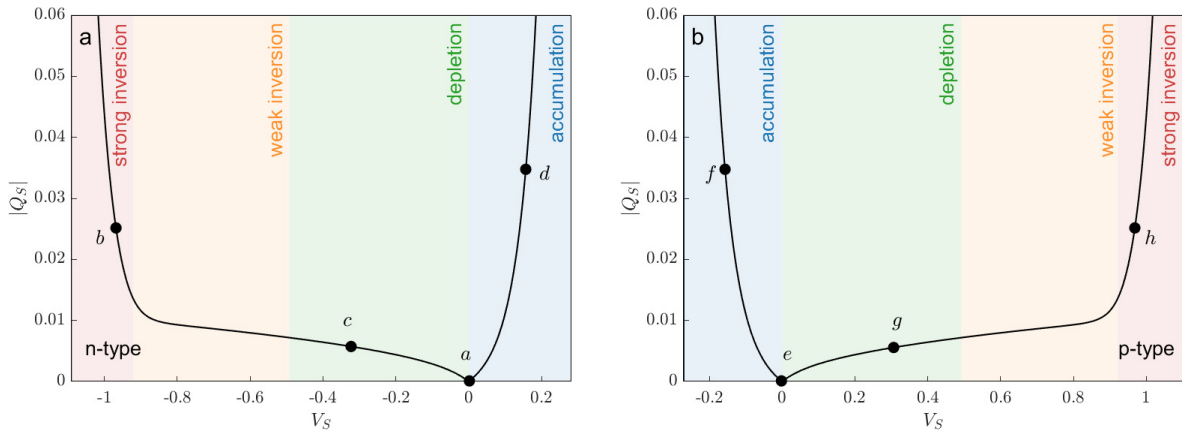


Figure 2.7: **MIS capacitance curves.** $|Q_S| - V_S$ curves for n-type (a) and p-type (b) MIS capacitors, where the interfacial capacitance $C_{int} = \partial Q_S / \partial V_S$. The labelled points correspond to (a-h) in Figure 2.6. The accumulation, depletion, weak inversion, and strong inversion regimes are coloured blue, green, yellow, and red, respectively. The capacitor parameters here correspond to those of Figure 2.6.

across the semiconductor (V_S) occurs at the surface (Q_S). (Note that the validity of the Maxwell-Boltzmann approximation in Equation 2.10 deteriorates as the absolute value of V_g increases, in strong and weak inversion. These effects are not considered here.) However, in the depletion and weak inversion regimes, w_d increases significantly with bias, meaning the charge $n_z(z)$ or $p_z(z)$ is not concentrated at the surface, so the exponential relationship between $|Q_S|$ and V_S is reduced.

2.4.3 MIS capacitor force

The potential continuity equations for the MIM capacitor (Equation 2.12) and the MIS capacitor (Equation 2.21) are in some sense identical. In Sections 2.3.1 and 2.3.2, the influence a hypothetical surface dipole V_{SD} could have on the MIM capacitor potential and force was briefly discussed. For the MIS capacitor, $V_{AB} = V_{ins} + V_S$, and V_S , therefore, is a specific case of V_{SD} . In the MIS capacitor, the surface charge density is not localized exactly at the semiconductor surface, but rather penetrates some distance into the semiconductor bulk, meaning that Equation 2.14 does not simplify to Equation 2.15. Following the result shown in Hudlet (1995)[5], the energy of a one-dimensional MIS capacitor is:

$$U = \frac{Q_M V_M}{2} + \frac{Q_S V_S}{2} + \frac{\epsilon}{2} \int_{V_S}^0 \frac{\partial V(z)}{\partial z} \partial V \quad (2.24)$$

where the subscripts M and S for Q and V denote the charge and potential of the metal and semiconductor surfaces, respectively. The force between the MIS capacitor plates is:

$$F = \frac{-Q_S^2}{2\epsilon_o} \quad (2.25)$$

Derivations of Equations 2.24 and 2.25 are provided in Appendices C.7 and C.8.

In comparison with the linearly V_g -dependent MIM potential and parabolic MIM force shown in Figure 2.5, the MIS potential is nonlinear with V_g , and consequently the force is non-parabolic. These bias dependencies are shown for an n-type and p-type MIS capacitor in Figure 2.8. More precisely, the $F(V_g)$ relationship for the MIS capacitor depends on its bias regime: In the accumulation and strong inversion regimes, $V_S(V_g)$ is approximately

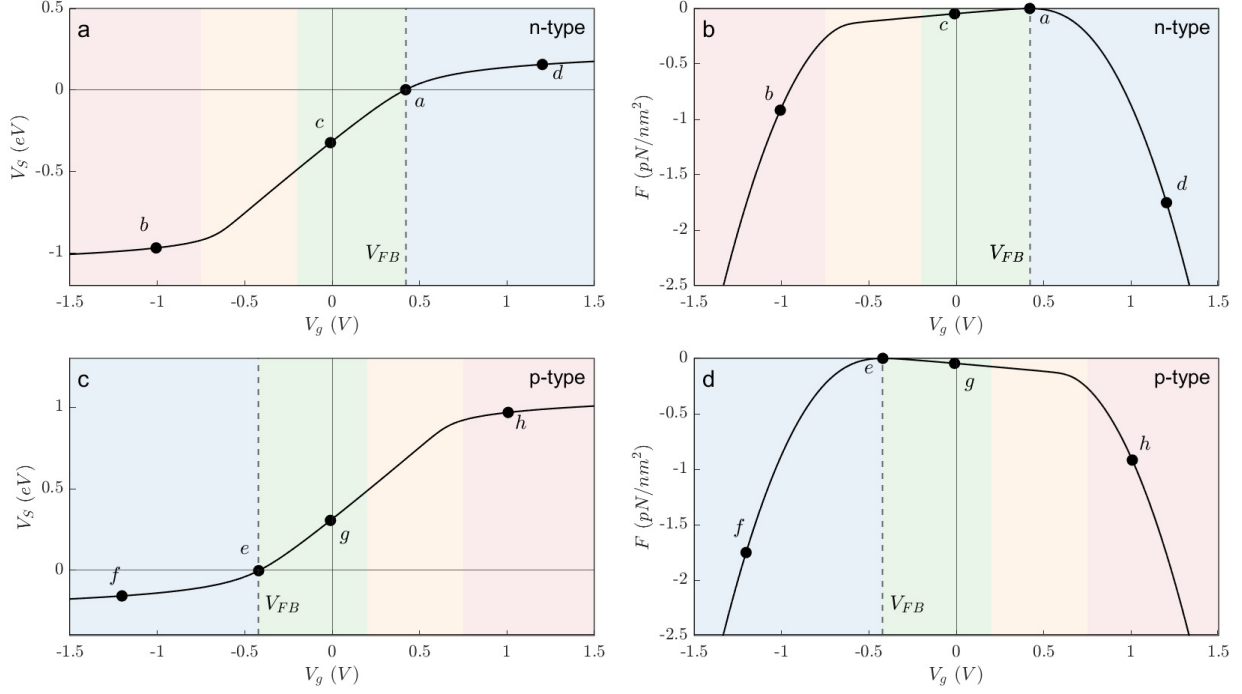


Figure 2.8: **The MIS capacitor surface potential and force.** Modelled bias dependencies corresponding to the n-type (a-d) and p-type (e-h) MIS capacitor in Figures 2.6 and 2.7. (a,c) MIS surface potential V_S and (b) force F . When $V_g = V_{FB}$, $F = 0$. As compared to the MIM capacitor (Figure 2.5), the MIS force has a non-parabolic bias dependence. The accumulation, depletion, weak inversion, and strong inversion bias regimes are coloured blue, green, yellow, and red, respectively. The capacitor parameters here correspond to those of Figure 2.6.

linear, so $F(V_g)$ is approximately parabolic; in the depletion and weak inversion regimes, the nonlinearity of $V_S(V_g)$ is maximized, and therefore so is the $F(V_g)$ non-parabolicity. When $V_g = V_{FB}$, $V_S = 0$ and $V_{AB} = 0$, meaning that $Q_S = -Q_M = 0$ and so $F = 0$.

The capacitor parameters for Figures 2.6, 2.7, and 2.8 are the same (listed in the caption of Figure 2.6), so they can be directly compared to each other. Most are “reasonable” values (E_g , m_e , m_h , T , N_d , and N_a) for typical semiconductors, but some are less realistic, and were chosen instead so that the semiconductor physics is easier to see on these axes. ϕ_M and χ , for example, tend to be about half an order of magnitude larger, but smaller values were chosen so that the total y-scale could be reduced in Figure 2.6. More significantly, the value of z_{ins} is much smaller than the experimental values that will be shown throughout this thesis – in fact, at such a small z_{ins} , various approximations in the measurement operating principle suffer. This will be revisited in Chapters 3 and 4. The extreme z_{ins} shown here essentially

leads to a stretch along the horizontal axis of the $V_S(V_g)$ curves in Figure 2.8, resulting in a much more extreme nonparabolicity in $F(V_g)$. This will be revisited in Chapters 4 and 7.

References

- [1] Sze, S. M. and Ng, K. K. *Physics of semiconductor devices*. 3rd ed. John Wiley & Sons, Inc, **2007**. Chap. 1,4. ISBN: 9780470068328. DOI: 10.1002/0470068329.
- [2] Ashcroft, N. W. and Mermin, N. D. *Solid state physics*. Harcourt College Publishers, **1976**. ISBN: 0030839939.
- [3] Robert F. Pierret. *Semiconductor fundamentals: Volume I*. 2nd ed. Pearson Publishing, **1988**. ISBN: 9780470068328. DOI: 10.1002/0470068329.
- [4] Dieter K. Schroder. *Advanced MOS devices*. Addison-Wesley Publishing Company, Inc, **1987**. ISBN: 9780470068328. DOI: 10.1002/0470068329.
- [5] Hudlet, S., Jean, M. S., Roulet, B., Berger, J., and Guthmann, C. “Electrostatic forces between metallic tip and semiconductor surfaces”. *Journal of Applied Physics* 77.7 (**1995**). DOI: 10.1063/1.358616.

Chapter 3

Frequency-modulated atomic force microscopy

The question *What is “contact”?* seems immediately obvious, but on reflection is non-obvious and even arbitrary. Are two objects in contact if there is a force interaction between them? If so, then all materials in the universe are in constant contact with one another, because the electromagnetic and gravitational forces have infinite range. Are materials in contact if they have significant wave function overlap? If so, what is the threshold for “significant”? Though elusive, the concept of “contact” must be real in some sense. After all, solids do not pass through one another: At some point, it is effectively impossible to push two objects closer together. This effect, macroscopically, is the “normal force”, which arises specifically when two solid objects are in “contact”.

One definition of “contact” considers the cumulation of forces acting between two solid objects as they are brought together from an infinitely large separation distance. At atomic-scale separations there is appreciable interference of their electron wave functions, leading to the creation of new orbitals which might have a higher and/or lower energy with respect to the energies when the objects are isolated. Due to the Pauli exclusion principle, only two electrons of opposite spin can occupy the lowest-energy states, so as the two materials are brought closer and closer together and more hybrid states are created, the electrons need

to occupy higher and higher levels. At some point, the average electron energy is much larger than when the two objects are isolated, so much so that the energy cost to continue to push the materials together diverges to infinity. “Contact” occurs at the distance where this effect, called Pauli repulsion, dominates all other forces. Pauli repulsion is the origin of the macroscopic “normal force”.*

The concept of “contact” is integrally significant for atomic force microscopy (AFM). The basic principle of AFM is to bring a probe (tip) close to a sample surface and observe the net force between them. There are three main AFM operating modes: Contact mode, in which the tip is very close to the sample, where the tip-sample force ($F_{ts}(z)$) is repulsive; tapping mode, where the tip oscillates between the attractive and repulsive regimes; and non-contact mode, in which the tip is comparatively far from the sample, where attractive forces (assuming a net charge-neutral, non-magnetic system) dominate. The dominant forces in contact mode are short-range forces (such as Pauli repulsion, the van der Waals force, chemical forces, adhesion, and the capillary force), whereas in non-contact atomic force microscopy (nc-AFM), long-range forces (such as electrostatic and magnetic forces) dominate. The AFM results in this thesis were measured in the non-contact regime, so $F_{ts}(z)$ is due

* Pauli repulsion is not technically a force, since it is not mediated by force carrier exchange, but it tends to be identified as one for convenience, and is here labelled F_P .

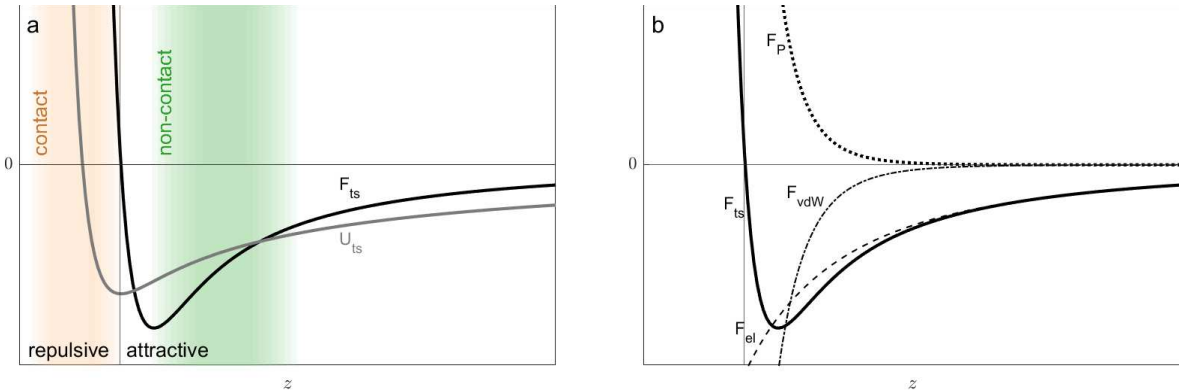


Figure 3.1: **AFM approach curve.** (a) The total distance-dependent F_{ts} and interaction energy U_{ts} , indicating the repulsive ($F_{ts} > 0$) and attractive ($F_{ts} < 0$) regimes. “Contact” and “non-contact” AFM operation regimes are highlighted. (b) The total tip-sample force F_{ts} , which depends on the tip-sample separation distance z , is the sum of different force contributions, in this case an attractive electrodynamic force F_{el} , an attractive van der Waals force F_{vdW} , and Pauli repulsion F_P .

to a long-range capacitive electrodynamic force $F_{el}(z)$, an intermediate-range van der Waals force $F_{vdW}(z)$, and a short-range Pauli repulsion force $F_P(z)$, as shown in Figure 3.1. The tip-sample separation z was large, in the regime where F_{el} dominates, so for the remainder of this thesis, contributions from F_{vdW} and F_P (and all other forces) are assumed to be zero.

In the most common version of an AFM apparatus, the tip is situated at the end of a flexible cantilever. The tip is small, so by scanning it over the surface, a map of the spatial nonhomogeneity of the tip-sample interaction $F_{ts}(x, y, z)$ can be measured by detecting the deflection of the cantilever as it interacts with the sample. (This is discussed in greater detail in Section 3.1.1.) The forces measured in contact mode AFM are generally large, and so can be detected simply by measuring the static deflection of the cantilever. In nc-AFM, however, the forces are generally too small to be detected as a static measurement. Therefore, in nc-AFM, the cantilever is oscillated, and variations in the oscillation amplitude or frequency are used to determine $F_{ts}(x, y, z)$. The version of nc-AFM used for all AFM measurements in this thesis is frequency-modulated AFM (fm-AFM), in which the oscillation amplitude is held constant and variations in F_{ts} lead to variations in the oscillation frequency[1]. The mechanism by which F_{ts} is measured using fm-AFM is described in Section 3.1.

3.1 fm-AFM operating principle

There are two common descriptions of the fm-AFM operating principle. The difference between them is central to this thesis, so both are presented here for clarity. In both, the fm-AFM cantilever is described as a damped, driven harmonic oscillator experiencing an external tip-sample force $F_{ts}(z, t)$:

$$m\ddot{z} + \xi\dot{z} + kz = F_{drive} + F_{ts}(z, t) \quad (3.1)$$

where m is the cantilever mass, ξ is the damping coefficient, k is the spring constant, and F_{drive} is the driving force. This is an experimental thesis, so re-written in terms of experimentally measurable variables (the free cantilever resonance frequency $\omega_o = \sqrt{\frac{k}{m}}$ and quality

factor $Q = \frac{m\omega_o}{\xi}$, see Figure 3.2), Equation 3.1 becomes:

$$\frac{k\ddot{z}}{\omega_o^2} + \frac{k}{Q\omega_o}\dot{z} + kz = F_{drive} + F_{ts}(z, t) \quad (3.2)$$

When the cantilever is very far from the sample surface, $F_{ts}(z, t) \approx 0$, the cantilever is “free” and the frequency-dependent amplitude and phase are related according to the equations derived in Appendix C.9. When $\omega = \omega_o$, A is maximized and $\theta = -90^\circ$. Modelled and measured free resonance curves of a typical fm-AFM cantilever are shown in Figure 3.2.

In fm-AFM, the cantilever oscillates at its resonance frequency ω , and the tip motion is assumed to be sinusoidal with constant amplitude A , such that $z = z_o + A \cos(\omega t)$. The amplitude is maintained using a self-excitation loop which applies a driving force at the resonance frequency that is 90 degrees out-of-phase with the position, such that $F_{drive} = F_d \sin(\omega t)$. (The controllers required to maintain this oscillation will be described in Section 3.1.1.) In the original derivation of fm-AFM[2–4], the driving force is assumed to exactly compensate for cantilever damping, and the tip-sample force is assumed to be in-phase with z , such that

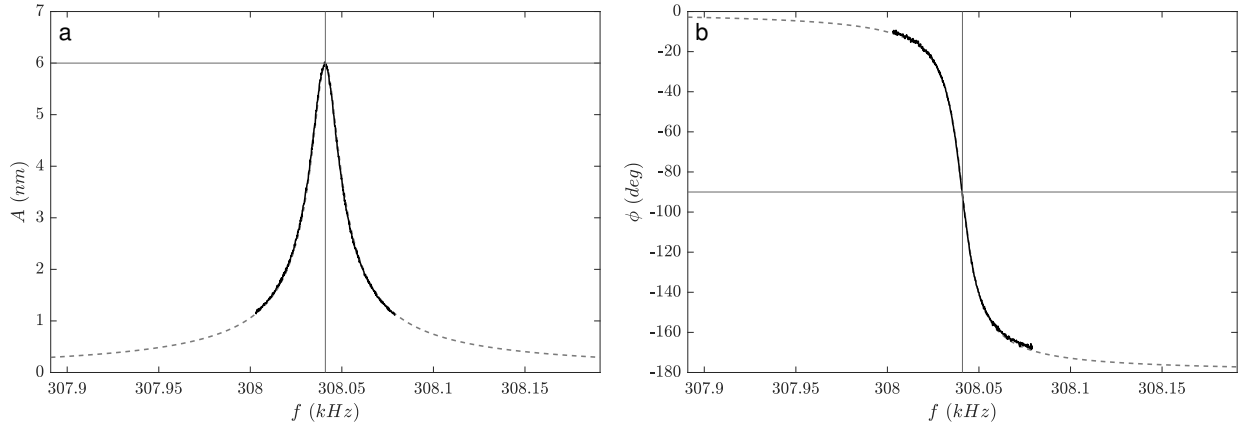


Figure 3.2: **Free resonance curves.** (a) Amplitude and (b) phase of a free ($F_{ts} = 0$) damped, driven harmonic oscillator. The dashed lines are modelled curves according to Equation 3.2 (specific equations given in Appendix C.9), and the solid lines are the experimental curves. This measurement is done for each cantilever to determine Q and ω_o by fitting. k is estimated by the cantilever manufacturers. For this cantilever, $k \approx 42 \text{ N/m}$, $Q \approx 20000$, and $\omega_o \approx 300 \text{ kHz}$.

Equation 3.2 becomes:

$$\frac{k}{\omega_o^2} \ddot{z} + kz = F_{ts}(z) \quad (3.3a)$$

$$\frac{k}{Q\omega_o} \dot{z} = F_d \sin(\omega t) \quad (3.3b)$$

$F_{ts}(z)$ is then expressed as a Taylor series:

$$F_{ts}(z) = F_{ts}(z_o) + \frac{\partial F_{ts}(z_o)}{\partial z}(z - z_o) + \frac{1}{4} \frac{\partial^2 F_{ts}(z_o)}{\partial z^2}(z - z_o)^2 + \dots \quad (3.4)$$

and the oscillation amplitude A is assumed to be small, such that $F_{ts}(z)$ is approximately linear (i.e. $\frac{\partial^n F_{ts}(z_o)}{\partial z^n} = 0$ for $n > 1$).

The shift in the cantilever resonance ($\Delta\omega = \omega - \omega_o$) due to $F_{ts}(z)$ is found by writing Equation 3.3a in terms of an effective spring constant $k_{eff} = k - \frac{\partial F_{ts}(z_o)}{\partial z}$ (by collecting powers of z), and taking the instantaneous cantilever resonance frequency $\omega = \sqrt{\frac{k_{eff}}{m}}$ and $1 - \sqrt{1 - x} \approx \frac{x}{2}$. The drive F_d required to maintain A (i.e. compensate for damping) is found given Equation 3.3b by making the approximation that $\omega \approx \omega_o$. So, in this first fm-AFM description,

$$\Delta\omega = -\frac{\omega_o}{2k} \frac{\partial F_{ts}(z)}{\partial z} \quad (3.5a)$$

$$F_d = \frac{kA}{Q} \quad (3.5b)$$

meaning that $\Delta\omega$ is proportional to the z -dependent force gradient and F_d is independent of F_{ts} and is constant, depending on the intrinsic damping properties of the cantilever.

The assumption between Equations 3.2 and 3.3 is that F_{ts} conservative – that is, F_{ts} and z are in-phase. However, this is not generically true. There could be, for example, displacement currents in the sample as the tip oscillates, leading to energy dissipation over every oscillation cycle by Joule heating[5–7]. Structural interactions between the tip and sample, such as adhesion[8] or changes in atomic positions in the tip or sample[9], can also lead to energy dissipation. In such cases, Equation 3.2 does not apply. There are a few ways to account for non-conservative effects in the equation of motion: Including a velocity

dependence in $F_{ts}(z, \dot{z})$ [10]; equivalently, the damping can be said to be dependent on the tip-sample position z , such that $\xi = \xi(z, t)$ [6]; or by using a more generic representation of $F_{ts}(z, t)$ which includes components which are in-phase and out-of-phase with z [11–13]. All of these are variants of the second prominent derivation of fm-AFM, which is more generic in that it accounts for non-conservative tip-sample interactions.

Using the latter approach, in the second fm-AFM description $F_{ts}(t)$ is expressed as a Fourier series with the orthonormal basis of $\sin(t)$ and $\cos(t)$. Since $F_{ts}(t)$ necessarily takes on the periodicity of $z(t)$ (i.e. $T = 2\pi/\omega$), in the approximately linear force regime (i.e. taking only the term at ω), $F_{ts}(t)$ can be expressed as:

$$F_{ts}(t) = F_{in} \cos(\omega t) + F_{out} \sin(\omega t) \quad (3.6a)$$

$$F_{ts}(t) = \left[\frac{\omega}{\pi} \int_0^{2\pi/\omega} F_{ts}(t) \cos(\omega t) \partial t \right] \cos(\omega t) + \left[\frac{\omega}{\pi} \int_0^{2\pi/\omega} F_{ts}(t) \sin(\omega t) \partial t \right] \sin(\omega t) \quad (3.6b)$$

where F_{in} and F_{out} in Equation 3.6a are the amplitudes of the in-phase and out-of-phase force contributions, with the Fourier coefficients written explicitly in Equation 3.6b.

Substituting $z = z_o + A \cos(\omega t)$, Equation 3.2 is:

$$F_{ts}(t) = \left[\frac{kA}{\omega_o^2} (\omega_o^2 - \omega^2) \right] \cos(\omega t) + \left[\frac{kA}{\omega_o^2} \frac{\omega \omega_o}{Q} - F_d \right] \sin(\omega t) \quad (3.7)$$

so that by comparing Equations 3.6 and 3.7 and assuming that $\omega \approx \omega_o$, solutions for $\Delta\omega$ and F_d are:

$$\Delta\omega = \omega - \omega_o = \frac{-\omega_o}{2kA} \frac{\omega_o}{\pi} \int_0^{2\pi/\omega} F_{ts}(t) \cos(\omega t) \partial t \quad (3.8a)$$

$$F_d = \frac{kA}{Q} - \frac{\omega_o}{\pi} \int_0^{2\pi/\omega} F_{ts}(t) \sin(\omega t) \partial t \quad (3.8b)$$

Consequently, components of the tip-sample force which are in phase with the cantilever position shift the cantilever resonant frequency, and the out-of-phase force components lead to an additive contribution to the drive signal. The most obvious distinction between Equa-

tions 3.5 and 3.8 is that F_d , in addition to the constant term kA/Q which is due to intrinsic damping, depends on $F_{ts}(z, t)$.

In the regime where $\omega \approx \omega_o$, the drive amplitude F_d may be expressed as an energy loss per cycle as[13, 14]:

$$E_{ts} = E_o \left[\frac{F_d - F_{do}}{F_{do}} \right] \quad (3.9)$$

where $E_o = \frac{\pi k A^2}{Q}$ is the intrinsic damping loss of a high quality factor oscillator[14] and $F_{do} = \frac{kA}{Q}$ is the intrinsic offset in Equation 4b. Experimental drive signals are sometimes reported in units of Hz , but can also be converted to units of energy loss per cycle given the above expression, where $F_d = A_{exc}$ is the measured drive amplitude in Volts and $F_{do} = A_{exco}$ is the drive amplitude in Volts measured in the absence of a tip-sample interaction (if the $A_{exc} : F_d$ transfer function for the drive piezo excitation is flat, i.e. frequency-independent[15]).

3.1.1 Optical detection and feedback controllers

In fm-AFM, the cantilever oscillates at its resonance frequency ω at a constant setpoint oscillation amplitude A_S . In the JEOL system (the apparatus used for all of the fm-AFM measurements in this thesis, introduced in Appendix A.1), the deflection of the cantilever is measured optically, by deflecting light from a photodiode off of the back of the cantilever and into a four-quadrant photodetector (FQ-PD). The FQ-PD measures the cantilever deflection by comparing the illumination intensity in the four quadrants of the photodetector: For example, when labelling the quadrants in clockwise direction starting from the top left as Q_1 , Q_2 , Q_3 , and Q_4 , then $(Q_1 + Q_2) - (Q_3 + Q_4)$ reveals the vertical deflection of the cantilever and $(Q_1 + Q_3) - (Q_2 + Q_4)$ reveals the horizontal deflection of the cantilever. Since in fm-AFM the cantilever is oscillating (and presumed to be oscillating vertically, with horizontal deflections due to lateral tip-sample forces approximately zero – the cantilever stiffness is orders of magnitude higher for its torsional modes[16]) the $(Q_1 + Q_2) - (Q_3 + Q_4)$ signal varies at ω with an amplitude related to the cantilever oscillation amplitude A .

The cantilever oscillation is controlled using a phase-locked loop (PLL) and an amplitude controller. A PLL compares the phase of a reference signal to the phase of the input signal: If the phase difference (θ) is constant, the signals necessarily have the same frequency. In

fm-AFM, the PLL is used to measure $\Delta\omega = \omega - \omega_o$ (which is due to a non-zero $F_{ts}(z)$ as shown in Equation 3.8). The amplitude controller maintains a constant oscillation setpoint A_S by applying a periodic drive at ω to the drive piezo. The drive is maintained at 90° out-of-phase with the tip position using a phase shifter, so that the system is necessarily always on resonance. The drive amplitude, which incorporates the voltage: deflection calibration constant of the piezo, is F_d . The oscillation control system therefore has four outputs: $\Delta\omega$ and F_d provide information about $F_{ts}(z)$, and θ and A are constant and equal to 90° and the setpoint amplitude A_S , respectively. For the remainder of this thesis, it is taken as true that $\theta = 90^\circ$ and $A = A_S$.

The second fm-AFM feedback mechanism is a proportional-integral (PI) controller, which is used to control the tip-sample separation z . A PI controller compares a process variable to a setpoint, and works to keep the process variable as close to the setpoint as possible. To do this, it continually applies a correction to the error value $e(t)$ (difference between the setpoint and process variable). The P-controller accounts for proportional errors (i.e. it works to keep a constant offset), and the I-controller brings the integral of $e(t)$ with respect

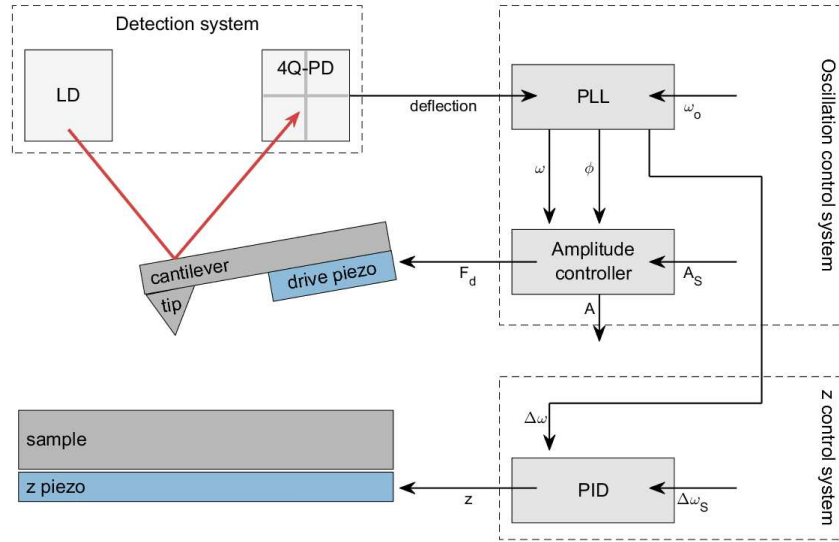


Figure 3.3: **fm-AFM operation systems.** Simplified block diagram of the fm-AFM detection system, oscillation control system, and z -control system. Input parameters are ω_o , A_S , and $\Delta\omega_S$, and the measured variables are $\Delta\omega$ (or ω), F_d , θ , A , and z . Several components of the circuitry, such as amplification of the deflection signal and multiplications by the voltage: deflection piezo constants, are not shown.

to time to zero (i.e. it works to make the offset zero). In the fm-AFM z-control system, the PID controller compares a setpoint frequency shift $\Delta\omega_S$ to $\Delta\omega$. Since, according to Equation 3.8, $\Delta\omega$ is related to $F_{ts}(z)$, $e(t) = \Delta\omega - \Delta\omega_S = 0$ is maintained by adjusting the tip-sample separation z_{ins} .

In fm-AFM experiments, the oscillation control system is always activated, but the z-control system might be on or off depending on the experiment that is being performed. One common fm-AFM experiment is to raster scan the tip over the sample surface, to measure an area spanning the x and y directions. In such experiments, when the z-control system is turned on, $\Delta\omega$ is constant and z reflects variations in F_{ts} over the sample surface. (If the electrostatic force $F_{el} = 0$, F_{ts} is generally assumed to be due to height variations in the sample, which is why these tend to be called topography images. This point will be revisited in Section 3.2.) Another common fm-AFM experiment is to maintain the tip at a constant x, y position and vary a parameter such as the tip-sample bias, V_g . These are sometimes called fm-AFM spectroscopy experiments. In bias spectroscopy experiments, the z-control system is usually turned off, such that z is constant (assuming negligible drift) and $\Delta\omega$, which is related to F_{ts} , varies with V_g . In a grid bias spectroscopy experiment, a bias spectrum is collected for every (x, y) pixel in a specified grid. While the tip moves between pixels, the z-controller is on, but while the bias spectrum is being measured, the z-controller is off. Another kind of measurement shown in this thesis is multipass bias imaging. In this experiment, the tip is passed multiple times over each line in the slow scan direction. On the first pass, the z-controller is on and z is recorded. On the remaining passes, the z-controller is turned off, V_g is set to a different value, and the tip traces the same path z as the first pass, and $\Delta\omega$ is recorded.

3.2 Electrostatic force microscopy

If the dominant force in an fm-AFM experiment is the electrostatic force F_{el} , $F_{ts} \approx F_{el}$ and the tip-sample junction resembles a capacitor. Electrostatic force microscopy (EFM) and Kelvin probe force microscopy (KPFM) are techniques that are used to characterize F_{el} . Standard descriptions of EFM and KPFM, presented here, describe a perfect MIM capacitor

(i.e. V_{SD} , if nonzero, is localized exactly at a surface and does not depend on V_g). Deviations from the MIM capacitor in the case of an MIS capacitor are discussed in Section 3.3.

3.2.1 fm-EFM operating principle

There are many variants of EFM. Universally, however, EFM involves applying a gate bias V_g to a tip or sample in order to characterize F_{el} . In an MIM capacitor, V_{AB} is z_{ins} -independent (Figure 3.4a), and F_{el} , given by Equation 2.16 is z_{ins} -dependent (Figure 3.4c). In fm-EFM (i.e. fm-AFM with an applied bias), as the cantilever oscillates above the surface, the insulator thickness z_{ins} varies sinusoidally in time. (In Figures 3.4a,c, the cantilever oscillates between the dashed lines.) Consequently, due to the z_{ins} -dependent capacitive gradient, F_{el} also varies in time. The constant V_{AB} and time-dependent F_{el} are shown for variable V_g in Figure 3.4b,d.

F_{el} leads to a modification of $\Delta\omega$, according to Equation 3.2. In a perfect MIM capacitor, the carrier mobility is essentially infinite, so the time to charge the capacitor equals zero and there is no lag between tip motion and F_{el} . Given Equation 3.5, F_d is constant and $\Delta\omega$ is:

$$\Delta\omega = \frac{\omega_o}{4k} \frac{\partial^2 C}{\partial z_{ins}^2} (V_g - V_{CPD})^2 \quad (3.10)$$

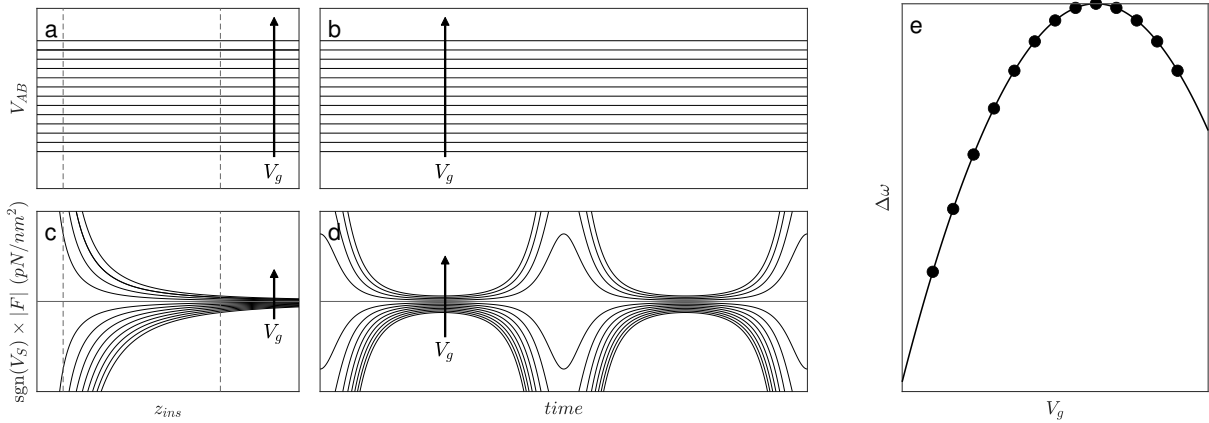


Figure 3.4: The MIM capacitor frequency shift. (a,c) z_{ins} and (b,d) time-dependent potential V_{AB} (a,b) and force F (c,d) of the MIM capacitor, and the resulting bias-dependent frequency shift $\Delta\omega$. Thirteen curves are shown for (a-d) corresponding to the biases indicated in (e), and the order of increasing bias is shown with arrows. All F values are below 0, but in (c,d) $F(V_g > V_{CPD})$ have been reflected above the x-axis so that they are visible. The cantilever is oscillating sinusoidally, such that z_{ins} varies between the dashed lines in (a,c).

Equation 3.10 shows that, like F_{el} , the MIM $\Delta\omega(V_g)$ is parabolic, and the parabola peak corresponds to V_{CPD} , as shown in Figure 2.5b.

In Figure 3.4, at certain biases the assumption that $F_{ts}(z)$ is approximately linear (required for the $\Delta\omega$ derivation, see Section 3.2.1) is untrue. This is clear in Figure 3.4c, where within the dashed lines F_{el} is clearly nonlinear. This manifests as a non-sinusoidal $F_{el}(t)$ in Figure 3.4d, which would violate both $\Delta\omega$ derivations shown in Section 3.1. This highlights the importance of conducting an fm-AFM experiment at a large enough z_{ins} (or a small enough A) that the linear $F_{ts}(z)$ approximation remains valid. The validity of the linear force approximation for the experimental tip-sample separations used in this work, which are much larger than that shown in Figure 3.4, is demonstrated in Chapter 4.

3.2.2 fm-ac-EFM operating principle

In another common variant of EFM, which is here referred to as ac-EFM, an oscillating bias $V_g = V_{DC} + V_{AC} \sin(\omega_{AC}t)$ is applied between the tip and sample. In the case of an MIM capacitor, the potential becomes:

$$V_{AB} = (V_{DC} - V_{CPD}) + V_{AC} \sin(\omega_{AC}t) \quad (3.11)$$

and the electrostatic force becomes:

$$F_{el} = -\frac{1}{2} \frac{dC(z)}{dz} [(V_{DC} - V_{CPD}) + V_{AC} \sin(\omega_{AC}t)]^2 \quad (3.12)$$

such that in fm-ac-EFM, the cantilever frequency shift is $\Delta\omega = \Delta\omega_{DC} + \Delta\omega_{1\omega_{AC}} + \Delta\omega_{2\omega_{AC}}$ (derived in Appendix C.10), where:

$$\Delta\omega_{DC} = \frac{\omega_o}{4k} \frac{\partial^2 C}{\partial z^2} \left(\frac{V_{AC}^2}{2} + (V_{DC} - V_{CPD})^2 \right) \quad (3.13a)$$

$$\Delta\omega_{\omega_{AC}} = \frac{\omega_o}{4k} \frac{\partial^2 C}{\partial z^2} (-2V_{AC}(V_{DC} - V_{CPD})) \cos(\omega_{AC}t) \quad (3.13b)$$

$$\Delta\omega_{2\omega_{AC}} = \frac{\omega_o}{4k} \frac{\partial^2 C}{\partial z^2} \left(\frac{-V_{AC}^2}{4} \right) \sin(2\omega_{AC}t) \quad (3.13c)$$

Figure 3.5 illustrates the various signals involved in an fm-ac-EFM experiment. The signal which modulates the carrier frequency ω_o is F_{el} which, since it is quadratically related

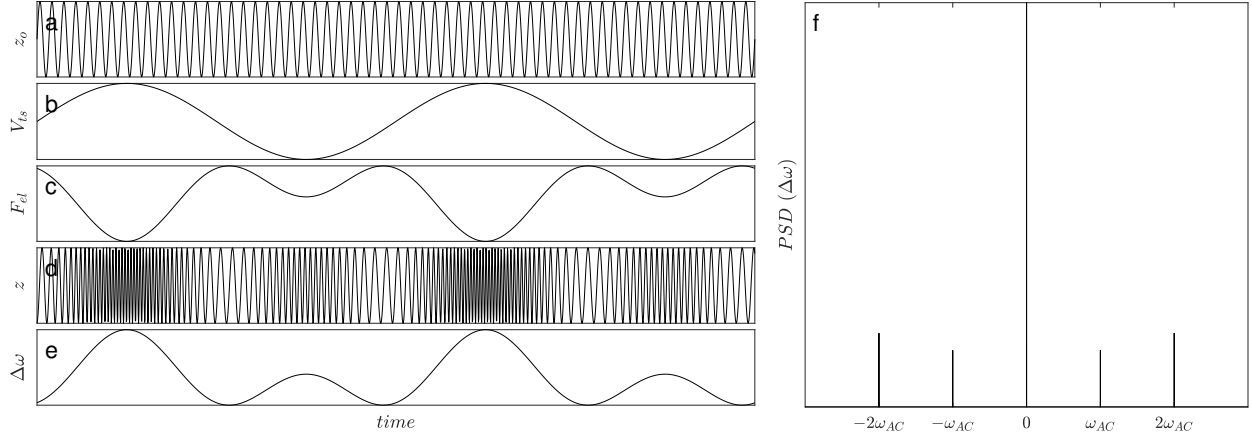


Figure 3.5: **fm-ac-EFM signals.** (a) Free oscillating cantilever position z_o (at its free resonance ω_o); (b) applied bias V_{AB} incorporating V_{CPD} and a modulated $V_g = V_{DC} + V_{AC} \sin(\omega_{AC}t)$; (c) the resulting electrostatic force F_{ts} according to Equation 2.16 (with contributions at $2\omega_{AC}$ due to the square exponent); (d) the resulting cantilever position z with variable instantaneous frequency (derived in Appendix C.10); (e) the resulting $\Delta\omega$, which is the measured fm-EFM signal; (f) Fourier transform of z , showing sidebands.

to V_{AB} , means that sidebands appear in the FFT at $f \pm f_{AC}$ and $f \pm 2f_{AC}$ (where $f = \omega/2\pi$ and $f_{AC} = \omega_{AC}/2\pi$). The sideband amplitudes, according to Equation 3.13, depend on experimental parameters V_{DC} , V_{AC} , k , and ω_o and sample parameters V_{CPD} and $\frac{\partial C}{\partial z}$.

3.2.3 fm-ac-KPFM operating principle

fm-ac-KPFM is a sub-category of fm-ac-EFM which is used specifically to measure V_{CPD} . In fm-ac-EFM, V_{DC} can take any value, but in fm-KPFM, a PID controller is used to set V_{DC} to the value at which the amplitude of the $\Delta\omega_{\omega_{AC}}$ sideband is minimized. At this bias, according to Equation 3.13b, $V_{DC} = V_{CPD}$, so V_{CPD} is known. (If there is a non-zero V_{SD} , the $\Delta\omega_{\omega_{AC}}$ sideband is minimized at $V_{DC} = V_{CPD} + V_{SD}$.) An intuitive demonstration of the bias-dependence of the $\Delta\omega_{\omega_{AC}}$ sideband amplitude is shown in Figure 3.6.

The main benefit of fm-ac-KPFM is that it allows for an accurate measurement of the sample topography. In a standard fm-AFM image with the z -controller on, $\Delta\omega$ is due to the net F_{ts} which, as demonstrated in Figure 3.1, is made up of many different contributions. If V_{CPD} is spatially heterogeneous, then if V_g is held constant, F_{el} is spatially heterogeneous as well. In this situation, the apparent “topography” (measured z signal) is due to a convolution

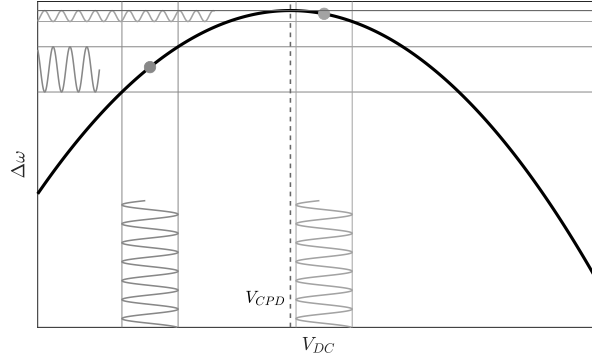


Figure 3.6: **MIM Kelvin parabola.** The amplitude of $\Delta\omega_{\omega_{AC}}$ is shown for two V_g . When $V_g = V_{CPD}$, the sideband amplitude is minimized.

of F_{el} and the true sample topography. Sometimes, F_{el} can be inverse to the sample height, meaning that the apparent “topography” measured in fm-AFM could be inverted. Both of these examples are demonstrated experimentally in Appendix A.2. During an fm-KPFM measurement, however, the Kelvin controller (which is used in conjunction with the fm-AFM oscillation and z -controllers) continually adjusts V_{DC} such that F_{el} is nulled (i.e. V_{DC} is set to correspond to the peak of the Kelvin parabola, see Figure 3.6). Consequently, even if V_{CPD} is spatially heterogeneous, since F_{el} is nulled, the z channel contains only information about the sample topography, so that the measured z (assuming negligible lateral tip-sample interactions) corresponds to the true sample topography. Furthermore, by recording V_{DC} for all (x, y) during a fm-ac-KPFM experiment, a map of the spatial variability of V_{CPD} is simultaneously measured. fm-ac-KPFM is the only type of KPFM measurement shown in this thesis; it will henceforth be referred to simply as KPFM.

Of course, V_{CPD} can also be measured by performing bias spectroscopy, since V_{CPD} corresponds to V_{DC} at the parabola peak. However, since V_{CPD} is not typically constant over the entire surface of non-homogeneous samples, measuring the spatial variability of V_{CPD} using bias spectroscopy would require grid bias spectroscopy, which tends to be slow and therefore prone to drift and limited spatial resolution. KPFM, comparatively, allows for a fast measurement of the V_{CPD} since the entire Kelvin parabola does not need to be measured, and it can be used in conjunction with the z feedback during imaging.

All fm-AFM data shown in this work were measured using a Nanonis control system. The typical oscillation amplitude was 6 nm, and typical setpoint frequencies for the z -controller

were between $-1 : -5 \text{ Hz}$. For KPFM images, the AC amplitudes and frequencies were between $200 : 700 \text{ mV}$ and $200 : 900 \text{ Hz}$. Imaging speeds were $1 : 100 \text{ nm/s}$, such that images took $\sim 2 : 6$ hours to measure. Grid spectroscopy and multipass data were longer measurements which took $\sim 8 : 20$ hours to measure. Bias spectroscopy sweep rates were $\sim 1 \text{ V/s}$ or more (such that each pass of the sweeps took $\sim 20 \text{ s}$ to measure).

3.3 The MIS fm-AFM model

The EFM derivation in Section 3.2.1 assumes that the tip-sample junction behaves like a perfect MIM capacitor. However, as was noted in Section 2.4.3, F_{el} for the MIS capacitor is not given by Equation 2.16, so the standard fm-EFM derivation does not apply. This section outlines the main experimental differences between the MIM and MIS capacitors.

3.3.1 MIS frequency shift

For the MIS capacitor, like for the MIM capacitor, F_{el} is distance dependent, so in an fm-AFM experiment, where the tip oscillates above the sample surface, F_{el} varies in time. Unlike the MIM capacitor, V_S also varies in time. Both of these time dependencies are shown for an n-type and p-type capacitor in Figure 3.7a-d,f-i (which is intended to be directly compared to Figure 3.4, which shows the MIM capacitor time dependencies). In the MIS capacitor, the force expression is not analytical, so $\Delta\omega$ must be found numerically. This is done by first calculating $F_{el}(t)$ over a full oscillation cycle given $z_{ins} = z_o + A \cos(\omega t)$ and integrating $F_{el}(t)$ according to Equation 3.8a. The distance dependence of V_S (Figure 3.7a,f), and correspondingly, the time-dependence of V_S as the cantilever oscillates (Figure 3.7b,g) correspond to the non-linearly distance-dependent band bending at the semiconductor surface. F_{el} also correspondingly depends on band bending. To model the V_g dependence of $\Delta\omega$, this process of calculation of the time-dependent F_{el} and subsequent integration, is repeated for an array of V_g values (i.e. -1.5 to 1.5 V in Figure 3.7e,j). $\Delta\omega(V_g)$, like $F_{el}(V_g)$, is non-parabolic.

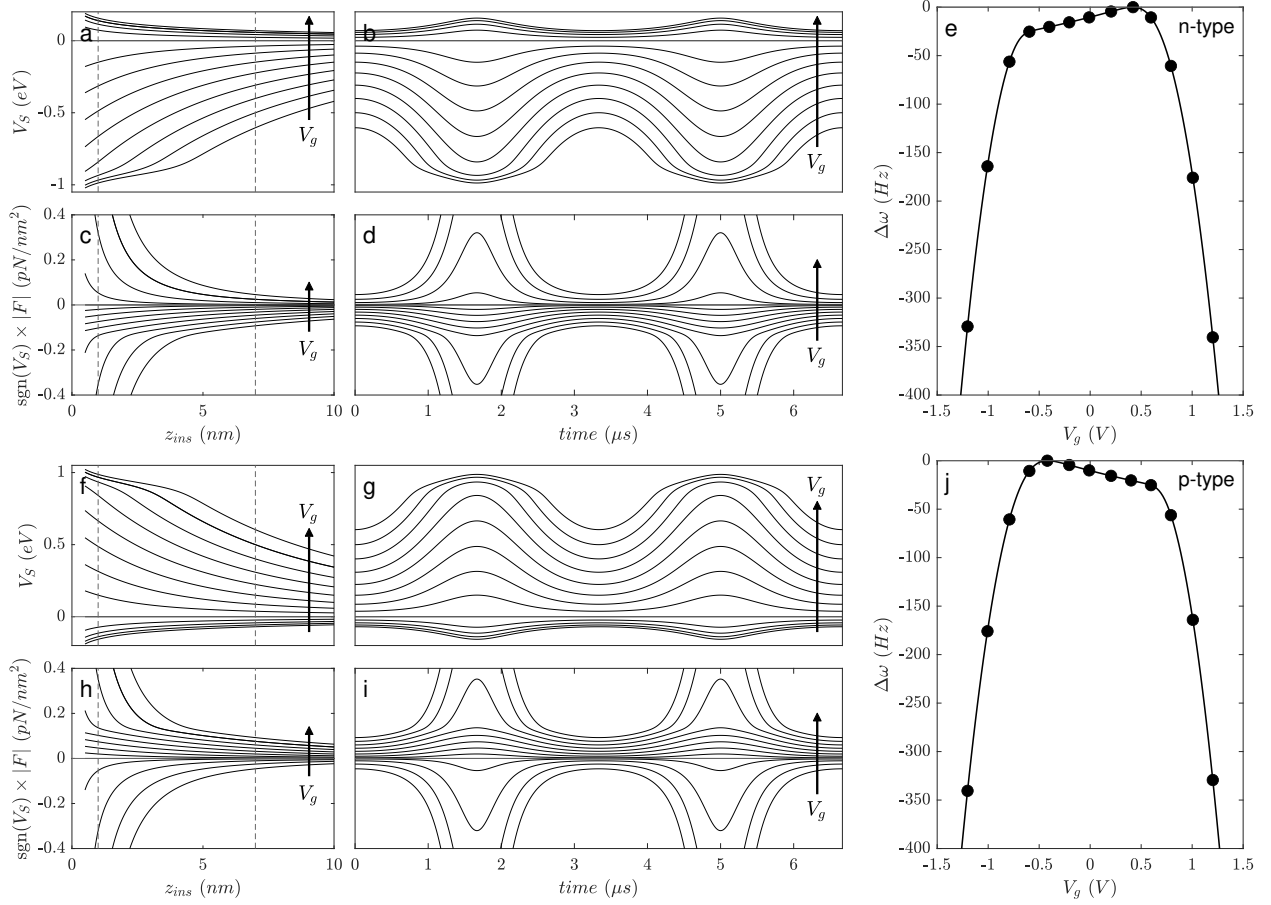


Figure 3.7: **The MIS capacitor frequency shift.** z_{ins} and time dependencies of V_S and F , and the resulting Δf for an n-type (top) and p-type (bottom) MIS capacitor. The capacitor parameters correspond to those of Figures 2.6-2.8. The AFM parameters are: $A = 6 \text{ nm}$, $\omega_o = 300 \text{ kHz}$, $k = 42 \text{ N/m}$, $a = \pi(5 \text{ nm})^2$, and $Q = 18,000$.

3.3.2 MIS excitation

In a semiconductor, the carrier mobility is finite, meaning that there is a finite timescale (τ) to establish the equilibrium V_S in a MIS capacitor. In other words, τ is the capacitor charging time constant, where if the equilibrium discharged capacitor has $V_S = 0$, then:

$$V_{S,discharging}(t) = V_{So}e^{-t/\tau} \quad (3.14a)$$

$$V_{S,charging}(t) = V_{So}(1 - e^{-t/\tau}) \quad (3.14b)$$

where V_{So} is the equilibrium charged surface potential. Given that the carrier organization at the semiconductor surface depends on z_{ins} (recall Figures 3.7a,f), if z_{ins} varies according

to $z_{ins}(t) = z_o + A \cos(\omega t)$ (Figure 3.8a,e), τ leads to a phase shift δ between $z_{ins}(t)$ and $V_S(t)$, according to $\delta = \omega\tau$. Consequently, there is also a phase shift δ between $F_{el}(t)$ and $z_{ins}(t)$ (Figure 3.8b,f). In a fm-AFM experiment, any non-zero δ manifests as an increase in F_d , according to Equation 3.8b (Figure 3.8c-d,g-h). The tip-sample junction, then, could be thought of effectively as an alternating current (AC) - resistor-capacitor (RC) circuit, where the resistance (i.e. the equivalent series resistance) corresponds specifically to carriers moving within w_d . A more rigorous description of τ will be presented in Chapter 4.

The approximate upper bound of measurable τ is imposed by the MIS capacitor model, which describes the static (equilibrium) state of the surface charge organization. If τ is small with respect to the oscillation period, the system can be described as quasi-static – that is, the MIS capacitor is essentially at equilibrium at every time t in its oscillation – so the equilibrium MIS capacitor model from Section 2.4 still applies. This sets $\tau \ll 2\pi/\omega_o$ as an upper limit on the measurable τ using this approach (or, equivalently, the minimum ω_o for the cantilever). The cantilevers used in this work have $\omega_o/2\pi \approx 300$ kHz, meaning that

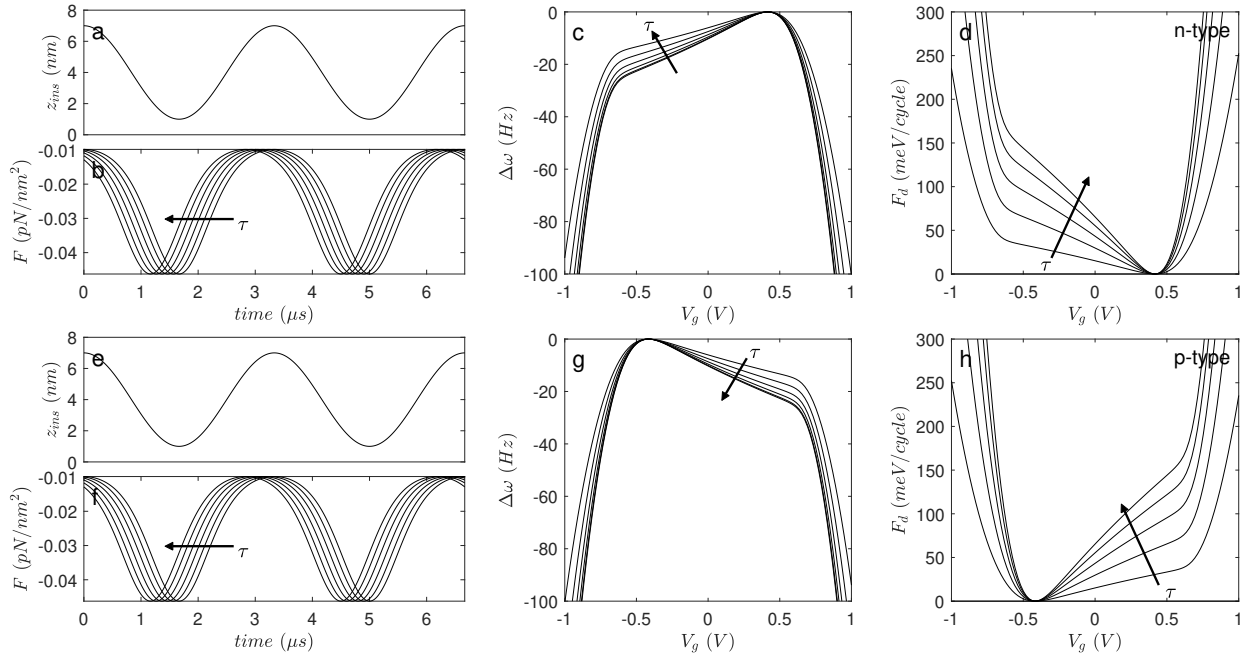


Figure 3.8: The MIS capacitor lag. Time-dependent $z_{ins}(t)$ (a,e) and $F_{el}(t)$ (b,f) at $V_g = 0$. Integrating $F_{el}(t, \tau)$ according to Equation 3.8 for variable V_g gives $\Delta\omega(V_g, \tau)$ and $F_d(V_g, \tau)$. The arrows indicate the direction of increasing τ , from 0 – 500 ns. (a-d) are an n-type MIS capacitor and (e-h) are a p-type MIS capacitor. The capacitor and AFM parameters here correspond to those of Figure 3.7.

they have an oscillation period of approximately $3 \mu s$, so it is necessary that $\tau \ll 3 \mu s$.

The approximate lower bound of measurable τ is mainly determined by the noise characteristics oscillation controller. The minimum measurable phase is theoretically $\delta = 0$ (i.e. $\tau = 0$), but practically, if the damping of the equivalent series resistor is smaller than the intrinsic cantilever damping, variations in τ will be within the measurement noise of the PLL. In the JEOL system, the sampling period is $20 ms$ and the phase noise amplitude $\Delta\theta < 0.2^\circ$. Since the phase noise is normally distributed (the uncertainty goes as $1/\sqrt{N}$), for a $\sim 1 s$ experiment ($N = 50$) and a $300 kHz$ cantilever, at small $|V_g|$, $\tau \approx 1 ns$ is an approximate lower bound on the measurable τ . Once again – these bounds, for now, are somewhat vague and are included in this section as more of an introduction. A more rigorous description of the measurement bounds of τ will be presented in Section 4.2.

3.3.3 Geometrical considerations

The MIS capacitor derivation in Section 2.4 describes a one-dimensional system, in which the surface charge organization occurs uniformly along $\hat{z}$ over the capacitor area a , with no edge effects. This geometry does represent the fm-AFM tip-sample junction to first order, but neglects force contributions from the tip shank and the cantilever, which are known[17] to contribute to the total electrostatic force, F_{el} . A cantilever contribution has been incorporated for this model by calculating the MIS force for a second metallic plate of large area ($a_c \gg a$) and large tip-sample separation ($z_{ins,c} \gg z_{ins}$). The total force, therefore,

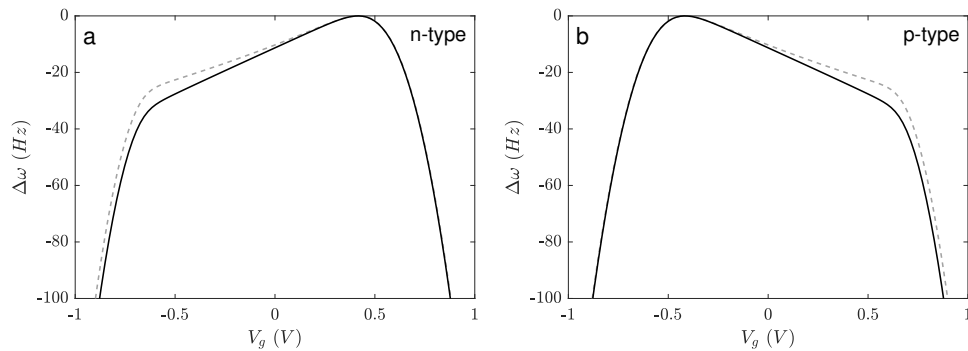


Figure 3.9: **Cantilever contribution to the MIS capacitor.** $\Delta\omega$ for an n-type (a) and p-type (b) MIS capacitor without (grey, dashed) and with (black, solid) a cantilever force contribution F_{cs} , where $z_{ins,c} = 8 \mu m$, $a_c = 3750 \mu m^2$, and $\tau = 0$. The model parameters here correspond to those of Figure 3.7.

is a combination of the tip-sample force and the cantilever-sample force, $F_{el} = F_{ts} + F_{cs}$. The metal-coated tips used for the JEOL system have a tip shank length of approximately $8\ \mu\text{m}$ and an area of approximately $125 \times 30\ \mu\text{m}^2$ [18]. A comparison of F_{el} with and without this F_{cs} contribution is shown in Figure 3.9. The cantilever correction is minor, and so the cantilever geometry ($z_{ins,c}$ and a_c) are not expected to be sensitive model parameters.

References

- [1] Albrecht, T. R., Grütter, P., Horne, D., and Rugar, D. “Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity”. *Journal of Applied Physics* 69.2 (**1991**). DOI: 10.1063/1.347347.
- [2] Giessibl, F. J. “Forces and frequency shifts in atomic-resolution dynamic-force microscopy”. *Physical Review B* 56.24 (**1997**). DOI: 10.1103/physrevb.56.16010.
- [3] Hölscher, H., Schwarz, U., and Wiesendanger, R. “Calculation of the frequency shift in dynamic force microscopy”. *Applied Surface Science* 140.3-4 (**1999**). DOI: 10.1016/S0169-4332(98)00552-2.
- [4] García, R. “Dynamic atomic force microscopy methods”. *Surface Science Reports* 47.6-8 (**2002**). DOI: 10.1016/S0167-5729(02)00077-8.
- [5] Denk, W. and Pohl, D. W. “Local electrical dissipation imaged by scanning force microscopy”. *Applied Physics Letters* 59.17 (**1991**). DOI: 10.1063/1.106088.
- [6] Suzuki, K., Kobayashi, K., Labuda, A., Matsushige, K., and Yamada, H. “Accurate formula for dissipative interaction in frequency modulation atomic force microscopy”. *Applied Physics Letters* 105.23 (**2014**). DOI: 10.1063/1.4903484.
- [7] Arai, T., Kura, D., Inamura, R., and Tomitori, M. “Resistivity change in Joule heat energy dissipation detected by noncontact atomic force microscopy using a silicon tip terminated with/without atomic hydrogen”. *Japanese Journal of Applied Physics* 57.8S1 (**2018**). DOI: 10.7567/jjap.57.08nb04.

- [8] Oyabu, N., Pou, P., Sugimoto, Y., Jelinek, P., Abe, M., Morita, S., Pérez, R., and Custance, Ó. “Single atomic contact adhesion and dissipation in dynamic force microscopy”. *Physical Review Letters* 96.10 (**2006**). DOI: 10.1103/physrevlett.96.106101.
- [9] Pishkenari, H. N. “Atomic interactions between metallic tips and surfaces in NC-AFM”. *Journal of Physics D: Applied Physics* 48.12 (**2015**). DOI: 10.1088/0022-3727/48/12/125301.
- [10] Hölscher, H., Gotsmann, B., Allers, W., Schwarz, U., Fuchs, H., and Wiesendanger, R. “Measurement of conservative and dissipative tip-sample interaction forces with a dynamic force microscope using the frequency modulation technique”. *Physical Review B* 64.7 (**2001**). DOI: 10.1103/physrevb.64.075402.
- [11] Kantorovich, L. N. and Trevethan, T. “General theory of microscopic dynamical response in surface probe microscopy: From imaging to dissipation”. *Physical Review Letters* 93.23 (**2004**). DOI: 10.1103/physrevlett.93.236102.
- [12] Sader, J. E., Uchihashi, T., Higgins, M. J., Farrell, A., Nakayama, Y., and Jarvis, S. P. “Quantitative force measurements using frequency modulation atomic force microscopy – theoretical foundations”. *Nanotechnology* 16.3 (**2005**). DOI: 10.1088/0957-4484/16/3/018.
- [13] Voightlander, B. *Atomic force microscopy*. Springer International Publishing, **2015**. Chap. 16. ISBN: 9783030136536.
- [14] Morita, S., Wiesendanger, R., and Meyer, E., eds. *Noncontact atomic force microscopy*. Springer International Publishing, **2002**. Chap. 20. ISBN: 9783642627729. DOI: 10.1007/978-3-642-56019-4.
- [15] Labuda, A., Miyahara, Y., Cockins, L., and Grütter, P. H. “Decoupling conservative and dissipative forces in frequency modulation atomic force microscopy”. *Physical Review B* 84.12 (**2011**). DOI: 10.1103/physrevb.84.125433.
- [16] Weymouth, A. J. “Non-contact lateral force microscopy”. *Journal of Physics: Condensed Matter* 29.32 (**2017**). DOI: 10.1088/1361-648x/aa7984.

- [17] Sadewasser, S. and Glatzel, T., eds. *Kelvin probe force microscopy: From single charge detection to device characterization*. Springer International Publishing, **2018**. ISBN: 9783319756868.
- [18] *Nanosensors PPP-NCHPt; Electrical, Tapping Mode AFM Probe*. <https://www.nanoandmore.com/AFM-Probe-PPP-NCHPt>. Accessed: 2023-07-07.

Chapter 4

The time-dependent interfacial polarization

A fundamental theme of solid state physics is to explore how materials respond to external electric fields. In general, the response can be understood in terms of the free and bound carriers in the solid. When free charges are subjected to an electric field, they move, leading to current flow through the material. Bound charges, by contrast, are in some way or another spatially restricted, and do not contribute to the current density. This is why wires, whose role it is to carry current between different elements in a circuit, are made out of metals, which have a high free charge density, instead of wood or plastic, which have low free charge densities. Bound charges give rise to a dielectric polarizations, or charge separations of some finite radius within the material. Charges can be bound in a material in various ways, and the resulting polarizations are categorized in terms of the radius of charge separation and the time required to establish the polarization. The total polarization is additively due to each polarization mechanism in the solid. Processes with characteristic frequencies less than the impinging field frequency are unable to establish and so do not contribute to the total polarization. In this way, the polarization response of a material is frequency-dependent[1].

Material polarizations tend to be categorized as electronic (orbital distortions around atomic nuclei, resonating at ultraviolet and optical frequencies), ionic/atomic (non-symmetric

distributions of electrons in the lattice, resonating at optical and infrared frequencies), dipolar (related to structural configurations of the lattice, such as is the case for acoustic or optical phonons, which resonate around GHz and the infrared, respectively), and interfacial polarizations (charge re-organization at interfaces between either like or unlike media, resonating around $kHz - MHz$)[1–5]. Band bending at the semiconductor surface of an MIS capacitor is an example of an interfacial polarization ($\vec{P}_{int}$), since the distribution of charge can be considered as a series of surface dipoles[6, 7].

In the time domain, the frequency dependence of $\vec{P}_{int}$ corresponds to a non-zero surface charge organization timescale τ . In different terminology, this frequency dependence can be understood as a nonzero equivalent series resistance[6], nonzero loss tangent, nonzero dissipation coefficient, dielectric loss, or even as the frequency-dependent interfacial capacitance. The latter comes from describing the MIS system as two capacitors in series:

$$\frac{1}{C_{tot}} = \frac{1}{C_{ins}} + \frac{1}{C_{int}} \quad (4.1)$$

where $C_{ins} = \epsilon_o a / z_{ins}$ (if the insulator is vacuum) and $C_{int} = \epsilon a / z_{int}$, where $z_{int} > w_d$. The interfacial capacitance has different high- and low-frequency limits under inversion[6], due to the finite response times of the carriers which establish the surface potential.

These various ways of representing the dynamic charge organization at a semiconductor surface span several different fields, principally condensed matter physics, materials engineering, and electrical engineering. This chapter aims, in part, to provide a wholistic understanding of the dynamic charge organization that occurs under a time-varying field at the semiconductor surface. Section 4.1 discusses the measured bias and dopant density dependencies of τ at the Si/SiO₂ surface, Section 4.2 outlines the relationship between τ , $\vec{P}_{int}$, and the susceptibility χ_e , and Section 4.3 extends these relationships to dielectric loss, and quantifies dielectric loss at the Si/SiO₂ surface. The MIS capacitor parameters for models of the Si/SiO₂ surface presented throughout this chapter will be given in Chapter 7. An assessment of the model fit is provided in Appendix B.2.

4.1 Measured lag time of Si/SiO₂

In Chapter 3, it was argued that an increase in F_d indicates an increase in the RC time constant τ of the interfacial capacitance. This section shows fm-AFM measurements of an n-type Si(001)/SiO₂ surface which exhibits non-instantaneous interfacial charging and discharging (band bending) as the cantilever oscillates.

4.1.1 Bias dependence

Figure 4.1 shows a bias spectrum measured above the Si/SiO₂ surface. Throughout, the experimental spectrum is in colour, and spectra corresponding to eight different MIS models with variable τ (between 1 – 100 ns) are shown in grey. Null curves, measured at a large tip-sample separation where $F_{ts} \approx 0$ are also shown. The frequency shift ($\Delta f = 2\pi\omega$) bias spectrum (Figure 4.1a) appears parabolic, certainly as compared to the spectra shown in Chapters 2 and 3. This is because – as was mentioned previously – $z_{ins,c}$ is comparatively large ($z_{ins,c} = 12$ nm here, as opposed to $z_{ins,c} = 1$ nm in Chapters 2 and 3). Despite

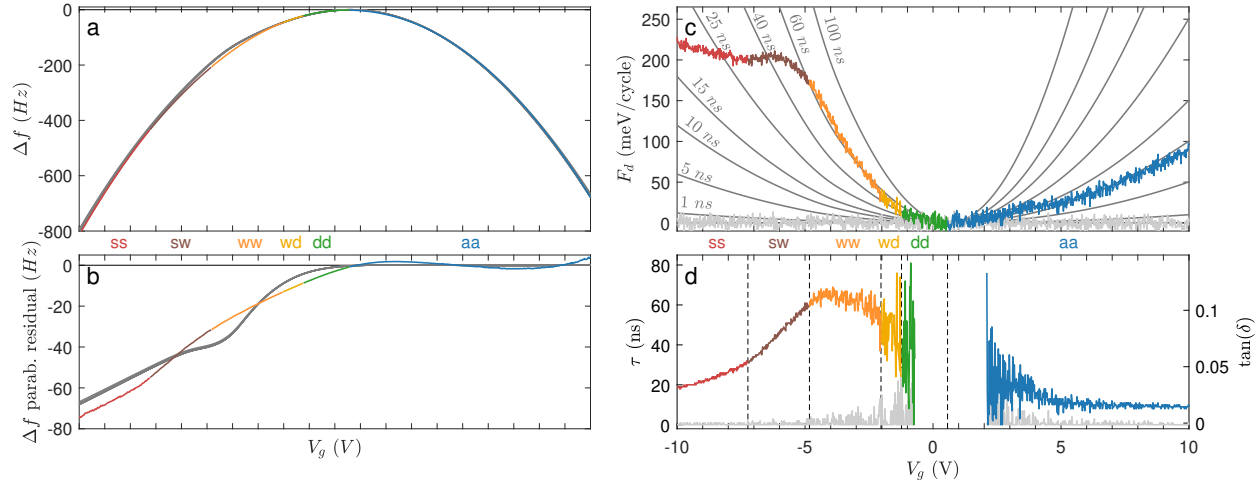


Figure 4.1: Bias dependencies at the Si/SiO₂ surface: Experimental bias spectra above Si/SiO₂ (colour) and null ($z_{ins} \sim 1$ μ m, grey). Eight modelled curves are also shown, corresponding to a MIS capacitor with variable τ , indicated. The colours indicate the six bias regimes, defined in the main text. (a) shows $\Delta f(V_g)$ spectra; (b) shows the residual of $\Delta f(V_g)$ with a parabola fitted to its right-hand side ($V_g > V_{FB}$); (c) $F_d(V_g)$ spectra; (d) best-fit τ results, where biases around V_{FB} (where the uncertainty diverges to infinity) omitted.

its apparent parabolicity, the measured $\Delta f(V_g)$ spectrum is non-parabolic. This is demonstrated in Figure 4.1b, which shows the residual of the experimental $\Delta f(V_g)$ spectrum with a parabola fitted for $V_g > V_{FB}$. The modelled $\Delta f(V_g)$ curves at variable τ , and their residual with their own fitted parabolas, are also shown, demonstrating similar trends. The modelled curves all essentially overlap, which indicates that Δf is very insensitive to τ .

F_d , on the other hand, is sensitive to τ . Figure 4.1c shows that the modelled curves at variable τ do not overlap. The experimental $F_d(V_g)$ spectrum does not follow any single modelled result, which indicates that τ is bias-dependent. Figure 4.1d shows $\tau(V_g)$, which was found by comparing the experimental $F_d(V_g)$ spectrum to MIS models with $\tau = 0 : 100 \text{ ns}$. For every V_g , the model which is closest to the measurement is recorded as the “best-fit” τ . ($\tan(\delta)$ on the right axis of Figure 4.1 will be discussed in Section 4.3.) The uncertainty in τ found with this method is highly bias-dependent, because the modelled F_d curves diverge as V_g increases. At large V_g , the uncertainty is small, which is reflected by the low “noise” in the best-fit τ spectrum. As $V_g \rightarrow V_{FB}$ (i.e. where F_d goes to zero), the uncertainty in τ goes to infinity, since here all of the F_d curves overlap. This region around V_{FB} (which is shown by a vertical line in Figure 4.1b) is therefore omitted. Note, however, that the experimental F_d spectrum does go to 0 when $V_g = V_{FB}$, so even in the omitted region, the model is consistent with experimental results.

The shapes of $\Delta f(V_g)$ and $F_d(V_g)$ are related to the bias-dependent charge organization in the MIS capacitor. As the cantilever oscillates, z_{ins} varies, which can be understood as the system moving along z_{ins} -dependent curves (this was shown in Chapter 3). For example, modelled $F_{ts}(z_{ins})$ curves at variable V_g corresponding to the the model in Figure 4.1 are shown in Figure 4.2. At large bias, $F_{ts}(z_{ins})$ is steeper, which (recall Equation 3.8) corresponds to an increase in Δf and the modelled F_d . Comparing Figures 4.1 and 4.2, however, the experimental τ at $V_g = -9 \text{ V}$ is smaller than τ at $V_g = -5 \text{ V}$. This means that, unlike Δf , τ is not directly related to the steepness of $F_{ts}(z_{ins})$. An intuitive explanation of the origin of the bias dependence of τ follows.

Figure 4.3a shows the z_{ins} dependence of V_S for an MIS capacitor with a closest tip-sample separation ($z_{ins,c}$) of 3 nm . In this example, the tip is oscillating sinusoidally between $z_{ins,c}$ and the farthest tip-sample separation $z_{ins,f} = z_{ins,c} + A$, where $A = 6 \text{ nm}$ is the constant

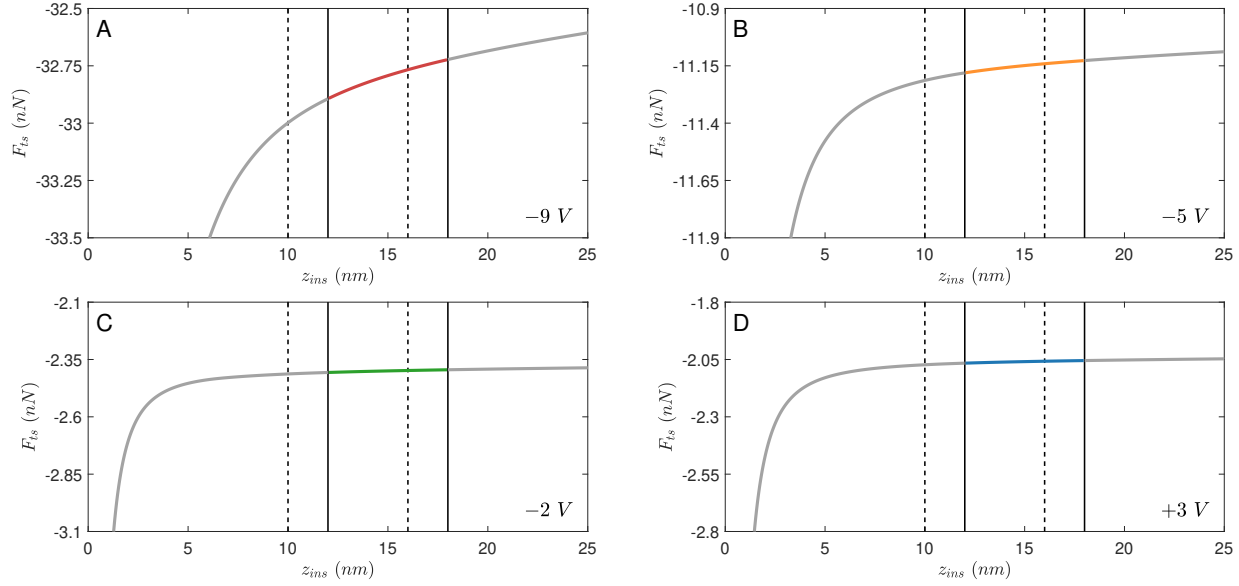


Figure 4.2: **Distance-dependent MIS force.** MIS model corresponding to the models in Figure B.1. $F_{ts}(z_{ins,c})$ for four values of V_g are shown (indicated). The measurement in Figure 4.1 corresponds to $z_{ins,c} = 12 \text{ nm}$ and $A = 6 \text{ nm}$, which is the region between the solid lines. The closest spectrum measured on Si/SiO₂ in this thesis is at $z_{ins,c} = 10 \text{ nm}$, which is the region between the dashed lines.*

fm-AFM oscillation amplitude. V_S is correspondingly time-dependent (Figure 4.3b), as is F_{ts} (Figures 4.3c,d). Figure 4.3 demonstrates that even though V_g is constant, as the cantilever oscillates, the MIS capacitor switches through different bias regimes. At the closest tip-sample separation, the system is under strong inversion (red, Figure 4.3e); at intermediate tip-sample separations, the system is under weak inversion (orange, Figure 4.3f); and at large tip-sample separations, the system is in depletion (green, Figure 4.3g).

In Figure 4.3, $z_{ins,c} = 3 \text{ nm}$, but for the fitted results in Figure 4.1, $z_{ins,c} = 12 \text{ nm}$. The 3 nm example is shown to exaggerate the regime-switching effect for illustrative purposes, but for the experimental Si/SiO₂ measurements shown in this work, the system only switches between two regimes over every oscillation cycle when $A = 6 \text{ nm}$. This is what the colours represent in Figure 4.1: Red indicates that the capacitor is under strong inversion over its entire oscillation cycle (ss), brown is the system switching between strong and weak inversion (sw), orange is under weak inversion over the whole cycle (ww), yellow is switching between

* A central assumption in the fm-AFM derivation shown in Chapter 3 is that the experiment is conducted in a linear force regime. The MIS capacitor model affords an opportunity to assess this approximation. Figure 4.2 shows that overall for the z_{ins} throughout this thesis, the linear approximation is reasonable.

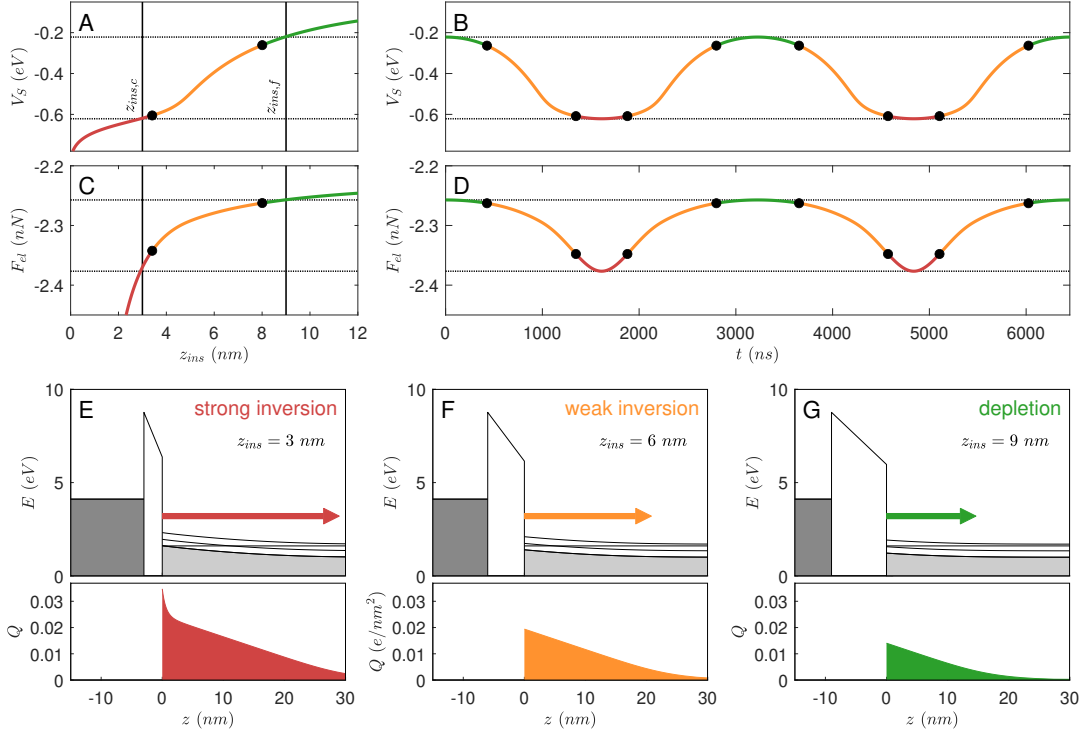


Figure 4.3: **Dynamic charge organization at the semiconductor surface.** (a) z_{ins} dependence of V_S and (b) corresponding time dependence of V_S as the cantilever oscillates between $z_{ins,c} = 3 \text{ nm}$ and $z_{ins,f} = z_{ins,c} + A = 9 \text{ nm}$. (c-d) Corresponding z_{ins} and time dependence of F_{ts} . (e-g) Band diagrams at various points during the oscillation cycle. $V_g = -2 \text{ V}$ for all plots. The arrows represent $\vec{P}_{int}$, to be discussed in Section 4.2.

weak inversion and depletion (wd), green is under depletion for the whole cycle (dd) and blue is under accumulation for the whole cycle (aa).

It is in this context that the bias dependence of τ (Figure 4.1b) can be understood: τ is maximized when the charge reorganization over every oscillation cycle is maximized. The charge organization is related to the separation of the charge from the interface, loosely w_d . In the aa and ss bias regimes, w_d is approximately constant and there is only a change in charge density very near to the surface as the cantilever oscillates; therefore, τ is small. In the dd, wd, and ww regimes, w_d varies significantly over every oscillation cycle; so τ is large.

4.1.2 Dopant density dependence

The Si/SiO₂ sample measured here contains squares patterned by hydrogen resist lithography. The fabrication methodology is described in detail in [8], but essentially, first hydrogen

is selectively removed from a passivated Si(001) surface by scanning tunneling microscopy (STM), and then the surface is exposed to arsenic, which favourably binds to de-passivated sites. 3 nm of Si are then epitaxially grown over the whole surface, which is subsequently terminated by 1 nm of native SiO₂. The patterned squares therefore have a higher dopant density than the background, where the dopant density is related to the number of hydrogen atoms that were de-passivated during the STM hydrogen removal step. Measurements of a similar sample using time of flight secondary ion mass spectrometry show that the dopants are confined to a layer ~ 2 nm thick beneath the epitaxially grown Si layer[8]. The Si wafer is n-doped with phosphorous, so the background (where the bias spectrum in Figure 4.1 was

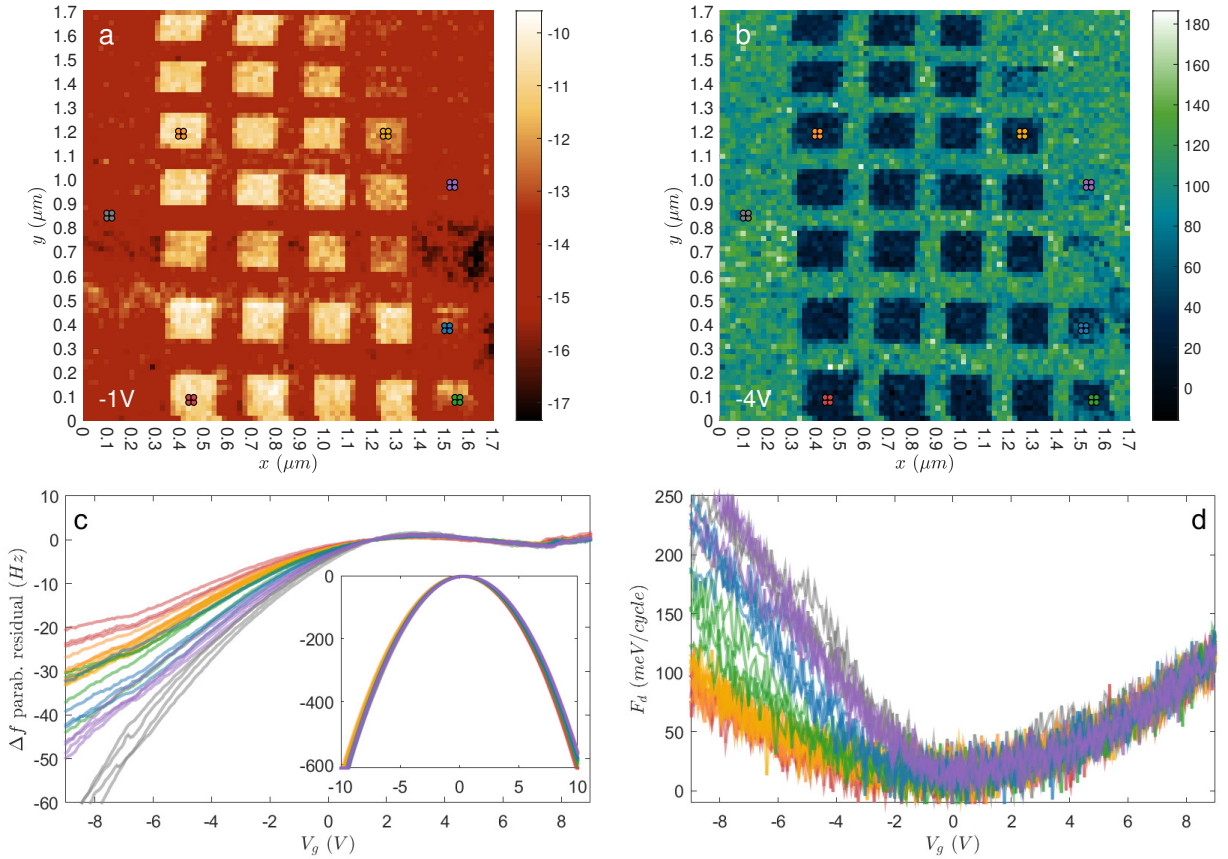


Figure 4.4: **Dopant density-dependent Δf and F_d .** Grid spectroscopy of variably-doped patterned squares in a Si/SiO₂ sample. At each pixel a full spectrum was acquired; the images shown here are “slices” at the V_g indicated. The dots correspond to the spectra shown in (b,c). The dopant density is highest at the bottom left and lowest at the top right, and the background has the lowest dopant density. (a) shows Δf at $V_g = -1$ V and (b) shows F_d at $V_g = -6$ V. (c) shows the residual of the $\Delta f(V_g)$ curves with parabolas fitted for $V_g > V_{FB}$ (the inset shows $\Delta f(V_g)$), and (d) shows $F_d(V_g)$.

measured) has the lowest dopant density.

Figure 4.4 a,b shows a grid spectroscopy image of these patterned squares. The dopant density is highest in the bottom left corner, and decreases first moving upward and then column-by-column moving right. The high dopant density squares can be easily distinguished from the substrate in both the Δf and F_d channels at negative bias, whereas the low dopant density squares are progressively indistinguishable from the background.

Figure 4.4c shows the residual of the $\Delta f(V_g)$ curve with a parabola fitted for $V_g > V_{FB}$ at various positions on the sample. Here, red corresponds to the highest dopant density, purple to the lowest patterned dopant density, and grey to the background (that is, the un-patterned region, which has the lowest dopant density). The non-parabolicity of the measured spectra gradually increases as dopant density decreases. This is because at larger dopant densities, $V_S(V_g)$ becomes increasingly linear over the bias range measured here. This is demonstrated in Figure 4.5a,b. Specifically, the increase in linearity of $V_S(V_g)$ for increasing dopant density is due to an increase in the screening capability of the Si surface (recall Appendix C.6). Practically, this means that high dopant density Si is more “metallic” than low dopant density Si. The increased linearity of $V_S(V_g)$ manifests as an increased constancy of the Δf

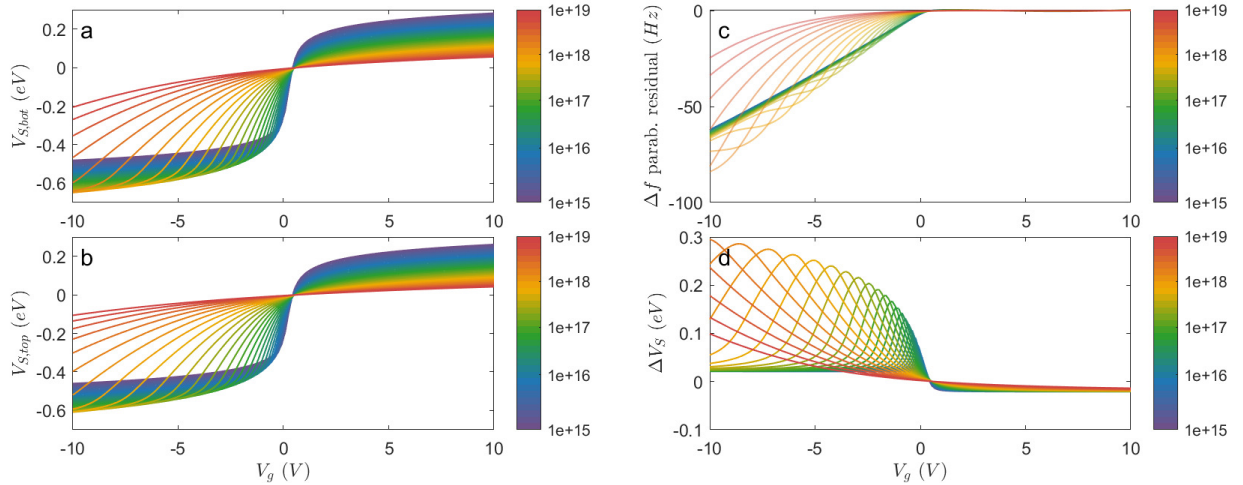


Figure 4.5: **Modelled dopant density-dependence.** (a-b) $V_S(V_g)$ curves at varying bias for the Si MIS capacitor with $z_{ins} = 12 \text{ nm}$ and $A = 6 \text{ nm}$ at the bottom (a) and top (b) of the cantilever oscillation. Purple corresponds to the lowest dopant density ($1e15/cm^3$) and red to the highest dopant density ($1e19/cm^3$). (c) Residual of each corresponding $\Delta f(V_g)$ spectrum with its parabolic fit for $V_g > V_{FB}$. (d) The change in V_S over the cantilever oscillation (ΔV_S) corresponding to the difference of (a) and (b).

parabola fit ($V_g < V_{FB}$) residual. This is shown in Figure 4.5c. By comparing Figures 4.4c and 4.5c, the effective dopant densities of the Si sample squares measured here are between $N_d = 1e17 - 1e19/cm^3$.

Figure 4.4d shows that as dopant density decreases, F_d at negative bias increases. This means that τ increases with decreasing dopant density at negative bias. This can be understood in terms of the discussion of the previous section: Since the change in V_S over every oscillation cycle (ΔV_S) is dopant density-dependent, so is τ . Figure 4.5d shows the modelled change in V_S over every oscillation cycle (ΔV_S). At high dopant densities at negative bias, ΔV_S is small in the experimental bias range, so τ is expected to be small. For the minimum N_d expected in this sample based on the discussion above, ΔV_S is large, so τ is expected to be large. At positive bias, ΔV_S is small for all dopant densities, so τ is expected to be small. These trends agree with the experimental results in Figure 4.4.

4.2 fm-AFM measurement of the interfacial polarization

The method of finding the “best-fit” τ given F_d , described in the previous section, implicitly contains several assumptions regarding the spatial and temporal characteristics of the surface charge density. A more thorough evaluation of this relationship follows in this section. The basis of this discussion is that the charge organization at the semiconductor surface can be described as $\vec{P}_{int}$, which accounts for both the magnitude of the charge and its distance from the interface. Specifically, if the semiconductor surface is located at $z = 0$, the spatial distribution of charge can be represented by writing the interfacial polarization at any time t as a sum of i dipole moments $q_i z_i$:

$$\vec{P}_{int} = \frac{\sum_i q_i z_i}{v} \quad (4.2)$$

where q_i are the charges contributing to band bending at the surface, z_i is their position, and v is the effective probing volume. In correspondence with the previous section, a large change in $\vec{P}_{int}$ over every oscillation cycle is expected to correspond to a large τ .

The temporally- and spatially-dependent polarization $\vec{P}(\vec{r}, t)$ of a dielectric material is not necessarily linearly related to the impinging electric field $\vec{E}(\vec{r}, t)$, and therefore is more generically described using a Taylor series expansion of $\vec{E}(\vec{r}, t)$ with proportionality constant χ_e , where χ_e is the spatially and temporally dependent susceptibility tensor[9]:

$$\begin{aligned} \vec{P}(\vec{r}, t) = & \epsilon_o \int \partial^3 \vec{r}' \int_{-\infty}^t \partial t' \chi_e^{(1)}(\vec{r}, \vec{r}', t - t') \vec{E}(\vec{r}', t') \dots \\ & + \epsilon_o \int \partial^3 \vec{r}_1' \int \partial^3 \vec{r}_2' \int_{-\infty}^t \partial t_1' \int_{-\infty}^t \partial t_2' \chi_e^{(2)}(\vec{r}, \vec{r}_1', \vec{r}_2', t - t_1', t - t_2') \vec{E}(\vec{r}_1', t_1') \vec{E}(\vec{r}_2', t_2') \dots \\ & + \epsilon_o \int \partial^3 \vec{r}_1' \dots \int \partial^3 \vec{r}_n' \int_{-\infty}^t \partial t_1' \dots \int_{-\infty}^t \partial t_n' \chi_e^{(n)}(\vec{r}, \vec{r}_1', \dots, \vec{r}_n', t - t_1', \dots, t - t_n') \vec{E}(\vec{r}_1', t_1') \dots \vec{E}(\vec{r}_n', t_n') \dots \end{aligned} \quad (4.3)$$

(and $\chi_e^{(n)}$ has units of dimensionality $1/\vec{E}^{(n-1)}$). The polarization response is not necessarily temporally instantaneous or spatially isotropic: The integrals* over t describe the non-instantaneity of the response, and the integrals over $\vec{r}$ represent the spatial anisotropy or non-locality of the response.

The nonlinear dependence of $\vec{P}$ on $\vec{E}$ is a central concept in the field of nonlinear optics, which describes electronic polarizations resulting from large-amplitude excitations at visible/ultraviolet frequencies. However, despite finding this description most commonly in optical textbooks, Equation 4.3 is simply the most generic description for dielectric polarization, and is not specific in and of itself to optical frequencies. In Sections 4.2.1 and 4.2.2, the significance and applicability of the temporal and spatial considerations of Equation 4.3 will be discussed specifically in the case of a fm-AFM MIS capacitor. Throughout these sections, a linear response is assumed; the non-linear bias-dependence is accounted for as a bias-dependent χ_e .

* For terms beyond the linear term, the electric field product is not only between electric field terms at the same position at the same time – the electric field product is taken over all positions and all times. This is why there is a distinction between integration constants $\vec{r}_1' \dots \vec{r}_n'$ and $t_1' \dots t_n'$, rather than simply integrating over space and time as $\vec{r}'$ and t' .

4.2.1 Temporal dispersion

In the fm-AFM MIS capacitor, $\vec{E}$ varies temporally due to the oscillation of the tip above the sample surface. (This was discussed in Chapter 3.) $\vec{E}$ is maximized when z_{ins} is minimized, and there is a lag due to the charging time of the MIS capacitor over every oscillation cycle. Consequently, F_{ts} has components which are out-of-phase with z_{ins} , manifesting as an increase in F_d . Experimentally, F_d is non-zero (as shown in Figure 4.1a), which means that the interfacial polarization is non-instantaneous. This section aims to quantify the nature of this non-instantaneity.

Assuming (for the moment) spatial isotropicity and homogeneity, $\vec{P}(t)$ for a linear material is:

$$\vec{P}(t) = \epsilon_o \int_{-\infty}^t \partial t' \chi_e^{(1)}(t - t') \vec{E}(t') \quad (4.4)$$

where the bounds $(-\infty, t)$ are an explicit causal restriction, since electric fields from the future (that is, $t' > t$) cannot influence $\vec{P}(t)$. If under some condition “ $\mathcal{C}$ ” there is a single timescale (“lag time”) $t - t' = \tau_c$ corresponding to a phase offset ϕ_c between $E(t)$ and $P(t)$, then the linear susceptibility is:

$$\chi_e^{(1)}(t - t') = \begin{cases} 0 & , t - t' \neq \tau_c \\ \chi_e^{(1)}|_{\mathcal{C}} & , t - t' = \tau_c \end{cases} \quad (4.5)$$

where $\chi_e^{(1)}|_{\mathcal{C}}$ is a scalar under the condition $\mathcal{C}$ and (by causal restraints) $\tau_c > 0$. Then, Equation 4.4 becomes:

$$\begin{aligned} \vec{P}(t)|_{\mathcal{C}} &= \epsilon_o \chi_e^{(1)}|_{\mathcal{C}} \int_{-\infty}^{\infty} \partial t' \delta((t - t') - \tau_c) \vec{E}(t') \\ &= \epsilon_o \chi_e^{(1)}|_{\mathcal{C}} \vec{E}(t - \tau_c) \end{aligned} \quad (4.6)$$

At first glance, Equation 4.6 takes the form of the standard linear, isotropic non-dispersive polarization equation (i.e. $\vec{P} = \epsilon_o \chi_e^{(1)} \vec{E}$). However, the explicitly stated lag τ_c between $\vec{E}$ and $\vec{P}$ is paramount for the interpretation of F_d in a fm-AFM measurement of a MIS capacitor, as was discussed at the end of Chapter 3.

Equation 4.5 is a simplification which must be treated with care. In general, the permittivity of a capacitor is *not* given by a scalar. This is necessarily true, since the time-dependent polarizations of a discharging and charging capacitor (as in Equation 3.14) are:

$$P_{discharging}(t) \approx P_o e^{-t/\tau} \quad (4.7a)$$

$$P_{charging}(t) \approx P_o(1 - e^{-t/\tau}) \quad (4.7b)$$

where P_o is charged system at equilibrium under a static electric field E_o , and where for the discharged system, $P = 0$. $P_{discharging}(t)$ is shown in Figure 4.6a for capacitors with variable τ under the influence of a step-function electric field. The characteristic τ for each curve is shown by a dashed line. This polarization response is clearly *not* generally well-represented by the delta-function response given in Equation 4.5; rather, capacitors have a Debye-like relaxation with timescale τ . However, as τ decreases, the delta function approximation improves. For the fm-AFM MIS capacitor, the τ for which Equation 4.5 is reasonable depend on the frequency of the cantilever. A 300 kHz cantilever has a period of $\sim 3 \mu s$, which corresponds to the x -axis range after $t = 0$ in Figure 4.6a. Highlighted in red is a discharging capacitor where $\tau = 100 ns$, which is here proposed as a reasonable limit where Equation 4.7 starts to approximate a delta function response. Figure 4.6b shows a capacitor with a delta function response at $\tau_c = 100 ns$. For a 300 kHz cantilever, then, Equation 4.5 is a reasonable approximation for $\tau_c < 100 ns$.

The linear frequency dependent permittivity of the delta function response is found by taking the Fourier transform of Equation 4.5[10]:

$$\begin{aligned} \tilde{\chi}_e^{(1)}(\omega) &= \int_{-\infty}^{\infty} \partial t \chi_e^{(1)}(t) e^{i\omega t} \\ &= \chi_e^{(1)}|_c \int_{-\infty}^{\infty} \partial t \delta(t - \tau_c) e^{i\omega t} \\ &= \chi_e^{(1)}|_c e^{i\omega\tau_c} \end{aligned} \quad (4.8)$$

If the material has an instantaneous polarization (i.e. $\tau_c = 0$), $\tilde{\chi}_e^{(1)}(\omega)$ is real. Conversely, if

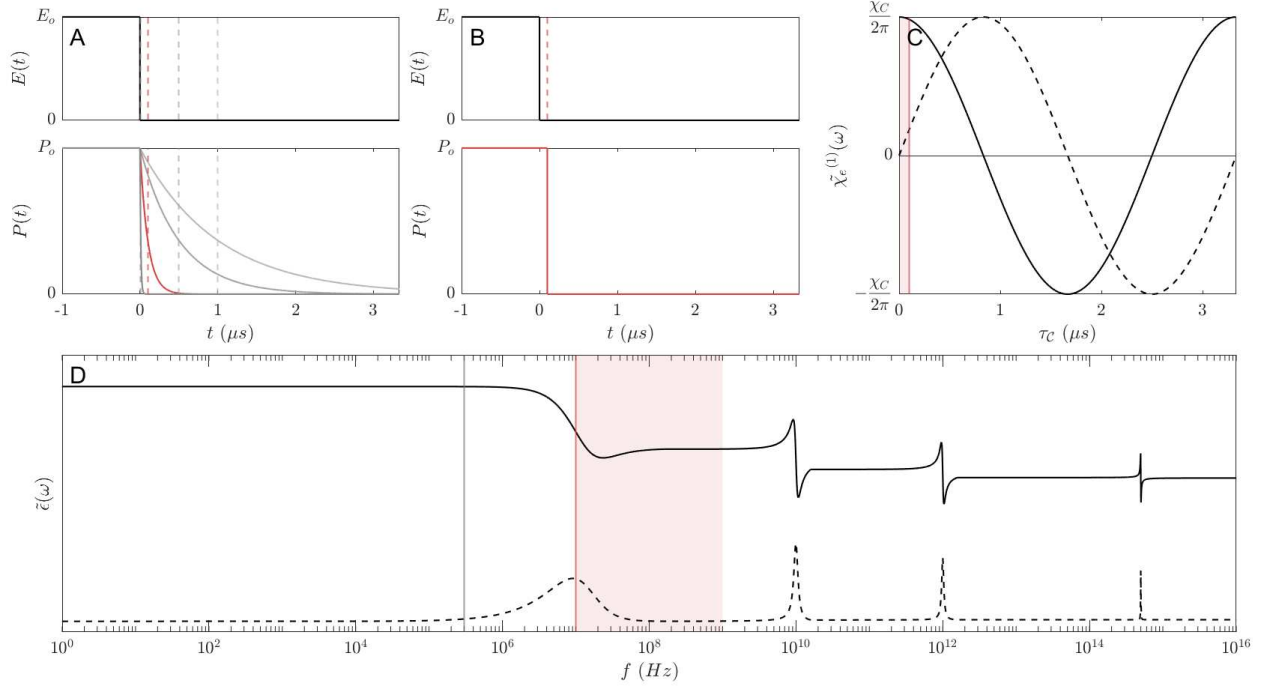


Figure 4.6: **Time-dependence of the polarization response.** (a) Time dependence of a step function electric field (top) and resulting discharging capacitor polarization (bottom) for four capacitors with variable τ (10, 100, 500, and 1000 ns). The $\tau = 100$ ns capacitor is highlighted in red. (b) Step function polarization response where $\tau_c = 100$ ns. (c) τ_c dependence of the real (solid) and imaginary (dashed) components of $\tilde{\chi}_e^{(1)}(\omega)$. $\tau_c \leq 100$ ns is highlighted in red. (d) Illustration of the real (solid) and imaginary (dashed) components of the frequency-dependent permittivity with resonances at 10 MHz, 1 GHz, 1 THz, and 500 THz. The grey line shows 300 kHz, and the red shaded region corresponds to $\tau = 100$ ns - $\tau = 1$ ns, which is the frequency range over which the delta function approximation is valid for a 300 kHz cantilever.

the response is non-instantaneous ($\tau_c > 0$), $\tilde{\chi}_e^{(1)}(\omega)$ is complex. In the quasi-static assumption required for the fm-AFM MIS capacitor model, $\tau_c \ll 2\pi/\omega$, which (since $\epsilon = 1 + \chi_e$) corresponds to the low-frequency limit of the permittivity. In other words, $\tilde{\epsilon}(\omega) \approx \epsilon$, where $\tilde{\epsilon}(\omega)$ is the complex, frequency-dependent permittivity and ϵ is the static permittivity used throughout Chapter 2. The real and imaginary components of $\tilde{\chi}_e^{(1)}(\omega)$ are shown for variable τ_c in Figure 4.6c, where the range $\tau < 100$ ns is highlighted in red. The x -axis range of Figure 4.6c corresponds to one period of a 300 kHz cantilever oscillation, showing that 100 ns is a reasonable upper limit at which the quasi-static approximation applies.

Another requirement for the quasi-static approximation is that the cantilever frequency is far from any natural resonances in the system. Examples of such natural resonances are

illustrated in Figure 4.6d. Dielectric loss (i.e. attenuation of the impinging electric field) is large at the resonance frequency of each polarization mechanism in the material. Therefore, near any resonance frequency, it is no longer true that $\tilde{\epsilon}(\omega) \approx \epsilon$. Figure 4.6d shows an illustration of the real and imaginary components of the frequency-dependent permittivity. The material has several resonances, intended to illustrate an interfacial polarization (10 *MHz*), an acoustic phonon (1 *GHz*), an atomic polarization (1 *THz*/infrared), and an electronic polarization (500 *THz*/optical). The 300 *kHz* cantilever is indicated by a grey line. In this example, the quasi-static approximation is valid, since the cantilever resonance is far from any material resonances, in a regime where $\epsilon(\tilde{\omega}) \approx \epsilon$ and the imaginary component of the permittivity is approximately zero.

4.2.2 Spatial anisotropy and non-homogeneity

The spatial integrals in Equation 4.3 describe any spatial anisotropy that might occur in the polarization response. Such anisotropy could be due either to a spatially non-homogeneous electric field $\vec{E}(\vec{r})$ or a spatially non-homogeneous susceptibility $\chi_e(\vec{r})$. Both of these two forms of spatial anisotropy manifest in an fm-AFM MIS capacitor experiment. First, the electric field at the semiconductor surface arises due to an applied V_g at the tip, and the tip-sample geometry is far from the ideal case of an infinite parallel plate capacitor for which $\vec{E}(\vec{r})$ would be spatially uniform. Such geometrical considerations were briefly addressed in Chapter 3, but ultimately the MIS capacitor model presented in Chapter 2 does not consider any spatial variability of $\vec{E}(\vec{r})$ due to non-planar/edge effects. In general, however, such geometric effects are treated as corrections in electrostatic nc-AFM experiments[11, 12]. The lateral motion of charges corresponding to variable lateral band bending as the cantilever oscillates is expected to contribute to τ [13]. These lateral effects should be explored in future work, in particular near features such as defects or step edges which might increase the lateral resistance, but these effects are also not considered here. Second, a spatial variability of $\chi_e(\vec{r})$ could occur if the semiconductor surface is non-homogeneous. If there is more than one type of material in the fm-AFM scan range, for example, of course χ_e depends on which material is immediately under the tip and contributing to the tip-sample force. Additionally, χ_e depends on imperfections or defects in the semiconductor, such as

adsorbates, vacancies, step edges, trap states, etc.

The small plate area a of the fm-AFM tip affords a unique opportunity to explore the spatially variable $\chi_e(x, y)$ at the surface. This, combined with the phase information which allows for a measurement of τ , discussed in the previous section, is the most important benefit of using fm-AFM to study semiconductor surfaces. The polarization $\vec{P}(r)$ of a linear material (neglecting, for the moment, temporal dispersion) is:

$$\vec{P}(r) = \epsilon_o \int \partial^3 r' \chi_e^{(1)}(r, r') \vec{E}(r') \quad (4.9)$$

where the spatial integral has bounds over all space. For the MIS capacitor, the spatial integral is over the capacitor volume $v = az_{int}$, where a is the capacitor plate area, and z_{int} is the spacecharge depth. If (as for the MIS capacitor derivation in Chapter 2) the semiconductor bulk is infinite, and if the semiconductor surface is located at $z = 0$, the bounds over depth are $(0, z_{int})$. The bounds over area are $(0, a)$, assuming that lateral tip-sample forces are zero. Here, perhaps, is a clearer way to explain the main benefit of the fm-AFM MIS capacitor measurement: In fm-AFM, a is small, meaning that there is less spatial averaging and more opportunity to look for spatial heterogeneity in $\chi_e^{(1)}(x, y)$, as long as that spatial heterogeneity occurs at a scale larger than the tip radius r . In this work, $r \approx 5 \text{ nm}$, so spatial heterogeneity on that scale can be measured. The spatial heterogeneity of $\chi_e^{(1)}$ for a one-dimensional MIS capacitor is then:

$$\chi_e(1)(x - x', y - y') = \begin{cases} 0 & , x \neq x', y \neq y' \\ \chi_e^{(1)}(x, y) & , x = x', y = y' \end{cases} \quad (4.10)$$

where $\chi_e^{(1)}(x, y)$ is a scalar which can vary depending on the tip position (x, y) . If (as discussed above) the electric field is assumed to be spatially uniform, $E(x, y) = E$ is constant over $a = x_a y_a$. The surface polarization can be defined to occur at $z = 0$, and since

$E(z = 0) = E_S$, Equation 4.9 becomes:

$$\begin{aligned}
 \vec{P}(x, y) &= \epsilon_o \int_0^{z_{int}} \partial z' \int_0^{x_a} \partial x' \int_0^{y_a} \partial y' \chi_e^{(1)}(x, y) \delta(z', x - x', y - y') \vec{E}(a') \\
 &= \epsilon_o \chi_e^{(1)}(x, y) \vec{E}_S \int_0^{x_a} \partial x' \int_0^{y_a} \partial y' \delta(x - x', y - y') \\
 &= \epsilon_o \chi_e^{(1)}(x, y) \vec{E}_S
 \end{aligned} \tag{4.11}$$

Equation 4.11, once again, at first glance takes the standard form $\vec{P} = \epsilon_o \chi_e^{(1)} \vec{E}$, but actually it acknowledges the fact that $\chi_e^{(1)}(x, y)$ might be spatially heterogeneous. The derivation of this section is perhaps roundabout, since the result in Equation 4.11 essentially exactly repeats the assumptions stated above, without providing any independent insights. This derivation is included to complete the generic polarization Equation 4.3, and to provide a framework for future extensions of this work incorporating lateral band bending effects.

4.2.3 fm-AFM MIS expression

Combining Equations 4.6 and 4.11, $\vec{P}_{int}$ at any tip position (x, y) and time t is:

$$\vec{P}_{int}(t, x, y)|_c = \epsilon_o \chi_e^{(1)}|_c(x, y) \vec{E}_S(t - \tau_c) \tag{4.12}$$

It might seem strange to express semiconductor band bending in this way, but quite a few important concepts “pop out” of this polarization representation: First, the response is not necessarily instantaneous; second, the susceptibility is not necessarily spatially homogeneous; and third, perhaps most importantly, there is a spatial distribution of charges near the surface which varies over every oscillation cycle.

The actual lag (it should be emphasized again) is a Debye process, as expressed in Equations 3.14 and 4.7. However, in the context of an fm-AFM experiment, it is much more complicated to model a capacitor which is continually relaxing over the continuous oscillation of the cantilever. The delta function approximation shown here (within the frequency constraints that have been discussed in detail above) allows for a much-simplified analysis methodology.

4.3 Dielectric loss at the Si/SiO₂ interface

“Dielectric loss” refers to electromagnetic energy dissipation in a material. This energy can be dissipated by a variety of different mechanisms, depending on various material parameters (including its natural polarization resonances, discussed previously) as well as the characteristics of the impinging electric field. Dielectric loss can in general also be characterized as energy absorption by the material, as attenuation of the impinging field, or dielectric dispersion.

One form of dielectric loss is related to the frequency-dependent complex susceptibility[1], as demonstrated in Figure 4.6, which describes the various polarization resonances of a material. If the excitation frequency corresponds to a resonance frequency, the amplitude of the resulting polarization oscillation is maximized, meaning that the excitation energy is maximally converted into kinetic energy. As the (material) oscillator relaxes, its kinetic energy is converted into/ “dissipated as” heat, which corresponds to the damping of each oscillator[1]. This is the basis of several spectroscopies, such infrared spectroscopy (where an absorption peak corresponds to an atomic polarization resonance, e.g. stretching of molecular bonds) or sub-band gap nonlinear optical microscopy (where an absorption peak corresponds to an electronic polarization resonance). Excitations which are not typically modelled as single harmonic potentials, such as excitons[14] (which are commonly measured by linear ultraviolet/visible absorption spectroscopy) may also have non-radiative relaxation pathways which lead to energy “loss”[15].

Additionally, there is loss associated with the finite conductivity of the material[1]. The electrical conductivity is principally limited by various scattering mechanisms in the material, such as electron-phonon scattering, electron-electron scattering, defect scattering, and surface roughness scattering[1, 4]. At the macroscopic scale, such conductivity losses are also called “Ohmic loss” or “Joule heating”. Note that these losses are not shown in Figure 4.6, which only describes the system as a series of oscillators; Ohmic loss leads to an increase in the imaginary component of the permittivity at low frequencies in the sub-microwave regime.

Dielectric loss goes hand-in-hand with temporal lag in the material response[10, 16]. For an oscillator, this is because the polarization requires some time to establish. For a

capacitive circuit, losses introduce an equivalent series resistance, which shifts the phase of the capacitive response. In the context of fm-AFM, then, it is intuitive that a $F_{ts}(t)$ (i.e. material response) which is lagged with respect to $z_{ins}(t)$ (i.e. the impinging field) is associated with energy dissipation[17–20]. (The total interaction might have both energy gaining and dispersive contributions[21], but the net effect is dissipative.) The increase in F_d due to this lag (required to maintain a constant oscillation amplitude, shown in Chapter 3) compensates for this energy loss in the material. In other words, if energy is being added to the system in the form of an increased drive amplitude (i.e. there is more energy than should be required to compensate for intrinsic cantilever damping) while the oscillation amplitude is constant, the energy must be going somewhere: It is being dissipated in the sample.

4.3.1 Loss tangent

The loss tangent ($\tan(\delta)$) is a parameter that is used to quantify the total electromagnetic energy loss in a material. $\tan(\delta)$ is usually defined in terms of high-frequency losses ($\tan(\delta_b)$) which are due to atomic and electronic polarizations of the intrinsic lattice, and low-frequency (including zero-frequency) losses ($\tan(\delta_c)$) which are due to Ohmic loss [22, 23], as:

$$\tan(\delta) = \frac{\text{im}[\tilde{\epsilon}]}{\text{re}[\tilde{\epsilon}]} = \tan(\delta_b) + \tan(\delta_c) \quad (4.13)$$

In other words, δ_b is due to bound charge carriers in the material, and δ_c is due to conduction losses according to[22, 23]:

$$\tan(\delta_c) = \frac{\sigma}{\omega \epsilon \epsilon_o} \quad (4.14)$$

where σ is the conductivity and ω is the frequency of the impinging field. $\tan(\delta)$ is frequency-dependent, given that the loss mechanisms are themselves frequency dependent, as discussed above. At low frequencies, $\tan(\delta_c) \gg \tan(\delta_b)$, and at high frequencies, $\tan(\delta_b) \gg \tan(\delta_c)$. At radio and microwave frequencies (which includes the 300 *kHz* fm-AFM cantilever), $\tan(\delta_b) \approx 0$ [23], meaning that the dominant loss mechanism associated with band bending in an fm-AFM experiment are Ohmic[7, 23, 24].

Given the discussion of Section 4.2, where the surface charge organization follows a

delta-function re-organization at lag time τ (Equation 4.12), in fm-AFM $\tan(\delta)$ is simply a measure of the phase difference between F_{ts} and z_{ins} , where the total tip-sample force is $F_{ts} = F_{in} \cos(\omega t) + F_{out} \sin(\omega t)$ as expressed in Equation 3.6. The “charge re-organization time” τ is equivalent to a measurement of this angle. Given the small-angle constraint described in Section 4.2, $\tan(\delta) \approx \delta$ and $\tan(\delta)$ and τ are proportionately related:

$$\tan(\delta) \approx \delta = \omega\tau \quad (4.15)$$

where ω is the cantilever oscillation frequency. The loss tangent is frequency and temperature-dependent, but the $\sim 0.01 - 0.1$ order of magnitude shown in Figure 4.1b is consistent with the generic observation that $\tan(\delta) < 1$ [10], as well as with previously reported values in the kHz to low GHz at room temperature (> 0.2 , measurement[25]; $\sim 0.01 - 0.1$, measurements[22], ~ 0.01 , calculation[26]). At low temperature ($< 15 K$), loss tangents are typically two to three orders of magnitude smaller ($< 1e-5$, measurements in the GHz [27, 28]), which can be attributed to the “freezing out” of free carriers. High-purity silicon, which has fewer scattering centers, also exhibits smaller loss tangents ($< 1e-5$, measurement in the GHz [23]).

4.3.2 Relation to mobility

The shape of the $F_d(V_g)$ curve is also related to the carrier mobility μ . In general, the mobility can be understood in the context of Matthiessen’s rule, where the total mobility μ_{tot} is due to the combination of all mobility-limiting mechanisms in the solid[29]:

$$\frac{1}{\mu_{tot}} = \frac{1}{\mu_{bulk}} + \frac{1}{\mu_{surface}} + \frac{1}{\mu_{defect}} \quad (4.16)$$

where μ_{bulk} represents the lattice phonon scattering-limited mobility, $\mu_{surface}$ is the mobility limited by surface scattering (including surface phonon scattering and surface roughness scattering), and μ_{defect} is the mobility limited by defect scattering. In terms of the average total scattering timescale for electrons and holes $\tau_{tot_{e,h}}$, the total electron and hole mobilities

$\mu_{tot_{e,h}}$ are:

$$\mu_{tot_{e,h}} = \frac{\tau_{tot_{e,h}} e}{m_{e,h}} \quad (4.17)$$

where $m_{e,h}$ are the effective electron and hole masses. (Note that τ_{tot} here corresponds to the scattering timescale, as compared to τ which is the RC time constant of the interfacial capacitance.) For an n-type semiconductor, for $V_g < V_{FB}$ the carrier reorganization at the surface is predominantly constituted of holes, μ_h dominates. For $V_g > V_{FB}$, μ_e dominates. Given Equation 4.17 and acknowledging this sensitivity to carrier type, Equation 4.16 can equivalently be written as:

$$\frac{1}{\tau_{tot}} = \frac{1}{\tau_{bulk}} + \frac{1}{\tau_{surface}} + \frac{1}{\tau_{defect}} \quad (4.18)$$

where τ_{bulk} , $\tau_{surface}$, and τ_{defect} are the scattering times in the bulk, at the surface, and near a defect, respectively.

The conductivity is related to the total mobility and number of carriers N according to:

$$\sigma = \mu_{tot} N e \quad (4.19)$$

and therefore, in comparing to Equation 4.14, an increase in μ_{tot} or N corresponds to an increase in loss. A complete understanding of loss as measured during an fm-AFM experiment accounts for variations in both N and in μ_{tot} . In the low-frequency regime of this work, then, τ is essentially an indirect measurement of N and μ_{tot} .

The measurement shown in Figure 4.1 was measured above an apparently featureless location at the surface, so μ_{defect} is presumed to be zero, and loss is dominated by $\mu_{surface}$. The bias dependence can be understood as being due to a bias-dependent N ; that is, at biases where ΔV_S (or Δw_d or $\Delta \vec{P}_{int}$) are large, N (the number of carriers that are actually participating in the surface charge re-organization) is large. Empirically, it has been observed that the surface mobilities of Si are significantly lower than those measured for the bulk: Bulk μ_e values are on the order of $1500 \text{ cm}^2/Vs$ [22], whereas surface sensitive mobility measurements at both room temperature[30] and cryogenic conditions[31] are about an order of magnitude smaller. This is attributed to significant surface scattering[29].

From Equation 4.18, it follows that an increase in τ_{tot} corresponds to an increase in $\tan(\delta)$. Far from any defects, τ_{tot} is expected to be dominated by τ_{bulk} and $\tau_{surface}$ according to Equation 4.18. This is the result presented in Figure 4.1. Measurements above impurities at the surface are expected to contribute a scattering timescale τ_{defect} , increasing τ_{tot} , which manifests as an increase in F_d . This will be demonstrated in Chapter 5.

References

- [1] J. D. Jackson. *Classical electrodynamics*. 3rd ed. John Wiley & Sons, **1999**. Chap. 7.5. ISBN: 0-471-30932-X.
- [2] Ari Henrik Shivola. *Electromagnetic mixing formulas and applications*. IEE Publishing, **1999**. ISBN: 9781849194006. DOI: 10.1049/PBEW047E.
- [3] Juan Martinez-Vega, ed. *Dielectric materials for electrical engineering*. John Wiley & Sons, **2010**. Chap. 1. ISBN: 978-1-84821-165-0.
- [4] Chambers, R. *Electrons in metals and semiconductors*. Springer Netherlands, **2012**. Chap. 8. ISBN: 9789400904231.
- [5] Sebastian, M. T. “Measurement of microwave dielectric properties and factors affecting them”. *Dielectric Materials for Wireless Communication*. Elsevier, **2008**. DOI: 10.1016/b978-0-08-045330-9.00002-9.
- [6] Sze, S. M. and Ng, K. K. *Physics of semiconductor devices*. 3rd ed. John Wiley & Sons, Inc, **2007**. Chap. 1,4. ISBN: 9780470068328. DOI: 10.1002/0470068329.
- [7] Qin, M., Zhang, L., and Wu, H. “Dielectric loss mechanism in electromagnetic wave absorbing materials”. *Advanced Science* 9.10 (**2022**). DOI: 10.1002/advs.202105553.
- [8] Stock, T. J. Z. et al. “Atomic-scale patterning of arsenic in silicon by scanning tunneling microscopy”. *ACS Nano* 14.3 (**2020**). DOI: 10.1021/acsnano.9b08943.
- [9] Peter E. Powers and Joseph W. Haus. *Fundamentals of nonlinear optics*. 2nd ed. CRC Press, **2017**. Chap. 10. DOI: 10.1201/9781315116433.

- [10] Jonscher, A. K. “Dielectric relaxation in solids”. *Journal of Physics D: Applied Physics* 32.14 (**1999**). DOI: 10.1088/0022-3727/32/14/201.
- [11] Sadewasser, S. and Glatzel, T., eds. *Kelvin probe force microscopy: From single charge detection to device characterization*. Springer International Publishing, **2018**. ISBN: 9783319756868.
- [12] Cockins, L. “Atomic force microscopy studies on the electrostatic environment and energy levels of self-assembled quantum dots”. PhD thesis. McGill University, **2010**.
- [13] Hasan, M. M., Arai, T., and Tomitori, M. “Mechanical energy dissipation of an oscillating cantilever close to a conductive substrate partly covered with thin mica films evaluated by frequency modulation atomic force microscopy”. *Japanese Journal of Applied Physics* 61.6 (**2022**). DOI: 10.35848/1347-4065/ac6b02.
- [14] Sumi, H. and Kayanuma, Y. “Is the Lorentz model applicable to optical absorption by excitons?” *Solid State Communications* 85.1 (**1993**). DOI: 10.1016/0038-1098(93)90906-4.
- [15] Chen, H.-Y., Sangalli, D., and Bernardi, M. “Exciton-phonon interaction and relaxation times from first principles”. *Physical Review Letters* 125.10 (**2020**). DOI: 10.1103/physrevlett.125.107401.
- [16] Jonscher, A. K. “Physical basis of dielectric loss”. *Nature* 253.5494 (**1975**). DOI: 10.1038/253717a0.
- [17] Kantorovich, L. N. and Trevethan, T. “General theory of microscopic dynamical response in surface probe microscopy: From imaging to dissipation”. *Physical Review Letters* 93.23 (**2004**). DOI: 10.1103/physrevlett.93.236102.
- [18] Hölscher, H., Gotsmann, B., Allers, W., Schwarz, U., Fuchs, H., and Wiesendanger, R. “Measurement of conservative and dissipative tip-sample interaction forces with a dynamic force microscope using the frequency modulation technique”. *Physical Review B* 64.7 (**2001**). DOI: 10.1103/physrevb.64.075402.
- [19] Denk, W. and Pohl, D. W. “Local electrical dissipation imaged by scanning force microscopy”. *Applied Physics Letters* 59.17 (**1991**). DOI: 10.1063/1.106088.

- [20] Arai, T., Kura, D., Inamura, R., and Tomitori, M. “Resistivity change in Joule heat energy dissipation detected by noncontact atomic force microscopy using a silicon tip terminated with/without atomic hydrogen”. *Japanese Journal of Applied Physics* 57.8S1 (**2018**). DOI: 10.7567/jjap.57.08nb04.
- [21] Sader, J. E., Uchihashi, T., Higgins, M. J., Farrell, A., Nakayama, Y., and Jarvis, S. P. “Quantitative force measurements using frequency modulation atomic force microscopy – theoretical foundations”. *Nanotechnology* 16.3 (**2005**). DOI: 10.1088/0957-4484/16/3/018.
- [22] Krupka, J., Breeze, J., Centeno, A., Alford, N., Claussen, T., and Jensen, L. “Measurements of permittivity, dielectric loss tangent, and resistivity of float-zone silicon at microwave frequencies”. *IEEE Transactions on Microwave Theory and Techniques* 54.11 (**2006**). DOI: 10.1109/tmtt.2006.883655.
- [23] Krupka, J., Kamiński, P., Kozłowski, R., Surma, B., Dierlamm, A., and Kwestarz, M. “Dielectric properties of semi-insulating silicon at microwave frequencies”. *Applied Physics Letters* 107.8 (**2015**). DOI: 10.1063/1.4929503.
- [24] Costato, M. and Reggiani, L. “Lattice-scattering Ohmic mobility of electrons in silicon”. *physica status solidi (b)* 38.2 (**1970**). DOI: 10.1002/pssb.19700380218.
- [25] Yang, R.-Y., Hung, C.-Y., Su, Y.-K., Weng, M.-H., and Wu, H.-W. “Loss characteristics of silicon substrate with different resistivities”. *Microwave and Optical Technology Letters* 48.9 (**2006**). DOI: 10.1002/mop.21786.
- [26] Mukherjee, S., Song, J., and Vengallatore, S. “Atomistic simulations of material damping in amorphous silicon nanoresonators”. *Modelling and Simulation in Materials Science and Engineering* 24.5 (**2016**). DOI: 10.1088/0965-0393/24/5/055015.
- [27] Checchin, M., Frolov, D., Lunin, A., Grassellino, A., and Romanenko, A. “Measurement of the low-temperature loss tangent of high-resistivity silicon using a high-Q superconducting resonator”. *Physical Review Applied* 18.3 (**2022**). DOI: 10.1103/physrevapplied.18.034013.

- [28] Liu, X., Spiel, C., Merithew, R., Pohl, R., Nelson, B., Wang, Q., and Crandall, R. “Internal friction of amorphous and nanocrystalline silicon at low temperatures”. *Materials Science and Engineering: A* 442.1-2 (**2006**). DOI: 10.1016/j.msea.2006.01.146.
- [29] Sah, C., Ning, T., and Tschopp, L. “The scattering of electrons by surface oxide charges and by lattice vibrations at the silicon-silicon dioxide interface”. *Surface Science* 32.3 (**1972**). DOI: 10.1016/0039-6028(72)90183-5.
- [30] Ando, T., Fowler, A. B., and Stern, F. “Electronic properties of two-dimensional systems”. *Reviews of Modern Physics* 54.2 (**1982**). DOI: 10.1103/revmodphys.54.437.
- [31] Niida, Y., Takashina, K., Ono, Y., Fujiwara, A., and Hirayama, Y. “Electron and hole mobilities at a Si/SiO₂ interface with giant valley splitting”. *Applied Physics Letters* 102.19 (**2013**). DOI: 10.1063/1.4803014.

Chapter 5

Defect scattering at Si/SiO₂ trap states

“Uniformity” and “homogeneity”, like “contact” which was discussed in the introduction to Chapter 3, are macroscopic concepts that begin to break down at the nanoscale. A large sheet of metal, for example, by eye is homogeneous, but when viewed under an optical microscope, some dirt or contamination might become visible at millimeter and micron scales. If the surface could be perfectly cleaned, then imaging it with AFM or STM at the nanoscale would show step edges corresponding to the planes of atoms that make up the material. If the surface could be perfectly polished so that it is atomically flat, then imaging it at the atomic scale (which is sub-Angstrom, i.e. less than one tenth of a nanometer) would show heterogeneity corresponding to the individual atoms that make up the material.

It is the lowest level of inhomogeneity which is at play in modern silicon devices. Silicon nanofabrication capabilities are unparalleled for any other material[1, 2] – that is, they are clean and they are flat – yet some degree of surface heterogeneity at the nanoscale is inherently unavoidable. This is because a bare Si surface is energetically unfavorable (it only exists at high temperatures in UHV), and so for the purpose of devices it necessarily has some kind of capping layer (which is most commonly SiO₂, which naturally grows at the Si surface when it is exposed to air). At the atomic scale, there is a “mismatch” of the two materials

at the interface. This natural “mismatching” leads to defects which modify the surface charge organization at the nanoscale[3]. In MOSFET devices, such defects are thought to be the origin of 1/f noise[4, 5] (to be discussed in Chapter 6), as well as other various instabilities[6]. In silicon-based quantum devices, such defects lead to heterogeneities in the dielectric dispersion, which will complicate the realization of many-qubit quantum sensors and computers[7]. This chapter demonstrates the defect-originated spatial heterogeneity of dielectric loss at the Si/SiO₂ interface.

5.1 Spatial heterogeneity of Si/SiO₂

There are many different types of trap states at the Si/SiO₂ interface which can be categorized in terms of their spatial position with respect to the interface and their energy with respect to the semiconductor band edges. The Si/SiO₂ interface has four main categories of traps: *Class a*, which are close to the Si surface and have energies between E_v and E_c ; *Class b*, which are also close to the Si surface but which have energies $> E_c$ or $< E_v$; *Class c*, which are present within the oxide, removed from the Si surface; and *Class d*, which are mobile within the oxide[8]. These categorizations of traps are relevant for many semiconductor-oxide interfaces, but have been most widely studied for Si/SiO₂. In general, these traps can be in various charge states (this is a central theme of Chapters 5 and 6, to be discussed in detail throughout).

Class a traps have many different names, including “interface traps”, “interface states”, “interface-trapped charge states”, “fast states”, and “surface states”[4, 6, 8, 9]. In general, these are due to the natural bond mismatch between Si and SiO₂ which gives rise to a silicon dangling bond[4, 6, 8, 9]. They are named P_b centers in Si(111) and P_{b0} and P_{b1} centers in Si(100). *Class a* trap states exist within the band gap, and therefore it is energetically “common” for them to interact with the surface charge density. They are either donor-like, meaning that they have an energy level in the lower half of the band gap and carry either neutral or negative charge, or acceptor-like, meaning that they have an energy level in the top half of the band gap and carry either positive or neutral charge[4, 6, 8, 10]. The occupancy of a *Class a* trap depends on the position of the trap with respect to E_f (due to V_S and therefore

V_g). Because they are located at the Si surface, they also participate appreciably in scattering with the surface charge density[8, 11]. Additionally, these interfacial dangling bond states can be either bonded or un-bonded with hydrogen. The of role hydrogen passivation for these defects will be discussed in Chapter 6.

Class b traps are called “fixed interface charges” or “fixed surface states”. Like *Class a* traps, they are located immediately at the Si/SiO₂ interface and are due to a bond mismatch between Si and SiO₂[8]. The only difference between *Class a* and *b* traps is their energies: *Class b* trap energies are far above E_c or below E_v . This means that capture and re-emission from the conduction and valence bands is energetically unfavorable under ‘typical’ experimental conditions, However, because of their spatial proximity to the surface, they are still believed to participate in scattering[8].

Class c traps, also called “border traps”, “near-interfacial oxide traps”, “slow states”, and “fixed oxide states” are due to imperfections in the bulk SiO₂[6, 8]. The main types of *Class c* traps in SiO₂ are oxide vacancies (E’ centers), which are common radiation-induced defects[4, 8], as well as oxide-trapped charge[4]. *Class c* traps are found deep within the oxide (i.e. several atomic distances up to nanometers removed from the Si surface), and so scattering at these states is expected to be minimal[11]. Additionally, the barrier for charge transfer between the surface charge density and *Class c* states is significant, and so occurs mainly by tunneling. This means that timescales related to filling and depleting these states depend on the position of the trap within the oxide (i.e. the barrier width), the trap energy, the Si/SiO₂ barrier height, and the concentration of Si surface carriers[4, 8, 12].

Class d traps, also called “mobile oxide charges”[8], are due to Group I contaminants (principally Na⁺) in the oxide, which have large lateral mobility[6, 8]. These contaminants are largely suppressed with modern Si microfabrication methods[8], and so are not presumed to contribute to any of the phenomena measured in this work. Finally, as for any material, grain boundaries are another type of defect, as are point vacancies and atomic impurities. Even “intentional” atomic dopants can be thought of as defects at the atomic scale, in the sense that their scattering coefficients and charge transfer characteristics will differ from the native material.

5.1.1 Loss peaks at an interfacial trap

Class a traps, henceforth in general referred to as interfacial traps (ITs), are expected to be the most significant origin of dielectric loss at the Si/SiO₂ interface because, for the reasons described above, they participate appreciably in scattering of the surface charge density. Furthermore, because interfacial trap energies are within the band gap, the interaction of the trap with the surface charge density is expected to be bias-dependent. This is because the electric potential energy at the trap location is determined by the potential continuity of the MIS capacitor (Equation 2.21), i.e. band bending.

Figure 5.1 demonstrates the bias-dependent IT energy associated with band bending at the Si/SiO₂ interface in an fm-AFM experiment. There is a continual re-organization of the surface charge density as the cantilever oscillates, as was described in Chapter 4. Figure 5.1 shows the band diagrams at the top (i.e. largest tip-sample separation, dashed lines) and bottom (i.e. smallest tip-sample separation, solid lines) of the cantilever oscillation at variable V_g . The bottom of the oscillation corresponds to $z_{ins,c} = 12 \text{ nm}$, and the top corresponds to $z_{ins,f} = 18 \text{ nm}$ ($A = 6 \text{ nm}$).

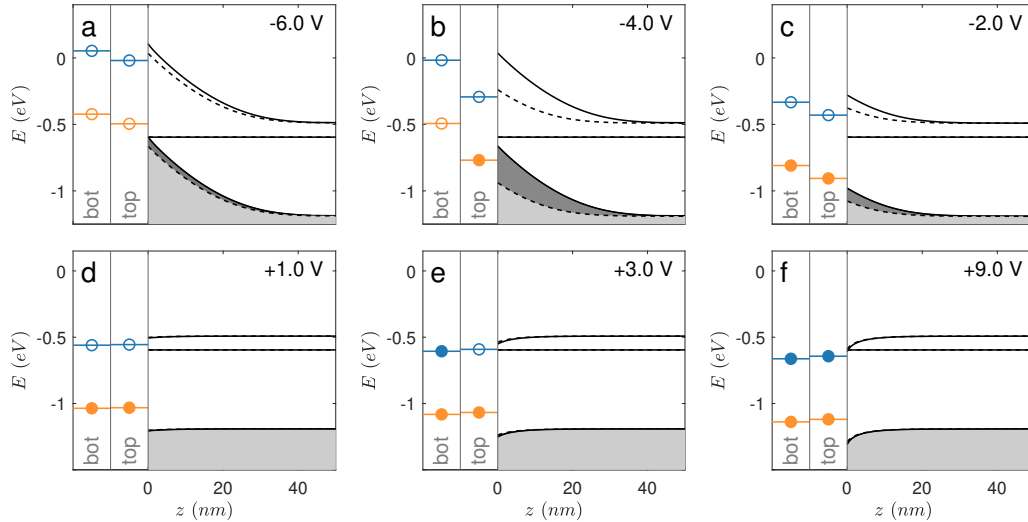


Figure 5.1: **Bias-dependent IT charge state:** Energy of a donor-like (orange) and acceptor-like (blue) IT at variable V_g at the bottom (bot) and top of the $A = 6 \text{ nm}$ cantilever oscillation where $z_{ins,c} = 12 \text{ nm}$, at variable V_g (indicated). Si surface band bending is also shown at the bottom (solid lines) and top (dashed lines). Filled circles correspond to the filled (electron) charge state. [Figure adapted from [13].]

A donor-like trap, which can become positively charged via capture of a hole from the valence band, is shown in orange; an acceptor-like trap, which can become negatively charged via capture of an electron from the conduction band, is shown in blue. Empty circles show empty (electron) states: That is, at large negative bias, the donor-like trap is positive and the acceptor-like trap is neutral; around 0 V both traps are neutral; and at large positive bias, the donor-like trap is neutral and the acceptor-like trap is negative.

Figure 5.2a,b show bias spectra measured at three positions on the Si/SiO₂ surface (on the order of 100 nm apart). In dark grey is the bias spectrum above an apparently “featureless” location (as shown in Chapter 4); the two other locations exhibit peaks in F_d at negative and positive bias. The Δf spectra in Figure 5.2a for each are the same (except for the orange spectrum – this difference can be attributed to ~ 2 nm drift between experiments. The light grey parabola in Figure 5.2a shows a model which is 2 nm closer).

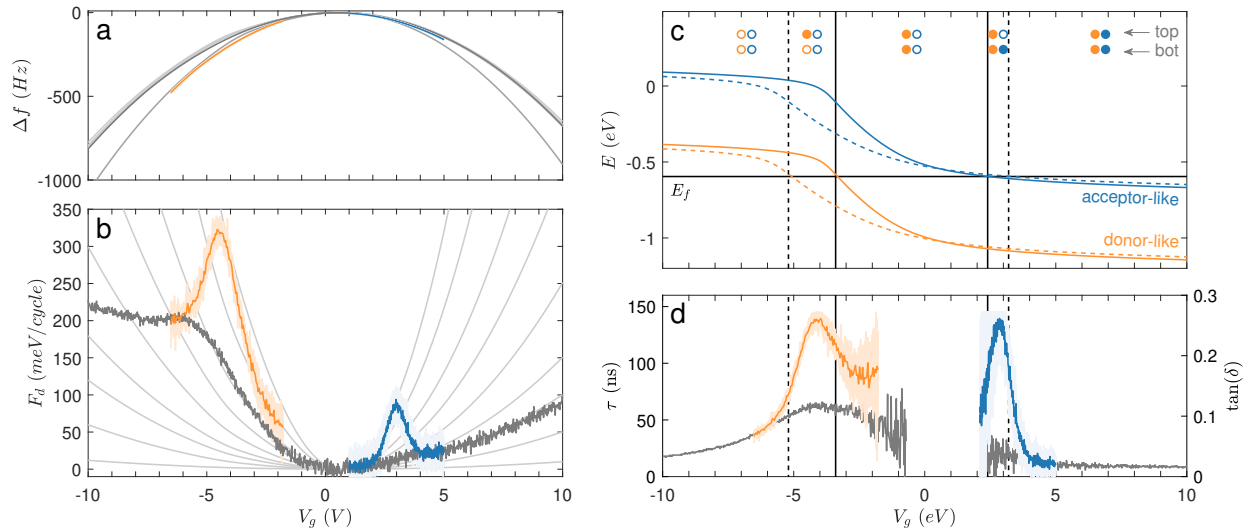


Figure 5.2: Bias-dependent loss. (a,b) Bias spectra measured above three locations at the Si/SiO₂ surface (grey, orange, blue), and models (also grey). (a) The measured and modelled ($z_{ins,c} = 12$ nm) $\Delta f(V_g)$ spectra all overlap, except for the orange curve; This difference can be attributed to 2 nm drift (a model with $z_{ins,c} = 10$ nm is also shown). (b) corresponding $F_d(V_g)$ curves, with models at variable τ (1, 5, 10, 15, 25, 40, 60, 100, and 150 ns). Models corresponding to the $z_{ins,c} = 10$ nm curve are not shown. (c) $E(V_g)$ dependence of the ITs in Figure 5.1 at the bottom (solid) and top (dashed) of the oscillation. Crossing points are shown by vertical lines, and a summary of the charge states in each bias regime is included at the top. (d) $\tau(V_g)$ and $\tan(\delta)$ corresponding to (b) above a donor-like trap (orange), acceptor-like trap (blue), and featureless region (grey). [Figure adapted from [13].]

The full bias dependence of the trap energies (corresponding to Figure 5.1) at the top and bottom of the oscillation are shown in Figure 5.2c. The shape of these curves comes from the shape of the bias-dependent V_S (recall Figure 2.8a). The value of V_g where the trap energy (E_T) equals E_f is called a crossing point. There are therefore four crossing points in total (corresponding to donor-like and acceptor-like traps at the bottom and top of the oscillation), which are indicated by vertical lines. When V_g is greater than (more positive than) the crossing point, the trap has a greater probability of being occupied (by an electron) than it does of being unoccupied, and within the crossing points, the trap has the highest probability of switching its charge state[14].

Figure 5.2d shows that there is a sharp increase in the τ at biases between the crossing points of each trap. The orange curve, therefore, corresponds to increased loss above a donor-like trap, and the blue curve to an acceptor-like trap. The trap energies shown in Figure 5.1 and Figure 5.2c were found by comparing the crossing points to these spectra. They are 0.17 eV above the valence band for the donor-like trap and 0.65 eV above the valence band for the acceptor-like trap. These values are in good agreement with accepted levels for Si/SiO₂ IT energies[9, 15].

A description of the energy loss mechanism for V_g between the donor-like crossing points follows. Figure 5.1b shows a donor-like trap which switches occupancy over every oscillation cycle. Beginning with the cantilever at the bottom of its oscillation, $E_T > E_f$ and the trap is unoccupied. As the cantilever moves upward, the trap energy decreases until $E_T = E_f$. From this point on, the trap energy is less than the average energy of electrons in the semiconductor, meaning that the capture probability is high[16]. When the electron is captured at $E_T > E_f$, it loses energy as it relaxes to its electronic ground state, in a process attributed to phonon scattering[14, 17, 18]. The trap, now occupied, has a low probability of electron emission while $E_T < E_f$, and therefore remains occupied until the cantilever reaches the top of its oscillation. The reverse occurs as the cantilever moves downward again: While $E_T < E_f$, the probability of losing the electron is low, but when $E_T \geq E_f$, the probability of losing the electron is high. Any difference of energy between the trap and the semiconductor ($E_T - E_f$) is dissipated in a phonon “cascade”[8, 14, 17, 18]. Between the acceptor-like crossing points (Figure 5.1e), an analogous loss mechanism occurs.

It should be noted that screening and the additional capacitance of the charged interfacial states[19] are not taken into consideration in the fm-AFM MIS capacitor model presented here. It is well-known that ITs lead to stretching (along the V_g axis) of the $V_S(V_g)$ curve[6]. MIS models which do account for this extra capacitance have been developed elsewhere[6, 8]. There are also fm-AFM models which include this capacitance[20, 21], though in these models the state occupancy is presumed to be static over each oscillation cycle, and there is no discussion of loss (that is, an equivalent series resistance is not included). Here, the in-phase force components (Δf) corresponding to the Figure 5.2h spectra overlap within measurement uncertainty (even at large biases where the traps are certainly charged). This indicates that screening effects (as well as the additional Coulomb force and image charges[22]) are not significant in this measurement, and would lead only to a minor correction of the values of τ and $\tan(\delta)$ in Figure 5.2h.

5.1.2 Dispersion rings

In fm-AFM experiments, spatially localized features which have a peak in a bias spectrum manifest as rings when imaged at constant height[23]. This is due to the spatial localization of the tip (i.e. top gate), which (presuming a spherically symmetric tip) introduces circular equipotential lines in the insulator and sample. For example, when the tip is positioned directly above a donor-like trap at $V_g \sim -4$ V, there is an increase in the measured F_d corresponding to the peak in Figure 5.2h. However, when the tip is moved some distance away in x , y , or z , the potential at the trap site decreases, and V_g must be increased to observe the F_d peak. Similar peaks[24] and rings[25] associated with ITs at semiconductor surfaces have been measured by STM. The contrast mechanism in STM, which measures tunneling between the tip and sample, is different than the contrast mechanism here (STM measures the local density of states at the Fermi energy), though it should be noted that similar dynamics are at play: The measured current in STM depends on the IT occupancy, and so is related to the lifetime of charged IT state; in the STM experiments where peaks and rings were observed, V_g was modulated[24, 25], and so frequency-dependent STM experiments on these systems should exhibit a frequency-dependent tunneling current. This is a possible future avenue to be explored.

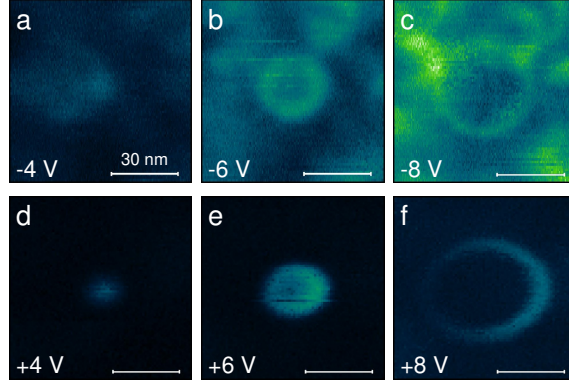


Figure 5.3: **Rings associated with dielectric loss.** Multi-pass images showing F_d at variable V_g above a donor-like IT (top, a-c) and an acceptor-like IT (bottom, d-f). The color scale is the same for all images: $F_d = [27 : 50] \text{ mV}$. [Figure from [13].]

Rings corresponding to a donor-like trap (negative bias) and an acceptor-like trap (positive bias) are shown in Figure 5.3. The ring shape is simply a consequence of raster imaging with a tip as discussed above, so in some sense these rings are nothing more than an imaging artefact (that is, these rings are not present in the sample when the tip is removed). However, measuring the surface in this way is informative on several fronts. First, ring imaging is a straightforward way to demonstrate the stability of ITs. The images shown in Figure 5.3 were measured over the span of hours at large biases (up to $\pm 8 \text{ V}$), which means that the traps are extremely robust as a function of both bias and time. Second, the IT number density at the surface can be estimated by counting the number of rings in an image. Third, ring imaging is an effective way to spatially locate traps. For example, consider a bias spectroscopy experiment for which there was no prior ring imaging. If the spectrum exhibits a peak, it would be challenging to know whether the measurement was done directly above the trap, or whether the peak corresponds to a ring edge at a shifted bias.

5.1.3 Peak shape dependencies

The three main characteristics of the loss peaks in Figure 5.2d are their position (along the V_g axis), width (along the V_g axis), and height (along the F_d , τ , or $\tan(\delta)$ axis). These characteristics depend on the oscillation amplitude A as well as the distance between the

tip and the trap. These dependencies are demonstrated experimentally in Figure 5.4, which shows bias spectra measured at varying A and tip lift (z_{TL}) above a trap.

The peak position is related to the tip-sample distance, as explained previously: At large distances, a larger V_g must be applied to excite the trap which leads to a peak. This is why, in Figure 5.4a-c, as z_{TL} increases, the peak position shifts toward larger negative biases. In Figure 5.4, when $z_{TL} = 6 \text{ nm}$ (i.e. when the distance between the tip and the trap increases by 6 nm), which corresponds to comparing the red and green spectra, the peak position shifts by $\sim 2.5 \text{ V}$ toward negative biases. This agrees with the rings shown in Figure 5.3, where when V_g increases by 2 V (which corresponds to comparing Figures 5.3a and b) the

* Figure 5.4e shows another experimental test of a fundamental fm-AFM assumption. In both of the fm-AFM derivations shown in Section 3.1, $\Delta\omega$ is independent of A (note that in the linear force regime, $F_{ts}(t)$ is directly proportional to A , which cancels with the denominator in Equation 3.8a). Figure 5.4e, which shows all of the $\Delta\omega(V_g)$ spectra corresponding to the $F_d(V_g)$ spectra in Figure 5.4a-c, demonstrates the $\Delta\omega$ is indeed A -independent within experimental uncertainty.

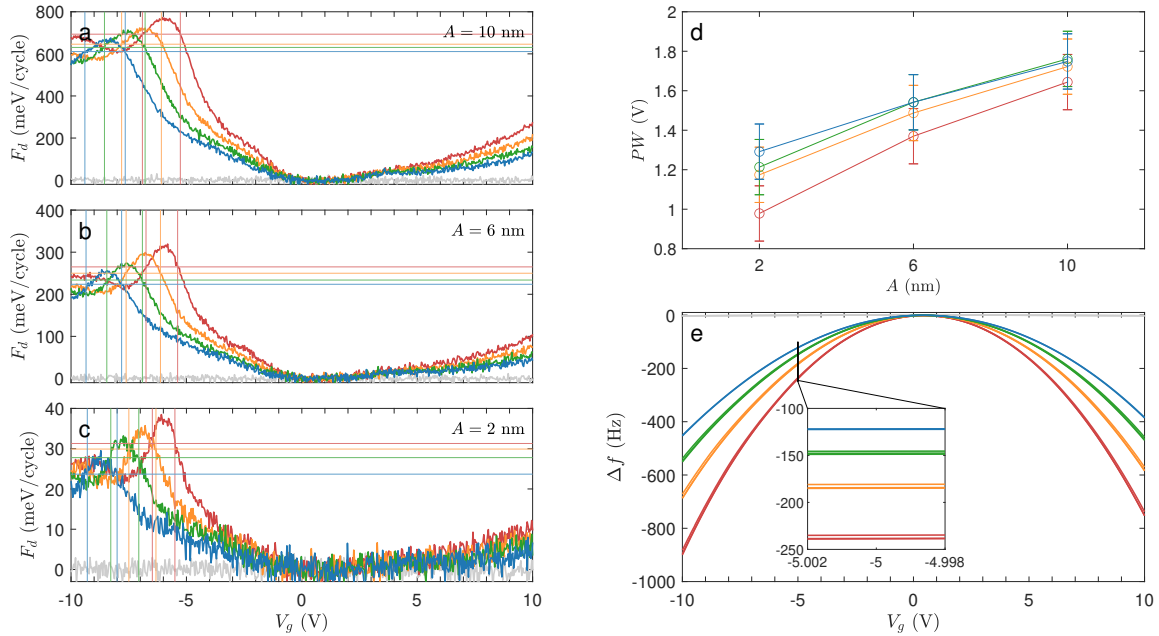


Figure 5.4: **Amplitude and distance dependencies.** (a-c) $F_d(V_g)$ spectra measured at a constant (x, y) at variable z_{TL} (red= 0 nm , yellow= 2 nm , green= 4 nm , and blue= 6 nm) and variable A , indicated. The horizontal lines show the peak half-maximum and the vertical lines correspond to the PW of each spectrum. (d) A -dependent PW . The trend here is not expected to be linear; the lines serve to guide the eye. (e) All 12 $\Delta f(V_g)$ spectra measured simultaneously with the $F_d(V_g)$ spectra in (a-c), overlaid. The inset demonstrates that Δf is amplitude-independent. [Figure from [13]].*

ring radius increases by $\sim 5 \text{ nm}$.

The peak width is related to the separation of the crossing points along the V_g axis: If there is a large separation, there is a larger bias range over which loss is enhanced, and so the peak broadens. In this experiment, the separation between the crossing points increases with A , since at large A there is a significant change in band bending over every oscillation cycle. Figure 5.4d demonstrates this dependence of the peak width (PW) on A . (Note that this relationship is not expected to be linear - the lines simply serve to guide the eye.) Finally, the peak height corresponds to the amount of energy that is dissipated at the defect site. When A is small and when z_{TL} is large, the change in surface potential as the cantilever oscillates is small, and by consequence there is less energy loss.

5.2 Dopant density-dependent defect scattering

All of the data shown in Section 5.1 were measured in the “background” (i.e. unpatterned, lowest dopant density) region of the Si/SiO₂ sample. This section demonstrates the dopant density dependence of IT scattering.

Figure 5.5 shows multipass images measured at negative and positive bias. Three of the squares in Figure 4.4 can be seen: Two at high dopant density (left) and one at intermediate dopant density (right). Figure 5.5a-c shows that there are many donor-like rings in the background region of the surface (the ring density is about 10 rings/100 nm^2) but in the highest dopant density squares only one ring is visible. Acceptor-like rings throughout are rare: Less than 5 are measured in Figure 5.5d-f, all in the background region.

ITs are presumably present at approximately the same density everywhere at the Si/SiO₂ interface, but the bias-dependent interfacial charge organization is highly dopant density-dependent (this was demonstrated in Section 4.1.2). In the high dopant density regions within this bias range, there is less band bending, so less surface charge organization overall, corresponding to less loss – this is why F_d in Figure 5.5a-c is lower overall. The absence of rings in the highly doped regions is also due to the reduced band bending: Due to the increased metallicity of the highly doped regions, larger biases required to shift the trap energy with respect to the band edges – i.e. the crossing points are shifted to more extreme

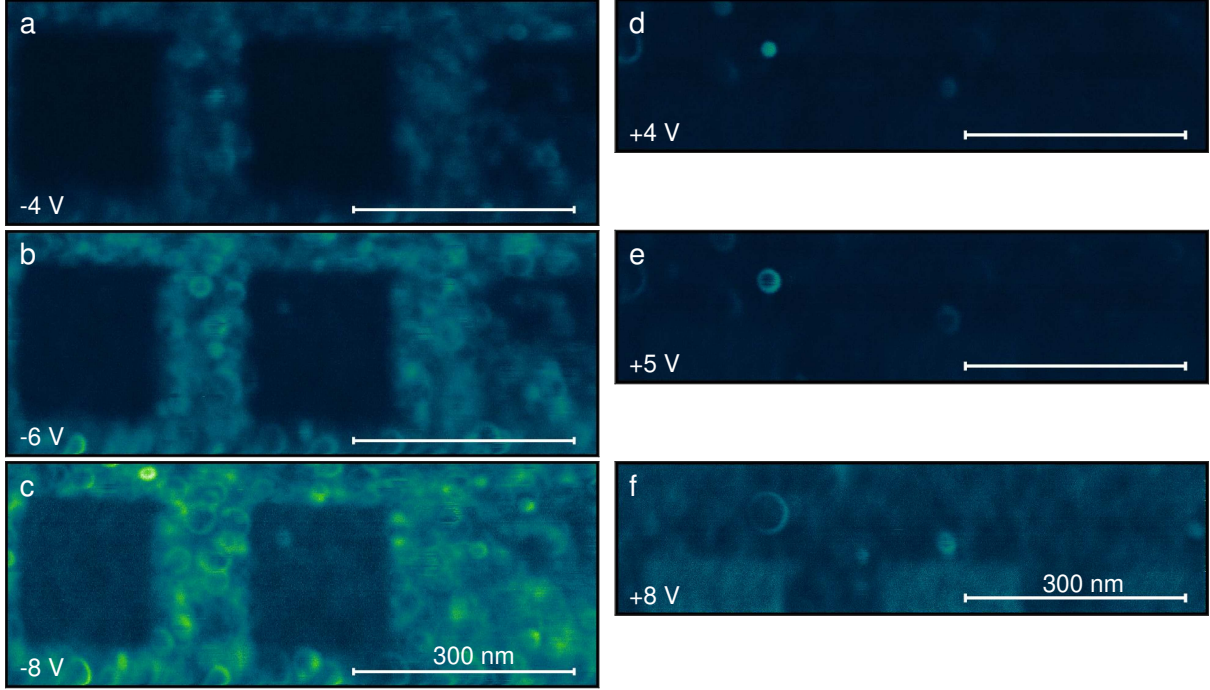


Figure 5.5: **Variable ring densities.** Multi-pass images showing F_d at variable V_g . (a-c) were measured simultaneously (line-by-line), and (d-f) were measured simultaneously. The scale bars are the same for both; but the (a-c) images are 300 nm tall whereas the (d-f) images are 200 nm tall. (Note that the images are vertically offset from one another.) The colour scale bar is the same for all images: $F_d = [25 - 43] \text{ mV}$. [Figure adapted from [13].]

biases.

The general lack of acceptor-like traps can be similarly understood: In the accumulation regime, ΔV_S is small for all dopant densities (see Figure 4.5). Therefore, only traps with energies very close to E_f are expected to participate in scattering. The bright ring in Figures 5.5d-f, therefore, corresponds to an acceptor-like trap with an energy close to E_f . This corresponds with the analysis in Figure 5.2, where the acceptor-like peak actually corresponds to this ring. Given the crossing points analysis discussed, the acceptor-like trap energy is merely 50 meV above E_f .

All of the spots in Figure 5.5a,d correspond to spatially localized decreases in mobility due to scattering by defects (i.e. τ_{defect}). The rings in Figure 5.5b,c,e,f demonstrate the bias-dependent spatial heterogeneity of dielectric dispersion in nanoscale semiconducting devices. For example, if in the development of a Si-based quantum computer two dopant atoms were buried $\sim 2 \text{ nm}$ apart beneath a 2 nm surface, Figure 5.5 shows that at intermediate dopant

densities ($\sim 1\text{e}17/\text{cm}^3$) the electrostatic environment of one is significantly impacted by the other, and additionally by the (unavoidable) Si/SiO₂ ITs.

In the future, measuring loss peaks and rings at variable temperature with this methodology would lead to valuable insights regarding the capture and emission mechanisms of ITs. Capture and emission by ITs are thermally activated processes which involve acoustic phonon interactions with the trap[8, 17, 26, 27]. The ground and excited IT states can be represented as displaced harmonic oscillators where, for an electron to be captured by the trap from the conduction band, it must first gain (vibrational) energy and then relax by phonon scattering or emission.

Specifically, two main mechanisms are proposed to dominate loss during electron capture and emission by ITs. In the first mechanism[17, 27], an IT close to the band edge captures an electron and vibrationally relaxes into its lowest electronic state via a phonon “cascade”, i.e. energy loss due to phonon-phonon scattering between the trap – which has a series of closely-spaced vibrational levels – and phonon modes in the conduction band. In this mechanism, as temperature increases, the probability that a the trap will relax is small because there is a substantial population of phonons in the bulk which can readily excite the electron back into the conduction band. The temperature dependence of this mechanism is T^c , where $1 < c < 4$ [17, 26]. In the second mechanism[28], the ground state IT energy is closer to mid-gap, and there is a significant relaxation after the electron is captured; this energy is lost by multiphonon emission. The temperature dependence of the second mechanism is $\exp(-(E_C - E_T)/(k_B T))$ [26].

References

- [1] Morton, J. J. L., McCamey, D. R., Eriksson, M. A., and Lyon, S. A. “Embracing the quantum limit in silicon computing”. *Nature* 479.7373 (**2011**). DOI: 10.1038/nature10681.
- [2] Gonzalez-Zalba, M. F., Franceschi, S. de, Charbon, E., Meunier, T., Vinet, M., and Dzurak, A. S. “Scaling silicon-based quantum computing using CMOS technology”. *Nature Electronics* 4.12 (**2021**). DOI: 10.1038/s41928-021-00681-y.

- [3] Ambal, K., Rahe, P., Payne, A., Slinkman, J., Williams, C. C., and Boehme, C. “Electrical current through individual pairs of phosphorus donor atoms and silicon dangling bonds”. *Scientific Reports* 6.1 (**2016**). DOI: 10.1038/srep18531.
- [4] Fleetwood, D. M., Pantelides, S. T., and Schrimpf, R. D., eds. *Defects in microelectronic materials and devices*. CRC Press, **2008**. DOI: 10.1201/9781420043778.
- [5] Kirton, M. and Uren, M. “Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise”. *Advances in Physics* 38.4 (**1989**). DOI: 10.1080/00018738900101122.
- [6] Sze, S. M. and Ng, K. K. *Physics of semiconductor devices*. 3rd ed. John Wiley & Sons, Inc, **2007**. Chap. 1,4. ISBN: 9780470068328. DOI: 10.1002/0470068329.
- [7] Wang, Y.-X. and Clerk, A. A. “Intrinsic and induced quantum quenches for enhancing qubit-based quantum noise spectroscopy”. *Nature Communications* 12.1 (**2021**). DOI: 10.1038/s41467-021-26868-7.
- [8] Engström, O. *The MOS system*. Cambridge University Press, **2014**. ISBN: 9780511794490. DOI: 10.1017/CBO9780511794490.
- [9] Poindexter, E. H., Gerardi, G. J., Rueckel, M.-E., Caplan, P. J., Johnson, N. M., and Biegelsen, D. K. “Electronic traps and P_b centers at the Si/SiO₂ interface: Band-gap energy distribution”. *Journal of Applied Physics* 56.10 (**1984**). DOI: 10.1063/1.333819.
- [10] Gray, P. V. and Brown, D. M. “Density of SiO₂ interface states”. *Applied Physics Letters* 8.2 (**1966**). DOI: 10.1063/1.1754468.
- [11] Fleetwood, D. M. “Interface traps, correlated mobility fluctuations, and low-frequency noise in metal-oxide-semiconductor transistors”. *Applied Physics Letters* 122.17 (**2023**). DOI: 10.1063/5.0146549.
- [12] Christensson, S., Lundström, I., and Svensson, C. “Low frequency noise in MOS transistors - I: Theory”. *Solid-State Electronics* 11.9 (**1968**). DOI: 10.1016/0038-1101(68)90100-7.

- [13] Cowie, M., Stock, T. J. Z., Constantinou, P. C., Curson, N., and Grütter, P. *Spatially resolved dielectric loss at the Si/SiO₂ interface*. **2023**. DOI: 10.48550/ARXIV.2306.13648.
- [14] Piscator, J., Raeissi, B., and Engström, O. “Multiparameter admittance spectroscopy for metal-oxide-semiconductor systems”. *Journal of Applied Physics* 106.5 (**2009**). DOI: 10.1063/1.3213384.
- [15] Ragnarsson, L.-Å. and Lundgren, P. “Electrical characterization of Pb centers in (100)Si–SiO₂ structures: The influence of surface potential on passivation during post metallization anneal”. *Journal of Applied Physics* 88.2 (**2000**). DOI: 10.1063/1.373759.
- [16] Nicollian, E. H. and Goetzberger, A. “The Si-SiO₂ interface - Electrical properties as determined by the metal-insulator-silicon conductance technique”. *Bell System Technical Journal* 46.6 (**1967**). DOI: 10.1002/j.1538-7305.1967.tb01727.x.
- [17] Lax, M. “Cascade capture of electrons in solids”. *Physical Review* 119.5 (**1960**). DOI: 10.1103/physrev.119.1502.
- [18] Wang, Y. and Cheung, K. P. “Carrier capture at the SiO₂-Si interface: A physical model”. *Applied Physics Letters* 91.11 (**2007**). DOI: 10.1063/1.2778354.
- [19] Schulz, M. “Coulomb energy of traps in semiconductor space-charge regions”. *Journal of Applied Physics* 74.4 (**1993**). DOI: 10.1063/1.354656.
- [20] Izumi, R., Miyazaki, M., Li, Y. J., and Sugawara, Y. “High–low Kelvin probe force spectroscopy for measuring the interface state density”. *Beilstein Journal of Nanotechnology* 14 (**2023**). DOI: 10.3762/bjnano.14.18.
- [21] Dâna, A. and Yamamoto, Y. “Electrostatic force spectroscopy of near surface localized states”. *Nanotechnology* 16.3 (**2005**). DOI: 10.1088/0957-4484/16/3/023.
- [22] Kantorovich, L. N., Livshits, A. I., and Stoneham, M. “Electrostatic energy calculation for the interpretation of scanning probe microscopy experiments”. *Journal of Physics: Condensed Matter* 12.6 (**2000**). DOI: 10.1088/0953-8984/12/6/304.

- [23] Cockins, L., Miyahara, Y., Bennett, S. D., Clerk, A. A., Studenikin, S., Poole, P., Sachrajda, A., and Grutter, P. “Energy levels of few-electron quantum dots imaged and characterized by atomic force microscopy”. *Proceedings of the National Academy of Sciences* 107.21 (**2010**). DOI: 10.1073/pnas.0912716107.
- [24] Labidi, H., Taucer, M., Rashidi, M., Koleini, M., Livadaru, L., Pitters, J., Cloutier, M., Salomons, M., and Wolkow, R. A. “Scanning tunneling spectroscopy reveals a silicon dangling bond charge state transition”. *New Journal of Physics* 17.7 (**2015**). DOI: 10.1088/1367-2630/17/7/073023.
- [25] Turek, N., Godey, S., Deresmes, D., and Mélin, T. “Ring charging of a single silicon dangling bond imaged by noncontact atomic force microscopy”. *Physical Review B* 102.23 (**2020**). DOI: 10.1103/physrevb.102.235433.
- [26] Raeissi, B. et al. “High-k-oxide/silicon interfaces characterized by capacitance frequency spectroscopy”. *Solid-State Electronics* 52.9 (**2008**). DOI: 10.1016/j.sse.2008.04.005.
- [27] Alekseev, P. S., Kipa, M. S., Perel, V. I., and Yassievich, I. N. “Cascade theory of electron capture in quantum wells”. *Journal of Experimental and Theoretical Physics* 106.4 (**2008**). DOI: 10.1134/s1063776108040213.
- [28] Henry, C. H. and Lang, D. V. “Nonradiative capture and recombination by multi-phonon emission in GaAs and GaP”. *Physical Review B* 15.2 (**1977**). DOI: 10.1103/physrevb.15.989.

Chapter 6

Interfacial noise

In experimental science, “noise” means that measured processes fluctuate in a random way about a value. “Noise” is often contrasted with “signal”, where the signal is the “meaningful” information of the measurement and the noise is some undesired effect. However, in any process, the noise itself has some physical origin, and understanding the noise can lead to powerful insights about the system as a whole.

Noise is commonly characterized by measuring its power as a function of frequency (i.e. its power spectral density, or PSD). White noise has equal power at all frequencies, meaning that its PSD is flat. The name is derived from “white light”, which also has equal power at all visible frequencies. In the time domain, this means that the mean of the noise (i.e. the average of the measurement with the signal subtracted) is zero, and the variance is constant. The noise of a fan, for example – which is due to variations in air pressure as the fan blades, which cover a wide range of tangential speeds, impact air molecules – has a flat PSD. Two common kinds of white noise in electronic signals are Johnson noise, which is due to random thermal motion of electrons in metals, and shot noise, which is due to the quantized nature of electrons which make up electrical current.

Pink noise, in contrast, exhibits higher power at low frequencies. Pink noise is often referred to as “ $1/f$ noise”, since its PSD is proportional to $1/f^\alpha$, where $0 < \alpha < 2$ [1]. The name “pink noise” again comes from an analogy with light: The limit where $\alpha = 2$ is called

“red noise” or “brown noise”, since red is the lowest frequency of the visible spectrum; $\alpha = 0$ is flat (i.e. white); so “pink noise” has values of α somewhere between the two. In processes that give rise to $1/f$ spectra, the mean and variance of the noise vary as a function of time. $1/f$ noise manifests in a huge variety of systems, from traffic flow rates to ocean current velocities, and even loudness fluctuations in radio and music (such as Scott Joplin’s piano rags)[2, 3]. $1/f$ noise also ubiquitously manifests in semiconductor devices, which can reduce the efficiency of CMOS technologies and presents challenges for the continued development of semiconductor qubits.

It is widely accepted that $1/f$ noise in semiconducting devices emerges principally at surfaces[3, 4]. In particular, $1/f$ noise is thought to be due to two-state fluctuations (i.e. random telegraph signals, RTSs) associated with individual defect states. The overall $1/f$ trend in the PSD emerges when there are many traps in the device, each with their own characteristic timescales. Huge research efforts have been devoted to understating the origin of random telegraph noise (RTN) in semiconductors, but pinpointing specific reaction pathways responsible for RTN is challenging. This is in part because RTN is typically measured as drain current or resistance fluctuations in MOSFET devices[3, 5, 6], which are microscopic and so it is hard to extract the behaviour of individual traps.

It is also challenging, in MOSFET measurements, to disentangle whether the measured RTSs are due to fluctuating carrier numbers or fluctuations in the mobility. The spatially localized fm-AFM MIS model presented in Chapter 4 allows for spatially localized measurements of noise, as well as the ability to directly measure variations in mobility associated with RTN at semiconductor surfaces. This chapter demonstrates spatially resolved RTN of individual trap sites at the Si/SiO₂ surface. The noise is shown to have a variety of timescales (Section 6.1) and to manifest as correlated number and mobility fluctuations (Section 6.2). Possible reaction pathways are identified (Section 6.3) and related to the observation that RTN timescales and amplitudes are bias-dependent (Section 6.4).

6.1 Defect-localized noise

Noise at the Si/SiO₂ surface is localized to defect sites. Figure 6.1 shows a multi-pass image where, like in Chapter 5, the F_d channel shows rings with a bias-dependent radius. In this measurement, the rings exhibit noise with a variety of timescales: The T_a ring does not appear to switch at all (this is the kind of ring that was presented in Chapter 5); the T_b ring switches several times as the tip passes over it (the scan speed in these images is 5 nm/s, so this slow switching corresponds to a timescale on the order of Hz); and the T_c ring switches many times as the tip passes over it (which corresponds to fast (ms) timescales on the order of kHz). The data shown in this chapter were measured in the “background” (lowest dopant density) region of the Si/SiO₂ sample. The doped regions, which did not exhibit rings, also did not exhibit noise above the base measurement noise. The analysis to follow was measured near the center of a T_c -type ring.

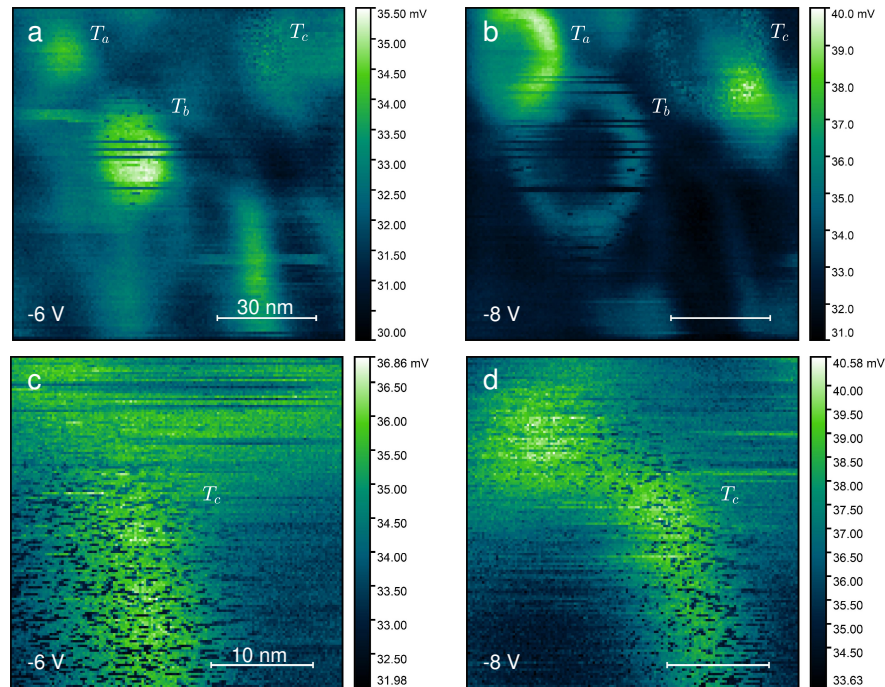


Figure 6.1: **Spatially localized noise.** Multi-pass F_d measurements at the Si/SiO₂ surface at variable bias (indicated). (a-b) and (c-d) were each measured simultaneously. In (a-b), three traps are identified: T_a , which has no apparent switching; T_b , which has slow switching; and T_c (hard to see), which has fast switching. (c-d) are “zoom in” measurements of the T_c trap.

6.2 Quantifying RTSs

Figure 6.2 shows two-state noise in the Δf and F_d channels, measured simultaneously at $V_g = -6.5$ V near the center of a “fast-switching” site. Various quantitative analyses of the time traces are also shown – these analyses will be explained in detail in this section. The two states of this system are labelled “0” and “1”: this will be revisited in Section 6.3. There was some drift (< 1 nm) over time as the noise time traces were measured, so the running mean (where the time average of the mean was much larger than the fluctuation period) was subtracted from the measurements. $\Delta(\Delta f)$ and $\Delta(F_d)$ correspond to the difference in the measured Δf and F_d from this background.

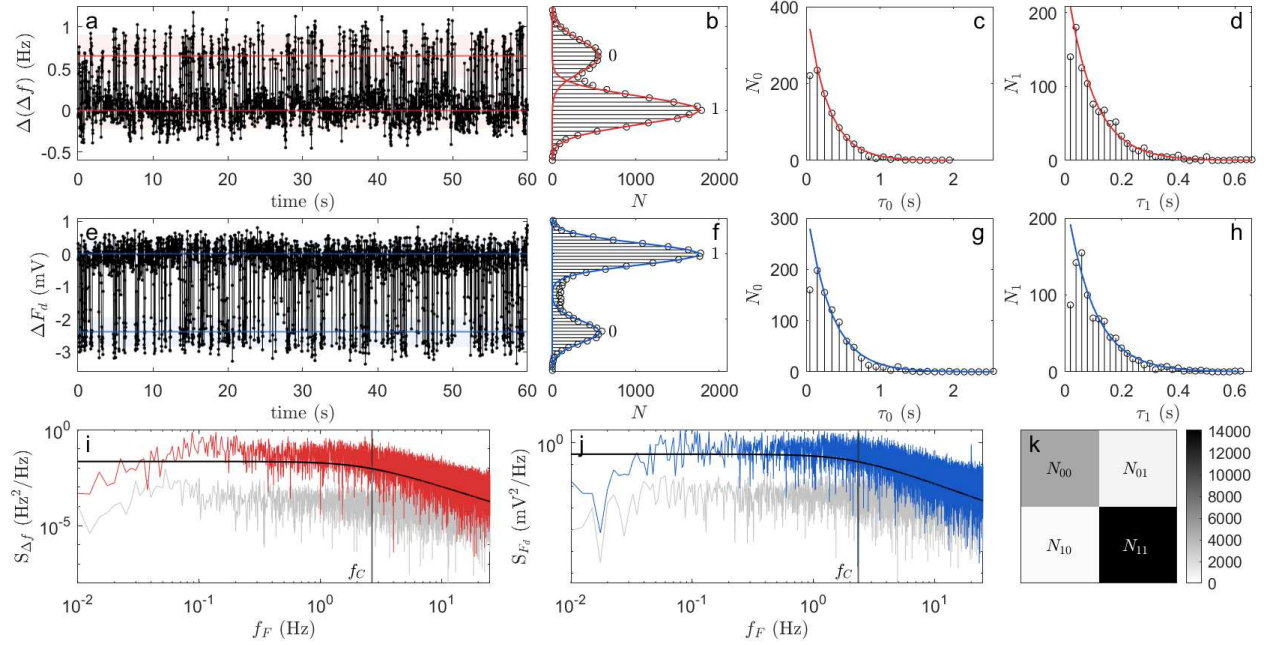


Figure 6.2: **RTS above an IT at $V_g = -6.5$ V.** Simultaneously measured (a-d,i) Δf (red) and (e-h,j) F_d (blue) above an IT. 60 s of the 400 s time trace are shown (a,e). Histograms of the time trace show two clearly distinguishable states (b,f). The mean and FWHM of the Gaussian fits (colour) in (b,f) are shown in (a,e) as coloured lines and shaded regions, respectively. The wait time between $0 \rightarrow 1$ transitions (τ_0 ; c,g) and $1 \rightarrow 0$ transitions (τ_1 ; d,h) are exponentially distributed; exponential fits are shown. (i,j) PSD of the Δf and F_d time traces, respectively (colour) and the base instrument noise (grey). Lorentzian fits are overlaid, and the corner frequency f_C is shown. (k) Correlation matrix of the Δf and F_d signals.

6.2.1 Fluctuation amplitude

The amplitude of the RTS (A_{01}) is measured by fitting Gaussian distribution functions to the time traces (Figure 6.2a-b,e-f) such that:

$$A_{01} = |\bar{0} - \bar{1}| \pm \sqrt{(\delta\bar{0})^2 + (\delta\bar{1})^2} \quad (6.1)$$

where $\bar{0}(\bar{1})$ is the mean value of 0(1) peak and $\delta\bar{0}(\bar{1})$ is the standard deviation of the Gaussian fit. In Figure 6.2a,e, the mean of each Gaussian fit is a solid coloured line, and the standard distributions are shaded coloured regions. Note that in using a Gaussian fit, there is an assumption that the noise in each state is normally distributed. This is a good assumption: The R^2 values for the fits are close to 1 (for the four fits in Figure 6.2b,f, from top to bottom they are 0.9971, 0.9973, 0.9983, and 0.9929). The peak widths are also approximately constant, and correspond to the measurement noise in a non-fluctuating system at this bias.

6.2.2 Exponential wait times

Next, the “wait time”, or time between state transitions, is measured. The random telegraph fluctuations studied here are two-state Poisson point processes (discussed further in Section 6.3), and therefore the time between transitions is exponentially distributed (derivation in Appendix C.11):

$$P_{0 \rightarrow 1} = \frac{1}{\bar{\tau}_0} \exp\left(-\frac{\tau_0}{\bar{\tau}_0}\right) \quad (6.2a)$$

$$P_{1 \rightarrow 0} = \frac{1}{\bar{\tau}_1} \exp\left(-\frac{\tau_1}{\bar{\tau}_1}\right) \quad (6.2b)$$

where $P_{0 \rightarrow 1(1 \rightarrow 0)}$ is the probability of a transition occurring at a given $\tau_{0(1)}$, which is the wait time in the 0(1) state before a $0 \rightarrow 1$ ($1 \rightarrow 0$) transition. $\bar{\tau}_{0(1)}$ is the average wait time in the 0(1) state. The uncertainty of $\bar{\tau}_{0(1)}$ corresponds to the 95% confidence interval of the

exponential fit. The rates of the $0 \rightarrow 1$ and $1 \rightarrow 0$ transitions are correspondingly:

$$k_{01} = \frac{1}{\bar{\tau}_0} \pm \frac{\partial \bar{\tau}_0}{\bar{\tau}_0^2} \quad (6.3a)$$

$$k_{10} = \frac{1}{\bar{\tau}_1} \pm \frac{\partial \bar{\tau}_1}{\bar{\tau}_1^2} \quad (6.3b)$$

Figure 6.2c-d,g-h show the wait time distributions with exponential fits. In each histogram, the first data point is omitted from the fit because it is artificially lowered due to the finite sampling period (20 ms). A description of the analysis to evaluate the timescales in Figure 6.2c-d,g-h is provided in Appendix B.5.

6.2.3 Lorentzian power spectral density

The power spectral density (PSD, Figure 6.2i,j) of a two-state RTS time trace is Lorentzian in the form[7, 8]:

$$S(f_F) = \frac{4A_{01}^2}{(\bar{\tau}_0 + \bar{\tau}_1) \times \left[\left(\frac{1}{\bar{\tau}_0} + \frac{1}{\bar{\tau}_1} \right)^2 + (2\pi f_F)^2 \right]} \quad (6.4)$$

where f_F is the Fourier frequency. The corner frequency f_C is:

$$f_C = \frac{1}{\bar{\tau}_0 + \bar{\tau}_1} \quad (6.5)$$

such that $S(f_F)$ exhibits a f_F^{-2} dependence for $f_F > f_C$. In Figure 6.2i,j, the Lorentzian curves and f_C values shown (Equations 6.4 and 6.5) correspond to the A_{01} , $\bar{\tau}_0$, and $\bar{\tau}_1$ values which are found from the RTS analysis (Equations 6.1 and 6.2), and there are no free fit parameters. (This serves as a way to check that the signal only has one RTS, within noise, and that the RTS analysis is accurate.) The corner frequency is shown by a vertical line.

The $1/f$ trend in the PSD of macroscopic devices is an emergent effect that arises in when there are many different two-state fluctuators with different corner frequencies[9–11]. For example, Figure 6.3 shows power spectra corresponding to two different traps at the Si/SiO₂ surface, along with their Lorentzian fits (Equation 6.4) and corner frequencies (Equation 6.5). Two dashed black lines show $S = f_F^\alpha$ for $\alpha = -1$ and $\alpha = -2$. The slope of PSD of each individual trap (at frequencies higher than its corner frequency, i.e. where the second

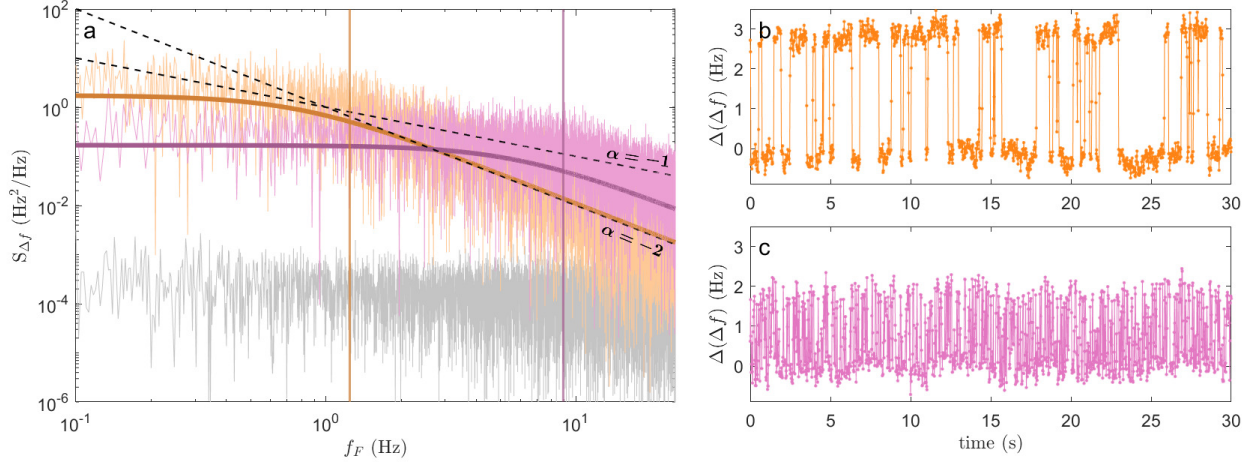


Figure 6.3: **Emergent $1/f$ trend for multiple RTSs.** (a) Power spectra of the Δf channel for two traps (orange and pink) and the base sample PSD (grey). Lorentzian fits and corner frequencies are shown for both traps in their respective colours. The black lines show $S = f_F^\alpha$ for $\alpha = -1$ and $\alpha = -2$ (indicated). (b-c) Time traces of both traps in their respective colours. The total time traces were measured for 500 s; only the first 30 s are shown. Note that the x - and y -axes are the same for both traps to allow for easy comparison between them.

term dominates the denominator in Equation 6.4) is proportional to -2 (i.e. $\alpha = -2$), but together, Figure 6.3 shows that the combined PSD slope (for f_F between the corner frequencies) is proportional to -1 (i.e. $\alpha = -1$).

The two measurements shown in Figure 6.3 were not chosen arbitrarily. The orange measurement has one of the longest sets of up and down timescales measured on the sample, and the pink measurement has one of the shortest sets of timescales. These particular measurements were selected for this figure to maximize the difference in their corner frequencies, so as to better visualize the effect of adding the two power spectra. The fact that these are the smallest and largest corner frequencies observed on this sample is due to experimental constraints: The longer measurement times required to study lower-frequency traps would be prohibitively long because drift artifacts would become appreciable; the measurement of higher-frequency traps is prohibited by the finite sampling period, which was 20 ms ($= 50 \text{ Hz}$). It can be noted, however, that in the absence of these experimental constraints, the frequency range over which a PSD exhibits an $\alpha = -1$ trend is related to the range of characteristic timescales of its individual fluctuators.

6.2.4 Correlated number and mobility fluctuations

Figure 6.2k shows the correlation matrix of the Δf and F_d time traces in Figure 6.2a,e, where N_{11} , N_{10} , N_{01} , and N_{00} are the number of instances (over a time trace) where Δf and F_d , respectively, are in the given state. For example, N_{01} is the number of points where Δf is in state 0 and F_d is in state 1. The off-diagonal elements of the correlation matrix are close to zero, which indicates that the two variables are correlated, and $N_{11} \gg N_{00}$, which means that the lifetime of the 1 state is longer than the lifetime of the 0 state (see Figures 6.2c-d,g-h for comparison).

The correlation between Δf and F_d is quantified as the phi coefficient:

$$\Phi = \frac{N_{11}N_{00} - N_{10}N_{01}}{\sqrt{(N_{11} + N_{10})(N_{10} + N_{00})(N_{00} + N_{01})(N_{01} + N_{11})}} \quad (6.6)$$

where $\Phi = 1$ indicates perfect correlation, $\Phi = -1$ perfect anti-correlation, and $\Phi = 0$ indicates no correlation. For this RTS, $\Phi = 0.87$, so Δf and F_d are strongly correlated. In these measurements, F_d is related to a phase lag, scattering, mobility, and loss (recall Chapter 4), so two-state fluctuations in F_d mean that the system is switching between a state where there is much inelastic scattering and a state where there is little. Δf is a measure of the charge at the semiconductor surface, or V_S (recall Chapter 2), so two-state fluctuations in Δf correspond to the system switching between two different $V_S(z)$ curves as a result of variations effective dopant density, band gap, or screening of the surface charge density by the trap.

The large Φ value found here, therefore, indicates that at a single site at the Si/SiO₂ interface, surface potential fluctuations and mobility fluctuations are correlated. Consider the results in Figure 6.2 as an example: When the system is in state 0, F_d is lower, meaning that timescales associated with scattering in state 0 are very short. In fact, considering the ring fluctuations in Figure 6.1, the “low F_d ” state approximately matches the background, so it is a reasonable approximation that defect scattering is negligible in the 0 state. The 1 state, however, exhibits a higher F_d , meaning that in the 1 state there is more energy loss and non-negligible defect scattering.

RTSs are most commonly measured in MOSFET devices as a fluctuating drain current I_d . In the correlated mobility fluctuation (CMF) model, the change in drain current when a carrier is captured (i.e. the RTS amplitude) is [3, 12–15]:

$$\Delta I_d = -\frac{I_d}{WL} \left(\frac{1}{N} \pm \alpha \mu \right) \quad (6.7)$$

where L and W are the MOSFET channel length and width, N is the channel carrier density per unit area, α is the carrier-defect scattering coefficient, and μ is the mobility. The sign in front of the mobility depends on whether the trap becomes charged (mobility decreases due to increased Coulomb scattering) or neutral (mobility increases) when a carrier is captured. In this simplest model, α is assumed to be constant (that is, independent of the trap charge state) [15]. Because both N and μ manifest in the same signal I_d , it can be difficult to disentangle the two. The measurement in Figure 6.2 is a straightforward demonstration of correlated number and mobility fluctuations at a single trap.

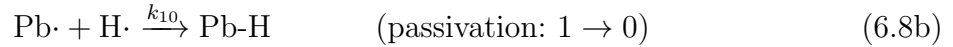
6.3 Multi-state reaction pathways

There has been significant debate regarding whether $1/f$ noise in semiconducting devices is dominated by charge trapping and emission at defect states (i.e. number fluctuations) or variable carrier scattering (i.e. mobility fluctuations) [8]. It is increasingly recognized that number and mobility fluctuations likely both play critical roles in low-frequency noise [13, 14], and there is generally consensus that defects play a role in $1/f$ noise, but the question of what kinds of defects are involved in RTN persists [13]. Oxide traps (OTs), which are removed some distance from the Si/SiO₂ interface, are consistent with tunneling-originated RTSs, where the tunneling rate decays exponentially with the distance of the OT from the interface. However, these traps are too far removed from the interface to participate appreciably in scattering, and therefore cannot explain mobility fluctuations [13]. Furthermore, slow (Hz) $1/f$ noise has been demonstrated in MOSFETS with very thin oxides [16], and even at oxide-free Si surfaces [17], which is inconsistent the fast (MHz) tunneling rates that would be predicted over such small distances [5, 16, 18]. Interface traps (ITs) such as Pb₀ and Pb₁

centres in Si(001), by comparison, are located exactly at the Si/SiO₂ interface, and therefore are expected to participate appreciably in scattering. However, also due to their being located at the interface, tunneling rates are expected to be much faster than the $ms - s$ ($Hz - kHz$) RTS fluctuations which have been empirically observed. Recently[3, 19, 20], it has been proposed that many-state pathways are likely associated with RTN, which further complicates identification of their origin. Two possible mechanisms evaluated below.

6.3.1 Hydrogen passivation / activation of Pb centers

In one proposed mechanism, Pb centers – which, recall Section 5.1, are ITs corresponding to Si dangling bonds (DBs) – are slowly kinetically activated and passivated by hydrogen[13, 19, 21]. In this sample, the Si surface is capped by SiO₂, so the hydrogen participating in these reactions is presumably interstitial. Interstitial hydrogen can exist in molecular or atomic positive, negative, or neutral charge states[21], but in the atomic form reacts with Pb centers according to[21–26]:

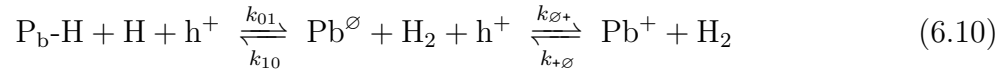


where k_{01} and k_{10} are the activation and passivation rates, respectively. The passivated trap (Pb-H, i.e. $\text{Si}_3 \equiv \text{Si-H}$) is here defined as the “0” state, and the activated trap (Pb·, i.e. $\text{Si}_3 \equiv \text{Si}\cdot$) is the “1” state. The DB state in Equation 6.8 is neutral (i.e. in the notation that will be used here, Pb· is equivalent to $\text{Pb}^\varnothing$), but donor-like DBs also exist in the positive charge state:



where $k_{\varnothing+}$ and $k_{+\varnothing}$ are the charging and discharging rates, respectively. Equation 6.9 was the reaction identified in Chapter 5. Combining Equations 6.8 and 6.9, there is a four-fold competition of rates. The simplified balanced reaction of the Pb center with atomic

hydrogen, where the dissociation of H_2 is not written explicitly but rather is included in the rate constant k_{10} , is:



In this description, the number of DBs in the device fluctuates. Additionally, as the DB is randomly activated and passivated by hydrogen, the mobility also fluctuates. In the activated (1) state, the trap is able to be charged and discharged, and so has associated loss according to the description in Chapter 5 (that is, loss is maximized at the crossing points V_C). The passivated state (0) does not experience this charge switching and associated energy loss. Correspondingly, in Figure 6.2 (by definition) F_d is larger in the 1 state than in the 0 state. In this description, it is expected that various rate constants are bias-dependent. This will be discussed in Section 6.4

Differential conductance measurements[17] point to $k_{+\emptyset}$ and $k_{\emptyset+}$ rates on the order of MHz , which corresponds to the ~ 100 ns charge re-organization timescales attributed to IT charging and discharging in Chapter 5. Bi-stabilities attributed to hydrogen adsorption and desorption at a DB site have been measured by STM at the Si(001) surface with timescales on the order of milliseconds to seconds[27, 28], indicating slow k_{01} and k_{10} rates on the order of Hz , which correspond to the Hz timescales that are commonly observed in RTSs, including in this chapter. In fact, the body of STM literature on activation of Si DBs by hydrogen desorption is appreciable, granted that such experiments are performed under cryogenic conditions in the absence of an SiO_2 overlayer. An informative experiment would be to selectively activate a single site by STM and subsequently use the methodology and understanding outlined in Chapters 4 and 5 to measure a corresponding increase in loss at that site using fm-AFM. If a sample can be prepared with appreciable concentrations of interstitial hydrogen, the site may also exhibit RTN which could be associated with hydrogen activation and passivation directly. Furthermore, this experiment would be performed in the absence of SiO_2 , meaning that it could at last resolve the debate over how or whether the oxide participates in RTN.

6.3.2 Structural meta-stabilities of oxide traps

Another multi-state RTS mechanism involves oxide traps (OTs)[19]. In this description, slow RTN is attributed to tunneling between the trap and Si surface charge density. There are subsequent structural relaxations of the trap because its electrostatic environment changes when the charge state switches. Bias dependencies are due to the bias-dependent tunneling probability and the variable structural response of the trap (dipole) under the influence of an electric field. Some OTs can additionally be passivated or de-passivated by hydrogen. This model, however, provides no obvious explanation of mobility fluctuations. In light of this, the first mechanism is considered for the remainder of this chapter. However, the most probable situation is that many kinds of traps and reaction pathways – including structural fluctuations[29], as in the mechanisms described above – contribute to $1/f$ noise in semiconductor devices.

6.4 Bias dependencies

RTS timescales and amplitude are bias-dependent. This section discusses these bias dependencies, with experimental results measured at the same site as in Figure 6.2 at variable V_g . Figure 6.4 shows bias spectra measured near a donor-like and acceptor-like trap. The

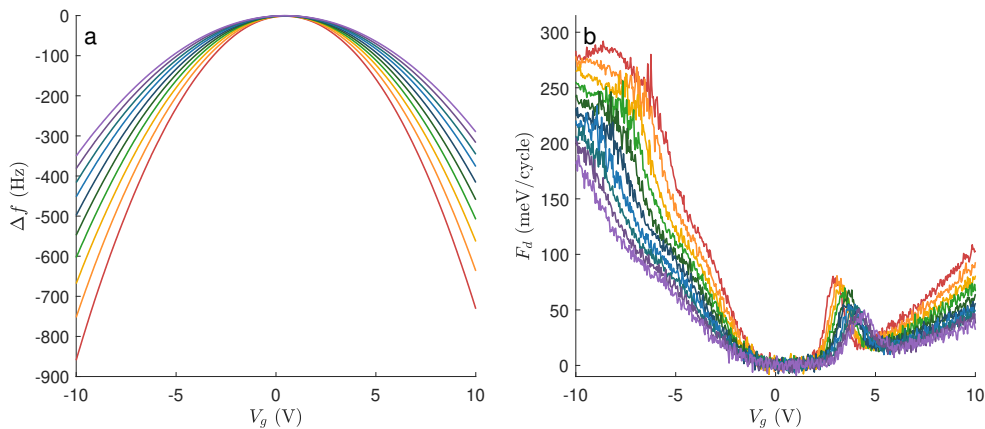


Figure 6.4: **Bias spectra at variable tip lift.** (a) Δf and (b) F_d measured simultaneously at variable tip lift (red = 0 nm, purple = 9 nm, in steps of 1 nm) above a -3 Hz setpoint with $A = 6$ nm. The location was chosen specifically to be near a donor-like and acceptor-like trap; in general, spectra do not exhibit peaks at both positive and negative bias.

acceptor-like peak is not noisy (and in fact, no RTSs corresponding to acceptor-like traps were found for this sample), but the donor-like peak here manifests instead as a noisy part of the spectrum. (The peak for the donor-like trap occurs at a slightly larger bias than that shown in Chapter 5, because this measurement is performed slightly away from the ring center.) This section aims to explain the noise on the donor-like peak.

In Figure 6.4, the noise associated with the donor-like trap is present at tip lifts up to 9 nm. This is definitive evidence that the noise mechanism is not due to tunneling between the tip and the sample, since this would produce a noise amplitude which decays exponentially with tip lift distance.

6.4.1 Reaction rate

Figure 6.5 shows the bias-dependent k_{01} and k_{10} rates (Equation 6.3) corresponding to the location in Figure 6.2, and extracted from the experimental data using the methodology described in Section 6.2. The x-axes for this figure account for the fact that there is a non-linear relationship between V_g and $E_f - E_v$ due to the bias-dependent V_S – this non-linearity is shown in the inset, with the shaded region corresponding to these axes. There appears to be an exponential relationship between the RTS rates and energy. This trend has

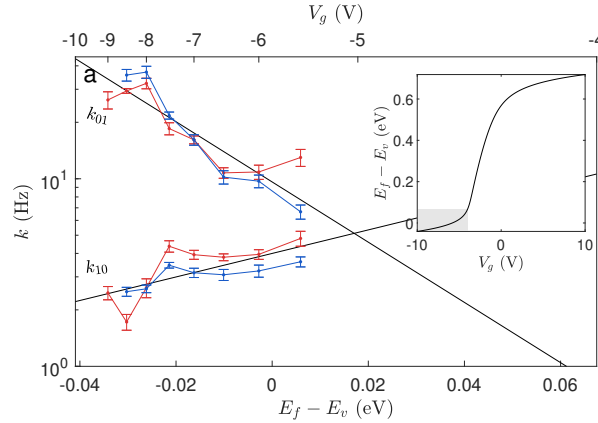


Figure 6.5: **Bias-dependent RTS rates.** Bias-dependent rates corresponding to the Δf (red) and F_d (blue) channels, measured simultaneously above a single trap, showing an exponential relationship between the rates and $E_f - E_v$. The inset shows the nonlinear relationship between the two x-axes.

been observed elsewhere[19, 30, 31], and is discussed below in the context of the hydrogen passivation/activation model presented in Section 6.3.1.

Equation 6.10 includes four processes – k_{01} , k_{10} , $k_{\varnothing+}$, and $k_{+\varnothing}$. Hydrogen absorption and desorption are thermally activated processes[24, 25, 32], and if charge transfer between the surface charge density and the IT occurs by “hopping” (which is the dominant process at room room temperature) rather than tunneling, all four reactions in Equation 6.10 are Arrhenius processes[9]:

$$k = k^{\circ} \exp\left(-\frac{E_A(V_g)}{k_B T}\right) \quad (6.11)$$

where k is the reaction rate (k_{01} , k_{10} , $k_{\varnothing+}$, or $k_{+\varnothing}$), k° is the corresponding attempt frequency (k_{01}° , k_{10}° , $k_{\varnothing+}^{\circ}$, or $k_{+\varnothing}^{\circ}$), E_A is the corresponding activation energy (E_{01} , E_{10} , $E_{\varnothing+}$, or $E_{+\varnothing}$) which may be bias-dependent, and T is the temperature. Figure 6.6 illustrates Equation 6.10 in terms of bias-dependent activation energies for each process. A proposed explanation for the origin of these bias dependencies follows.

With the RTS analysis methodology outlined in Section 6.2, only k_{01} and k_{10} rates can be measured, because these explicitly correspond to the RTS fluctuations. However, given Equation 6.10, k_{10} actually also depends on $k_{\varnothing+}$ and $k_{+\varnothing}$. At large negative biases, $k_{+\varnothing}$ becomes the rate-limiting step for k_{10} – that is, the P_b^* state becomes increasingly energetically favourable at large biases. This is because there is an energy difference between the trap energy (E_T) and E_f such that $E_{+\varnothing} = E_T - E_f$ (as discussed in Chapter 5), and $E_{+\varnothing}$ increases as negative bias increases. The exponential behaviour of k_{10} (rate-limited by $k_{+\varnothing}$) is consistent with Equation 6.11: There is a linear relationship between E_A and $E_f - E_v$; therefore, $\log(k) \propto E_f - E_v$. There may also be bias dependencies in the $k_{\varnothing+}$ and hydrogen passivation (k_{10}) processes. These are not included in this explanation.

The bias dependence of k_{01} has the opposite trend: k_{01} decreases exponentially with $E_f - E_v$. This can be understood in terms of a bias-dependent energy for hydrogen desorption. E_{01} is empirically bias-dependent, given the widely-reported observation by the STM community that a DB can be selectively created at an n-type Si surface by applying a gate bias[17, 32–35]. In STM, desorption is attributed to inelastic scattering of the tunneling electron[32]. In STM, the tip position is (presumed to be) static so it is the tunneling electrons

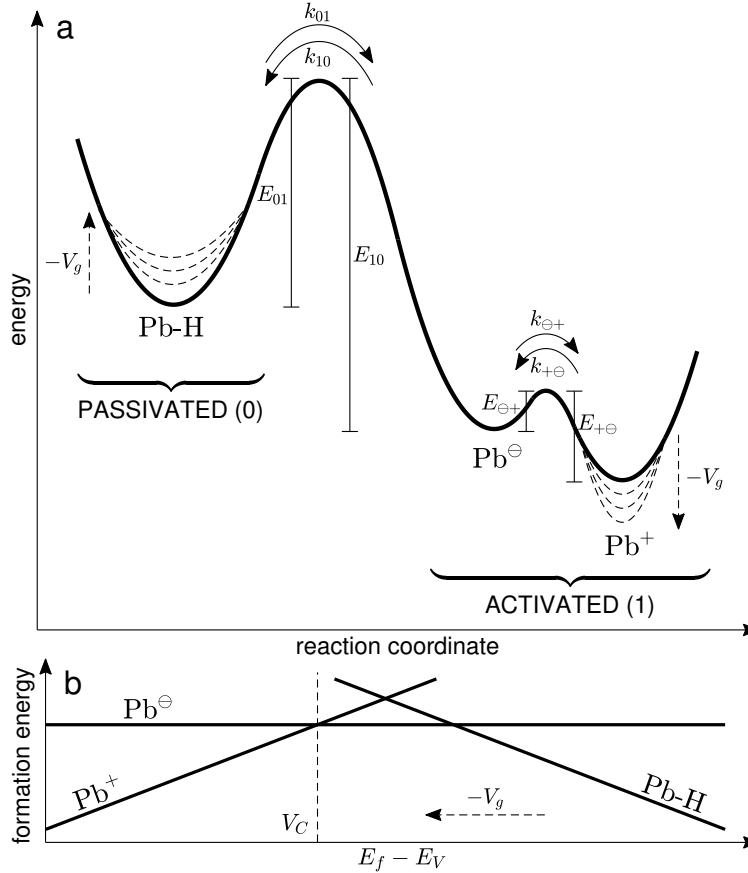


Figure 6.6: **Reaction pathway.** (a) Reaction diagram for hydrogen activation and activation and charging and discharging of an IT (Pb center). The dashed lines illustrate the effect of increasing the negative bias on barrier heights. (b) Formation energy diagram for Pb-H, Pb $^{\ominus}$, and Pb $^{+}$. The dashed arrow indicates the direction of increasing negative bias. The crossing point V_C is shown as a vertical dashed line.

which participate in scattering, but in fm-AFM, the tip position is dynamic, so the surface carriers are mobile and able to scatter. Acknowledging this allows for comparison between the STM literature and these fm-AFM results.

Two hydrogen desorption mechanisms have been identified in STM. In the first mechanism, a hydrogen atom desorbs when $E_T = E_f$, due to the resonant excitation (as charge transfers between the trap and the surface charge density) of a stretching mode of the Si-H bond[27, 28, 32, 36]. In other words, the Si-H vibrational energy, which peaks at a particular bias as identified in Chapter 5, is sufficient to overcome the activation barrier for hydrogen desorption[28]. In the second mechanism, the Si-H vibration is simply excited by dipole scattering when the excitation energy is greater than the vibrational activation energy[36].

For this fm-AFM measurement, hydrogen desorption similar to the second mechanism (a non-resonant effect related to vibrational excitations by defect scattering[37]) would be more likely given the RTS rate dependencies studied here; Figure 6.5, does not exhibit the peak that would be expected in the first mechanism. The resonant excitation process, if it occurs, is likely overwhelmed by significant dipole scattering (which, as opposed to STM, is appreciable, since all of the surface charge density, rather than only the tunneling electrons, can contribute to this excitation). In the second mechanism, however, it is expected that k_{01} should go to zero at large negative bias, where the change in surface potential as the cantilever oscillates goes to zero. This does not agree with the trend in Figure 6.5. This could be due to a bias dependence of the hydrogen desorption activation energy, due to an increase in the dipole moment (i.e. the Si-H interatomic separation increases) at large electric fields.

It should be noted that the hydrogen desorption processes that have been studied in STM occur for crystalline passivated Si surfaces without an oxide overlayer, and desorption can occur at any site under the tip. The ring positions measured here are permanent – that is, the positions of the rings are stable over the course of months of measurements at large positive and negative bias ($V_g = -10 : 10$ V) at room temperature – which indicates preferential hydrogen passivation and desorption at particular sites. This might be due to the oxide overlayer in this sample, where at certain sites, the activation energy for hydrogen desorption and absorption is lowered by the electrostatic landscape of the SiO_2 .

The rates for these Arrhenius processes are temperature-dependent, and decrease significantly under cryogenic conditions. The temperature dependence of RTN rates is found to be dominated by tunneling at low temperatures and hopping at room temperature[3]. In the future, exploring the temperature dependence of single-defect loss would provide useful insights regarding the scattering mechanisms discussed here. Additionally, the activation and passivation rates depend on the interstitial hydrogen concentration. Samples that were prepared with particular attention to reduced interstitial hydrogen concentrations exhibit less noise[8]; this is further evidence that hydrogen plays a critical role in RTN.

6.4.2 Fluctuation amplitude

The two channels measured in this experiment – Δf and F_d – have the same rate dependencies because the signals are correlated with one another (see Section 6.2). However, their amplitudes are not necessarily expected to follow the same trends. This section explains the origin of the bias-dependent amplitude for each channel.

Figure 6.7a-b show bias spectra for the Δf and F_d channels with the mean values and uncertainties for the RTSs in Figure 6.5 overlaid. The Δf RTSs do not perfectly follow the bias spectrum. This is attributable to drift, since even drift of $< 1 \text{ nm}$ (which was the scale of drift in these measurements) leads to significant variability in Δf (see Figure 6.4 for comparison). The Δf results shown in Figure 6.7 have had a background subtraction to account for drift. Figure 6.7c shows the bias-dependent Δf RTS amplitude. In the hydrogen activation/passivation description, the Δf signal here corresponds to the change in surface potential when the Pb center is a DB or bonded to hydrogen. Within the approximations that have been made here, the surface potential is expected to be approximately equal in both states. The $< 1 \text{ nm}$ Δf amplitudes in Figure 6.7 are very small as compared to the absolute values of Δf at these biases ($\sim -400 \text{ Hz}$). To first order, the Δf A_{01} is approximately constant as a function of bias.

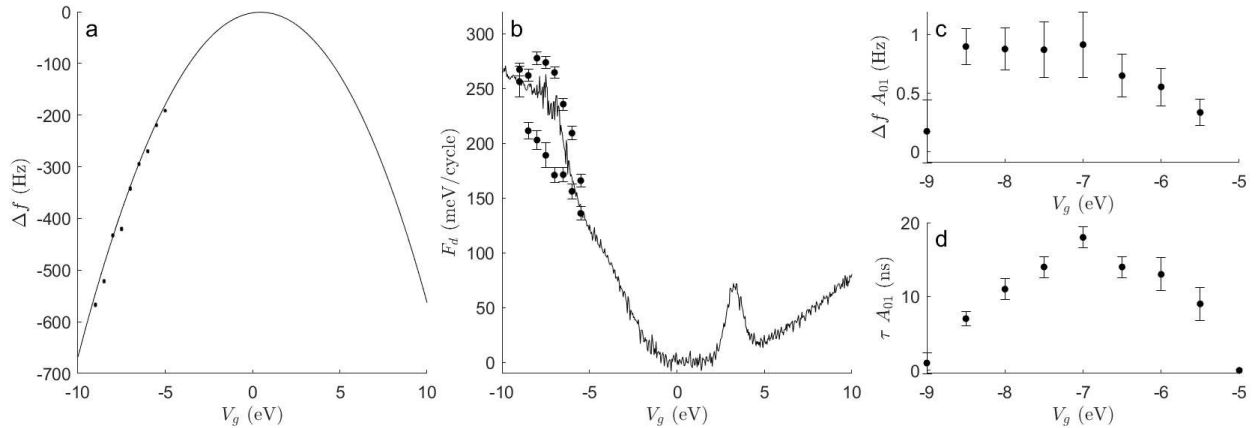


Figure 6.7: **Bias-dependent RTS amplitudes.** (a-b) Δf and F_d bias spectra with RTS measurements overlaid (points, where the value is the gaussian fit mean and the uncertainty is its full-width half-maximum). (c) Δf RTS amplitude and (d) τ RTS amplitude (calculated from the F_d RTSs).

Figure 6.7b shows the bias-dependent F_d and corresponding RTS measurements. The 0 state measurements (bottom) appear to follow one trend, and the 1 state measurements follow another. This is consistent with the observation that rings appear to fluctuate between a high value and a value matching the background (see Figure 6.1). In Chapter 5, it was shown that loss peaks at biases corresponding to the crossing points, due to charge transfer between the trap and surface charge density. When the state is passivated, according to Equation 6.10, this charge transfer does not occur, and so no additional loss (as compared to the background) is expected. Correspondingly, the maximum τ RTS amplitude (Figure 6.7d), where τ is measured given F_d according to the description in Chapter 4, occurs where the system switches between peak sample loss and zero sample loss.

References

- [1] Mandelbrot, B. “Some noises with 1/f spectrum, a bridge between direct current and white noise”. *IEEE Transactions on Information Theory* 13.2 (**1967**). DOI: 10.1109/tit.1967.1053992.
- [2] Press, W. H. “Flicker noises in astronomy and elsewhere.” *Comments on Astrophysics* 7.4 (**1978**).
- [3] Grassler, T., ed. *Noise in Nanoscale Semiconductor Devices*. Springer International Publishing, **2020**. DOI: 10.1007/978-3-030-37500-3.
- [4] Kirton, M. J., Uren, M. J., Collins, S., Schulz, M., Karmann, A., and Scheffer, K. “Individual defects at the Si:SiO₂ interface”. *Semiconductor Science and Technology* 4.12 (**1989**). DOI: 10.1088/0268-1242/4/12/013.
- [5] Haartman, M. von and Östling, M. *Low-frequency noise in advanced MOS devices*. Springer Netherlands, **2007**. DOI: 10.1007/978-1-4020-5910-0.
- [6] Ralls, K. S., Skocpol, W. J., Jackel, L. D., Howard, R. E., Fetter, L. A., Epworth, R. W., and Tennant, D. M. “Discrete resistance switching in submicrometer silicon inversion layers: Individual interface traps and low-frequency (1/f?) noise”. *Physical Review Letters* 52.3 (**1984**). DOI: 10.1103/physrevlett.52.228.

- [7] Machlup, S. “Noise in semiconductors: spectrum of a two-parameter random signal”. *Journal of Applied Physics* 25.3 (**1954**). DOI: 10.1063/1.1721637.
- [8] Kirton, M. and Uren, M. “Noise in solid-state microstructures: A new perspective on individual defects, interface states and low-frequency (1/f) noise”. *Advances in Physics* 38.4 (**1989**). DOI: 10.1080/00018738900101122.
- [9] Costanzi, B. N. and Dahlberg, E. D. “Emergent 1/f noise in ensembles of random telegraph noise oscillators”. *Physical Review Letters* 119.9 (**2017**). DOI: 10.1103/physrevlett.119.097201.
- [10] Mueller, H. and Schulz, M. “Individual interface traps at the Si-SiO₂ interface”. *Journal of Materials Science: Materials in Electronics* 5.6 (**1994**). DOI: 10.1007/bf00215568.
- [11] Wel, A. P. van der, Klumperink, E. A. M., Kolhatkar, J. S., Hoekstra, E., Snoeij, M. F., Salm, C., Wallinga, H., and Nauta, B. “Low-frequency noise phenomena in switched MOSFETs”. *IEEE Journal of Solid-State Circuits* 42.3 (**2007**). DOI: 10.1109/jssc.2006.891714.
- [12] Hung, K., Ko, P., Hu, C., and Cheng, Y. “Random telegraph noise of deep-submicrometer MOSFETs”. *IEEE Electron Device Letters* 11.2 (**1990**). DOI: 10.1109/55.46938.
- [13] Fleetwood, D. M. “Interface traps, correlated mobility fluctuations, and low-frequency noise in metal-oxide-semiconductor transistors”. *Applied Physics Letters* 122.17 (**2023**). DOI: 10.1063/5.0146549.
- [14] Ioannidis, E., Dimitriadis, C., Haendler, S., Bianchi, R., Jomaah, J., and Ghibaudo, G. “Improved analysis and modeling of low-frequency noise in nanoscale MOSFETs”. *Solid-State Electronics* 76 (**2012**). DOI: 10.1016/j.sse.2012.05.035.
- [15] Shi, Z., Mieville, J.-P., and Dutoit, M. “Random telegraph signals in deep submicron n-MOSFET’s”. *IEEE Transactions on Electron Devices* 41.7 (**1994**). DOI: 10.1109/16.293343.
- [16] Contaret, T., Romanjek, K., Boutchacha, T., Ghibaudo, G., and Bœuf, F. “Low frequency noise characterization and modelling in ultrathin oxide MOSFETs”. *Solid-State Electronics* 50.1 (**2006**). DOI: 10.1016/j.sse.2005.10.035.

-
- [17] Labidi, H., Taucer, M., Rashidi, M., Koleini, M., Livadaru, L., Pitters, J., Cloutier, M., Salomons, M., and Wolkow, R. A. “Scanning tunneling spectroscopy reveals a silicon dangling bond charge state transition”. *New Journal of Physics* 17.7 (**2015**). DOI: 10.1088/1367-2630/17/7/073023.
 - [18] Srinivasan, P. “Impact of gate material on low-frequency noise of nMOSFETs with 1.5 nm SiON gate dielectric: Testing the limits of the number fluctuations theory”. *AIP Conference Proceedings*. AIP, **2005**. DOI: 10.1063/1.2036738.
 - [19] Goes, W., Wimmer, Y., El-Sayed, A.-M., Rzepa, G., Jech, M., Shluger, A., and Grassler, T. “Identification of oxide defects in semiconductor devices: A systematic approach linking DFT to rate equations and experimental evidence”. *Microelectronics Reliability* 87 (**2018**). DOI: 10.1016/j.microrel.2017.12.021.
 - [20] Kaczer, B., Toledano-Luque, M., Goes, W., Grassler, T., and Groeseneken, G. “Gate current random telegraph noise and single defect conduction”. *Microelectronic Engineering* 109 (**2013**). DOI: 10.1016/j.mee.2013.03.110.
 - [21] Herring, C., Johnson, N. M., and Walle, C. G. V. de. “Energy levels of isolated interstitial hydrogen in silicon”. *Physical Review B* 64.12 (**2001**). DOI: 10.1103/physrevb.64.125209.
 - [22] Cartier, E., Stathis, J. H., and Buchanan, D. A. “Passivation and depassivation of silicon dangling bonds at the Si/SiO₂ interface by atomic hydrogen”. *Applied Physics Letters* 63.11 (**1993**). DOI: 10.1063/1.110758.
 - [23] Brower, K. L. “Dissociation kinetics of hydrogen-passivated (111) SiO₂ interface defects”. *Physical Review B* 42.6 (**1990**). DOI: 10.1103/physrevb.42.3444.
 - [24] Liu, X.-H., Xie, W.-F., Liu, Y., and Zuo, X. “Passivation and dissociation of P_b-type defects at a-SiO₂ interface”. *Chinese Physics B* 30.9 (**2021**). DOI: 10.1088/1674-1056/ac0e20.
 - [25] Stesmans, A. “Dissociation kinetics of hydrogen-passivated P_b defects at the (111)Si/SiO₂ interface”. *Physical Review B* 61.12 (**2000**). DOI: 10.1103/physrevb.61.8393.

-
- [26] Tsetseris, L. and Pantelides, S. T. “Migration, incorporation, and passivation reactions of molecular hydrogen at the Si-SiO₂ interface”. *Physical Review B* 70.24 (2004). DOI: 10.1103/physrevb.70.245320.
- [27] Bellec, A., Riedel, D., Dujardin, G., Boudrioua, O., Chaput, L., Stauffer, L., and Sonnet, P. “Nonlocal activation of a bistable atom through a surface state charge-transfer process on Si(001)-2×1:H”. *Physical Review Letters* 105.4 (2010). DOI: 10.1103/physrevlett.105.048302.
- [28] Stokbro, K., Quaade, U. J., Lin, R., Thirstrup, C., and Grey, F. “Electronic mechanism of STM-induced diffusion of hydrogen on Si(100)”. *Faraday Discussions* 117 (2000). DOI: 10.1039/b003179h.
- [29] Ralls, K. S. “Tunnel-junction noise at low temperatures: Tunneling or local “heating””. *Physical Review B* 46.7 (1992). DOI: 10.1103/physrevb.46.4295.
- [30] Lee, J.-W., Shin, H., and Lee, J.-H. “Characterization of random telegraph noise in gate induced drain leakage current of n- and p-type metal-oxide-semiconductor field-effect transistors”. *Applied Physics Letters* 96.4 (2010). DOI: 10.1063/1.3292204.
- [31] Amarasinghe, N. V., Çelik-Butler, Z., and Vasina, P. “Characterization of oxide traps in 0.15 μm^2 MOSFETs using random telegraph signals”. *Microelectronics Reliability* 40.11 (2000). DOI: 10.1016/s0026-2714(00)00089-5.
- [32] Shen, T. C., Wang, C., Abeln, G. C., Tucker, J. R., Lyding, J. W., Avouris, P., and Walkup, R. E. “Atomic-scale desorption through electronic and vibrational excitation mechanisms”. *Science* 268.5217 (1995). DOI: 10.1126/science.268.5217.1590.
- [33] Møller, M., Jarvis, S. P., Guérinet, L., Sharp, P., Woolley, R., Rahe, P., and Moriarty, P. “Automated extraction of single H atoms with STM: tip state dependency”. *Nanotechnology* 28.7 (2017). DOI: 10.1088/1361-6528/28/7/075302.
- [34] Pavliček, N., Majzik, Z., Meyer, G., and Gross, L. “Tip-induced passivation of dangling bonds on hydrogenated Si(100)-2 × 1”. *Applied Physics Letters* 111.5 (2017). DOI: 10.1063/1.4989749.

- [35] Stock, T. J. Z. et al. “Atomic-scale patterning of arsenic in silicon by scanning tunneling microscopy”. *ACS Nano* 14.3 (**2020**). DOI: 10.1021/acsnano.9b08943.
- [36] Persson, B. N. J. “Inelastic vacuum tunneling”. *Physica Scripta* 38.2 (**1988**). DOI: 10.1088/0031-8949/38/2/034.
- [37] Jech, M. et al. “Quantum chemistry treatment of silicon-hydrogen bond rupture by nonequilibrium carriers in semiconductor devices”. *Physical Review Applied* 16.1 (**2021**). DOI: 10.1103/physrevapplied.16.014026.

Chapter 7

Comparison systems

Silicon, undoubtedly, is the most prominent and widely studied semiconductor today – it is truly the “textbook” semiconductor – but countless other semiconducting materials which exhibit exotic electronic properties are continually being discovered. Carbon nanotubes, for example, are one-dimensional materials which can be semiconducting or metallic depending on how they are wrapped[1]. Graphene, another famous material, can switch from conducting to semiconducting when it is doped, cut, or strained in particular ways[2].

Semiconductors can be one, two, or three dimensional; they can be organic or inorganic; large or small; but, by definition, they all share a common feature: Their Fermi level lies within a band gap. This was the starting premise of Chapter 2. Exotic semiconductors (which are, in some sense, anything that is not bulk silicon) might deviate in various ways from the basic description of semiconductor physics outlined in Chapter 2, but on the whole, they obey the same basic principles: They have n-type and p-type carriers, they can be doped, they have temperature-dependent carrier densities, and they have intermediate screening as compared to metals and insulators.

In this chapter, two semiconducting materials – MoSe_2 and pentacene – are measured by fm-AFM and compared to silicon. Bias spectra are shown and compared to the MIS model, and defects are identified as being associated with loss and RTN.

7.1 Other semiconductors

The samples studied here – MoSe₂ and pentacene, along with silicon – are three very different kinds of semiconductors. This section presents a brief overview of each material. The KPFM images that are shown were used to determine the sample heights indicated – the height measurement methodology is explained in detail in Appendix A.3.

7.1.1 MoSe₂

Transition metal dichalcogenides (TMDCs) are two-dimensional materials which are comprised of transition metal atoms (Mo, W, Ta, Ti, etc.) sandwiched between chalcogen atoms (O, S, Se, Te). TMDCs are usually found as layered structures, with strong interatomic binding within each layer and weak van der Waals coupling between the layers (like graphene). MoS₂ and WS₂ are two of the most widely studied TMDCs today, but MoSe₂ – which, due to the higher metallicity of the Se atom as compared to S, has higher conductivity – is garnering increasing interest because it is (perhaps) better-suited for an assortment of energy storage, catalysis, and optoelectronic applications[3].

The layers of MoSe₂ have a hexagonal lattice structure, and therefore MoSe₂ tends to be triangular in shape. Figure 7.1 shows a KPFM image of a triangular MoSe₂ island on a 300 nm SiO₂ substrate. The island is stepped – the left edges are lower than the right edges – and it has a decorated left edge.

Figure 7.1b shows that the flatband potential of the island is very spatially heterogeneous. The two lower layers have smaller V_{FB} , the decorated left edge has the largest V_{FB} , and the highest layer has patches and edges which have significantly lower V_{FB} . It is likely that these patches and edges are residual PMMA from the fabrication process (this sample was prepared by all-dry viscoelastic stamping, where layers of MoSe₂ are transferred between substrates using stamps and tape). Despite their obvious difference in V_{FB} as compared to the top layer, these patches did not appear to otherwise affect the electronic properties of this sample, so it is unlikely that they are charged, traps, or metallic in nature. (This will be revisited in Section 7.3.)

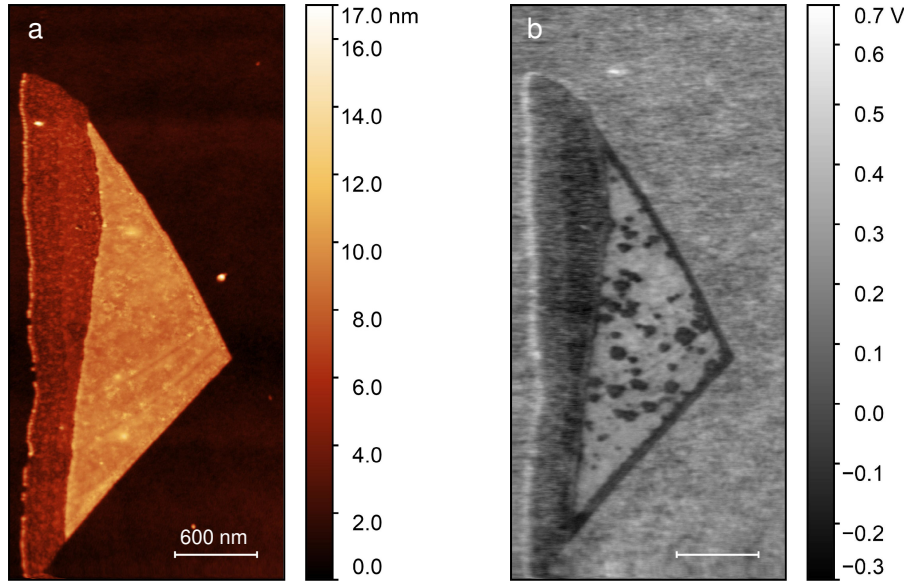


Figure 7.1: **MoSe₂ sample.** KPFM image of an exfoliated MoSe₂ island on 300 nm SiO₂, showing (a) Z and (b) V_{FB} . The left-most section of the island is the “first layer (height = 3.1 ± 0.5 nm) and the tallest section is the “top layer (height = 9.0 ± 0.4 nm). The island is slightly rougher than the substrate (height = 0.0 ± 0.2 nm).

The step between the first and second island layers is consistent with a single monolayer (0.7 nm), but the apparent first layer height relative to the substrate is 3.1 ± 0.5 nm. Measurements of other regions of this sample with Raman and photoluminescence spectroscopies were consistent with a monolayer[4], so it is likely that the apparently large height is due to either a rough overlayer or underlayer, which possibly also arose during the sample fabrication procedure. The surface roughness is also much greater than expected for an atomically flat sample, suggesting the presence of a rough over or underlayer. The top layer height (height = 9.0 ± 0.4 nm) undeniably signifies multilayer MoSe₂, of between 6-12 layers[5].

7.1.2 Pentacene

Organic semiconductors are made up of carbon-based pi-bonded molecules or polymers. In organic semiconductors, molecules are analogous to the atoms of inorganic semiconductors: When many molecules are brought together, the molecular orbitals interact and form energy bands. Specifically, the highest occupied molecular orbitals combine to form the valence band, and the lowest unoccupied molecular orbitals form the conduction band. The main

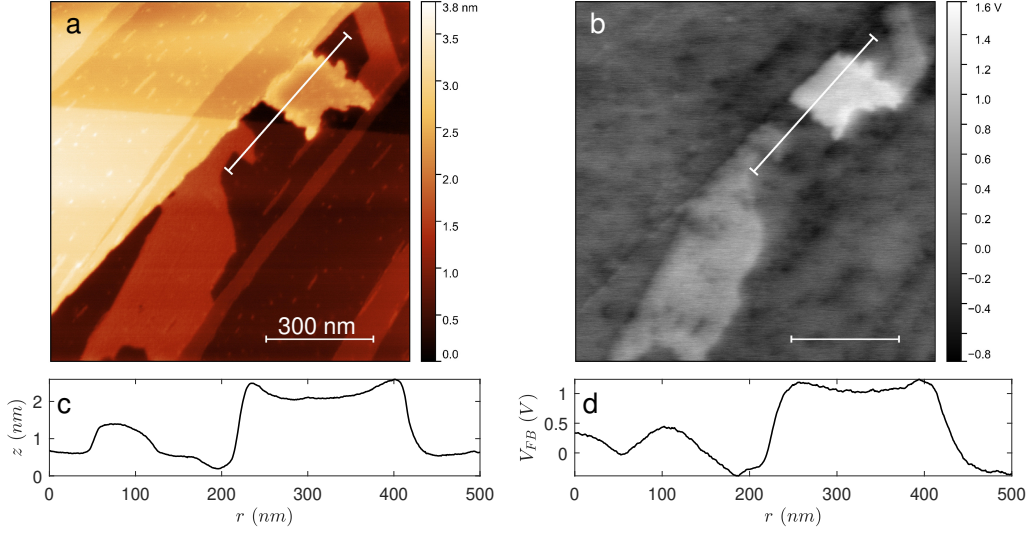


Figure 7.2: **Pentacene sample.** KPFM image of a pentacene island on KBr(001), showing (a) Z and (b) V_{FB} . The line traces (c,d) correspond to the white angled line in (a,b). Two islands can be seen: The left is thinner (height = 0.7 ± 0.1 nm) and the right thicker (height = 1.3 ± 0.2 nm). The approximately straight-edged areas are KBr terraces.

practical advantage of organic semiconductors is that they are lightweight, thin, and inexpensive to fabricate. In large part because of these characteristics, they are found in many existing technologies today, such as televisions, electronic paper displays, and solar panels.

Pentacene is an example of an organic semiconductor. Figure 7.2 shows a KPFM image of two pentacene islands on KBr. The small island is significantly taller than the large one, and they have different V_{FB} . The KBr step between the terraces underlying the islands have a step height of 0.44 ± 0.05 nm, which is consistent with a single atomic KBr step[6]. (This means that the calibration of the z -piezo is reliable.) The height of the thin pentacene island (0.7 ± 0.1 nm) is approximately consistent with a flat lying pentacene monolayer[6], which is observed in submonolayer deposition (as is the case here). The height of the tall island (1.3 ± 0.2 nm) corresponds either to a standing pentacene monolayer[6, 7] or to a flat-lying stacked bilayer. The tall island appears structurally non-uniform at its edges, with edge thicknesses that are ~ 0.5 nm taller than the island center. Figure 7.2c shows a line profile demonstrating this thickening at the island edges. Similar tall-edged islands have been measured for C60[8], and attributed to a different stacking number or molecular organization at the island edge.

V_{FB} for the KBr appears largely uniform, except that there are “pock marks” scattered throughout. These could be due to atomic point defects in the KBr substrate. This sample was fabricated entirely in UHV by cleaving bulk KBr along a (001) plane, annealing the KBr substrate, and subsequently evaporating pentacene. A quartz microbalance was used to calibrate the evaporation rate to achieve sub-monolayer coverage.

7.2 Bias spectroscopy

Figure 7.3 shows bias spectra of SiO_2 , MoSe_2 , and pentacene. The most striking feature, when comparing these spectra, is that $\Delta f(V_g)$ all have varying degrees of non-parabolicity. The non-parabolicity in a semiconductor force measurement is due to the nonlinearity of $V_S(V_g)$ (recall Chapter 2). The non-parabolicity is most obvious when the derivative of $V_S(V_g)$ is large, and when the ‘kink’ in $V_S(V_g)$ occurs far from $V_g = 0$ V (see the pentacene measurement). The most significant predictor of non-parabolicity in the MIS model is E_g : When E_g is large, the carrier populations in the conduction and valence bands exponentially decrease (Equations 2.8 and 2.10). Consequently, the carrier populations at the surface also decrease (Equation 2.18). This means that the screening of the semiconductor decreases, so the potential drop inside the semiconductor (V_S) increases. Also note that the Si/ SiO_2 and MoSe_2 curves are n-type, as evidenced by the fact that their non-parabolicity manifests at negative biases; the pentacene sample is p-type.

Table 7.2 lists all of the fm-AFM MIS capacitor parameters for each sample. There are many parameters, and so the risk of over-fitting is high, so where possible, parameters were constrained to known values. For the silicon sample, $E_g = 1.1$ eV[9] proved to be a significant over-estimation: $E_g = 0.7$ eV was the largest band gap which gave a relative uncertainty of less than 10%. It could be that this smaller-than-expected band gap is due to band gap narrowing at the surface due to the appreciable trap density at the surface of this sample[9–11]. $\epsilon = 11.7$ [9], and $\chi = 4.05$ [9] correspond to known values for silicon, N_d was a fit parameter, and $N_a = 0$ is an approximation given that the sample is n-type. For multilayer MoSe_2 , $E_g < 1$ eV[12], which is not to be confused with $E_g = 1.55$ eV for a monolayer[13, 14], since in MoSe_2 E_g is highly layer-dependent. $\epsilon = 7.5$ (for ~ 9 nm thick

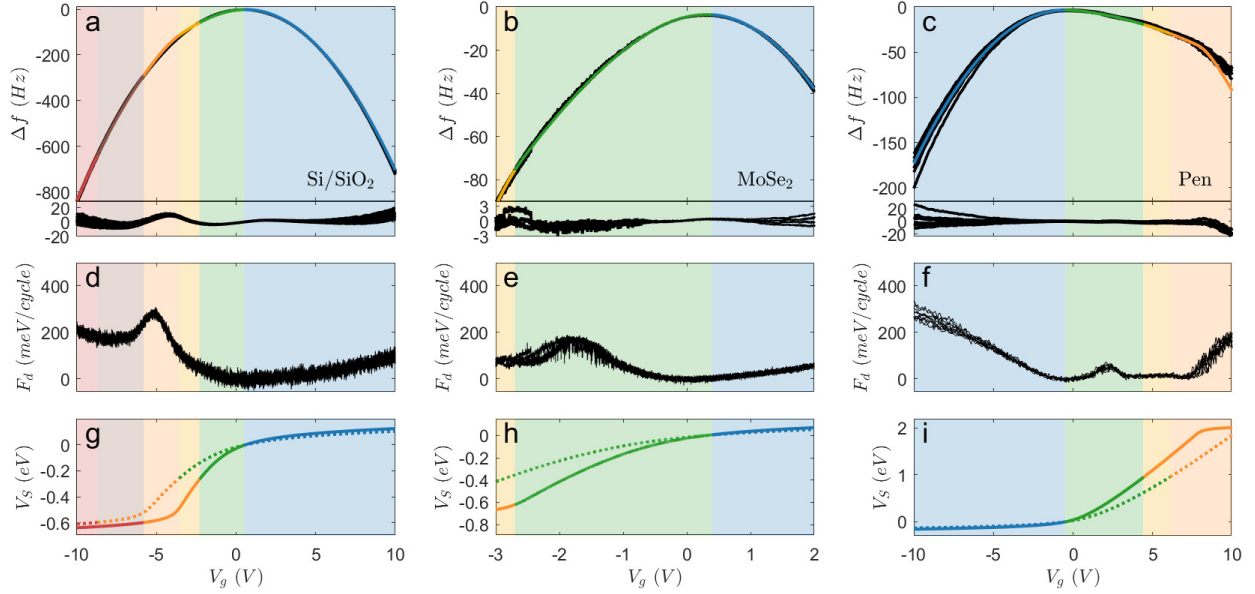


Figure 7.3: **Bias spectra for semiconducting samples.** Bias spectra for (a,d,g) SiO₂, (b,e,h) MoSe₂, and (e,c,i) pentacene. Data is in black (between 4-10 sweeps for each sample) and the modelled results are in colour (where red, brown, orange, yellow, green, and blue correspond to the ss, sw, ww, wd, dd, and aa regimes as defined in Chapter 4). (a-c) Δf and its residual ($\Delta f_{\text{model}} - \Delta f_{\text{data}}$, which are all $< 10\%$ relative difference), and (d-f) and V_S at the bottom (solid) and top (dashed) of the cantilever oscillation (model only). Note that the x-axis range is smaller for MoSe₂ than for the other samples.

<i>Sample p.</i>	Si/SiO ₂	MoSe ₂	Pen	<i>fm-AFM p.</i>	Si/SiO ₂	MoSe ₂	Pen
E_g (eV)	0.7	0.8	2.2	$z_{\text{ins},c}$ (nm)	12	9.0	9.1
ϵ	11.7	7.5	5.0	A (nm)	6	6	6
χ (eV)	4.05	3.72	2.88	Q	18000	18000	18000
N_d (cm ⁻³)	5.00e17	4.90e17	0	f_o (kHz)	310	330	310
N_a (cm ⁻³)	0	0	5.5e16	k (N/m)	42	42	42
m_e (m_o)	1	1	1	Φ_T (eV)	4.75	4.15	4.75
m_h (m_o)	1	1	1	r (nm)	5	5	5
α	0	0.05	0.6	z_c (μm)	6	6	6
T (K)	300	300	300	a_c (μm^2)	2900	2900	2900

Table 7.1: **MIS fm-AFM parameters for Si/SiO₂, MoSe₂, and pentacene (Pen).** Parameters corresponding to the models in Figure 7.3. The sample parameters (*Sample p.*) are: Band gap (E_g); permittivity (ϵ); electron affinity (χ), dopant concentration ($N_{a,d}$), where if $N_a = 0$ the sample is n-type and vice-versa for p-type; effective electron and hole masses ($m_{e,h}$); potential lever arm for backing substrate (α); and temperature (T). The fm-AFM parameters (*fm-AFM p.*) are: Closest tip-sample separation ($z_{\text{ins},c}$); oscillation amplitude (A); Q-factor (Q); free resonance frequency (f_o); tip work function (Φ_T); effective tip radius r ; effective cantilever height (z_c); and effective cantilever area (a_c).

MoS₂)[15], and $\chi = 3.5$ eV[16]. The sample is also n-type, so $N_a = 0$ and N_d is a fit parameter. For pentacene, $E_g = 1.2 - 2.8$ eV[17], $\epsilon = 2 - 5$ [18], and $\chi = 2.7$ eV[17], $N_d = 0$ and N_a is a fit parameter.

The effective masses were set as 1 for all samples. The effective masses are not actually expected to equal 1, but the effective mass is a very insensitive parameter (see Appendices B.2, B.3, and B.4 for an assessment of the Si/SiO₂, MoSe₂, and pentacene fit sensitivities for all sample parameters), and the samples studied here have anisotropic band edges, so a singular effective mass value is already an approximation. The value of 1 was chosen because it is a reasonable order of magnitude for the samples listed here. α is a fit parameter to account for the potential drop across the back insulator. In a “textbook” MIS capacitor, the back side of the semiconductor is connected by an Ohmic contact to a metal electrode. In the MoSe₂ and pentacene samples, however, the semiconductor is backed by a thick insulator (300 nm SiO₂ and > 500 μ m KBr, respectively). The potential drop across this backing capacitance is related to the charge density in the semiconductor; larger substrate thicknesses lead to a larger reduction in the potential across the vacuum gap and the semiconductor. This is accounted for by approximating $V_g = V_g^*(1 - \alpha)$, where V_g is the effective applied bias across the MIS capacitor and V_g^* is the total applied bias. For the silicon sample, $\alpha = 0$, since the back of the semiconductor has an Ohmic contact. For the top layer of the MoSe₂ island, $\alpha = 0.05$ (i.e. the surface charge density at the back of the substrate is 5% of that at the front), and for pentacene (which is expected to achieve similar surface charge densities as for MoSe₂, but which has a larger drop across the back insulator because it is much thicker), $\alpha = 0.6$. The temperature was assumed to be 300 K: Day-to-day fluctuations in room temperature are negligible within measurement sensitivity for the MIS model.

The fm-AFM parameters are also listed in Table 7.2. $z_{ins,c}$ varied from sample to sample and experiment to experiment (with a range of ~ 5 nm), depending on the setpoint frequency (which was generally between $-2 : -5$ Hz) and the value of V_g during approach. The setpoint amplitude A is assigned before every experiment. Q and f_o were measured by sweeping the drive frequency of a free cantilever, measuring the amplitude at constant F_d , and fitting a damped, driven harmonic oscillator model to the sweep (as in Figure 3.2). The probes used for these measurements were metal-coated (Pt-Ir) metallic Si tips (Nanosensors PPP-

NCHPt[19]). k was measured by the cantilever manufacturer. r , a_c , and z_c are fit parameters that were found by performing bias sweeps at variable height (shown in Appendix B.1): At very large tip lifts ($> 10 \text{ nm}$), the effective contribution to the total force from the tip is very small as compared to the cantilever background, which allows for the relative contributions of the tip and cantilever – i.e. r , $z_{ins,c}$, a_c , and z_c – to be determined. r found using this approach falls within the manufacturer-quoted specifications (“less than 7 nm ”), as do a_c and z_c [19]. Φ_T is a fit parameter, and is reasonable given that Pt and Ir have work functions on the order of 5 eV . The fm-AFM parameters are all the same, except Φ_T and f_o for MoSe₂. This degree of tip-to-tip variability is expected, since Φ_T depends on the tip geometry and f_o on the cantilever geometry.

It should also be noted that the 2 nm SiO₂ overlayer was simply included in the total insulator (vacuum gap) capacitance, meaning that it was assigned $\epsilon = 1$. In reality, its value is closer to 3.9. However, over such a small distance, this effect was found to be minor: The main consequence of not accounting for the SiO₂ in this way is that the $z_{ins,c}$ value reported here is likely a slight over-estimation. This discrepancy ($< 1 \text{ nm}$) falls within the uncertainty of the tip-sample position, given that every bias spectroscopy experiment is accompanied by some vertical drift.

All of the residuals correspond to relative uncertainties of $< 10\%$. The residual curves are somewhat flat, except for the notable bump in the Si/SiO₂ curve (which can be attributed to an over-estimation of the band gap, discussed previously), and the deviation in pentacene at large bias, which will be addressed in Section 7.3. The Si/SiO₂ sample was intentionally doped, but this is not the case for the MoSe₂ and pentacene samples. Rather, these systems were unintentionally doped by defects. This will be discussed in Section 7.3. The peaks in the F_d spectra in Figure 7.3 were not present everywhere in these samples: They, like for the SiO₂ sample, are only found at certain locations, and therefore are associated with defects.

7.2.1 Sample thickness dependence

The MoSe₂ and pentacene measurements were both sensitive to the sample thickness. The bias spectra presented in this chapter were generally measured on the thicker islands for both samples, except for Figure 7.4, which shows a large decrease in the total tip-sample force

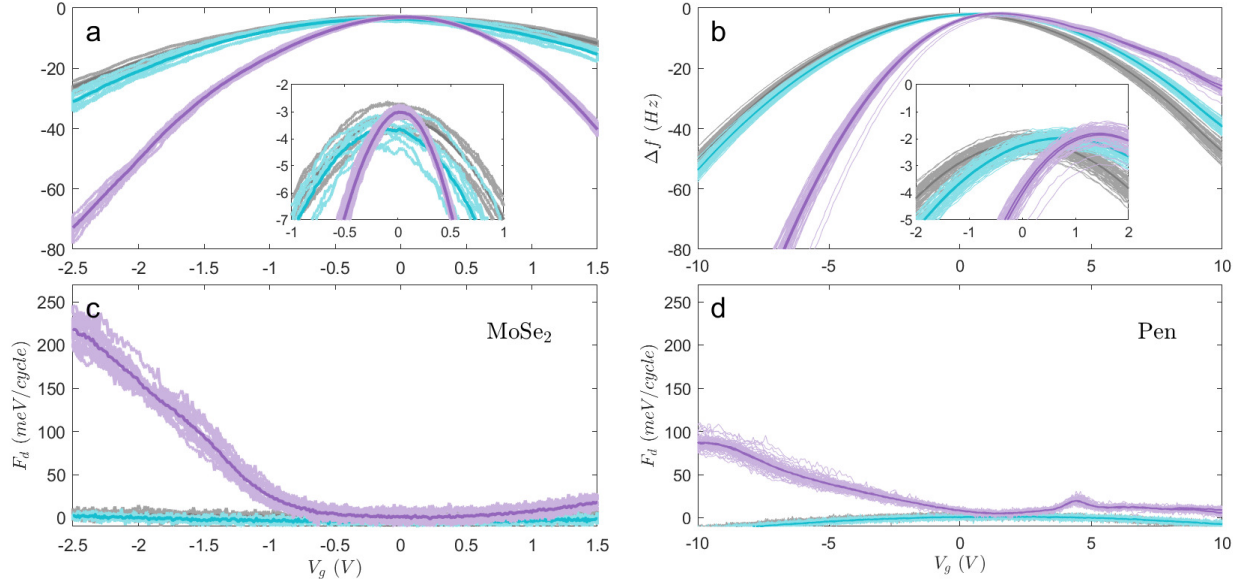


Figure 7.4: **Sample thickness-dependent screening.** Bias spectra measured for the MoSe₂ sample (left) and pentacene (Pen) sample (right). 10 sweeps are shown for MoSe₂ and 50 for pentacene, and their average is the dark curve. Grey corresponds to the substrate (300 nm SiO₂ and bulk KBr), blue is the thinnest part of the sample (the first MoSe₂ island step and a monolayer pentacene island), and purple is the highest part of the sample (the top layer of the MoSe₂ island and multilayer pentacene.) The insets show $\Delta f(V_g)$ near V_{FB} .

and a notable decrease in non-parabolicity for the thinner islands. This is an indication of a confinement effect: If the material is spatially confined, its ability to establish a surface potential is significantly reduced; and when V_S is smaller, the non-parabolicity in $F_{ts}(V_g)$ and $\Delta f(V_g)$ is smaller, as explained previously. This decreased capacity for screening is an important characteristic of dimensionally constrained semiconducting systems. If the sample is thinner than the depletion width (as is the case here for MoSe₂ and pentacene), this will significantly affect the charge densities at the front and back of the semiconductor, and therefore also the potential across each insulator; the MIS capacitor model does not account for this. Despite this obvious shortcoming of the MIS model for thin samples on insulating substrates, it still appears to capture the overall behaviour of the thicker systems well.

The insets of Figure 7.4 show V_{FB} for the various layers in each sample. In the MoSe₂ sample, V_{FB} is approximately equal on the substrate, first layer, and top layer of the island. This is consistent with Figure 7.1. In the pentacene sample, there is a slight difference between V_{FB} on the substrate, thin island, and thick island. This is consistent with Figure 7.2.

7.3 Defects

All three samples shown here – Si/SiO₂, MoSe₂, and pentacene – exhibited spatially heterogeneous dielectric loss (i.e. peaks in the $F_d(V_g)$ spectrum which were not present everywhere). This was discussed in Chapter 4 for Si/SiO₂, and is compared to MoSe₂ and pentacene below. For each system, this spatial heterogeneity can be attributed to spatially localized defects.

7.3.1 MoSe₂ rings

A multipass image of the corner section of the MoSe₂ sample is shown in Figure 7.5. In comparing with Figure 7.1, only the top layer appears “bright” in F_d , whereas the other two steps of the MoSe₂ island are indistinguishable from the SiO₂ background. The lower F_d of the thinner MoSe₂ steps corresponds with an observation in the previous section that for thin samples, $\Delta f(V_g)$ spectra are approximately parabolic: ΔV_S over the cantilever oscillation is necessarily small in the thinner regions, so there is reduced carrier motion, meaning less loss and smaller F_d .

Empirically, TMDC mobilities are lower than theoretical predictions, and impurity Coulomb scattering is thought to be a major contribution to this loss[20]. Figure 7.5 is an indication that this is likely the case. The top MoSe₂ island layer shows spots at low bias and rings at larger bias, which corresponds to scattering at defect sites. A defect which could be the origin of this loss is interstitial Mo or Se atoms, both of which introduce mid-gap states in the lower half of the band gap in MoSe₂[12]. Due to this energetic position, they are donor-like,

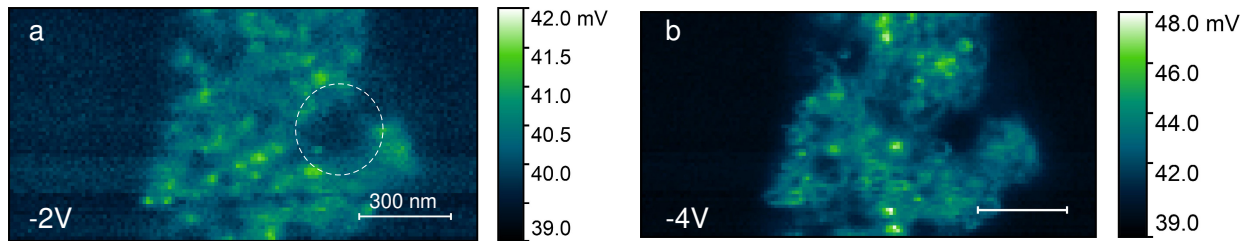


Figure 7.5: **MoSe₂ rings.** Multi-pass images of the island in Figure 7.1, showing F_d at variable V_g (indicated). The top layer of the island (except small areas, such as the circled region) appears bright; the two lower steps of the island, including the decorated edge on the far left of Figure 7.1, appear dark (and indistinguishable from the substrate).

so they are expected to interact with the surface charge density significantly at negative bias (so these states could be the origin of the unintentional n-doping of this sample).

Overall, interstitial atomic defects are expected to behave similarly to the donor-like Si/SiO₂ ITs discussed in Chapter 5, and the loss peaks in Figure 7.5 occur when there is the most interaction between the trap and the surface charge density when their Fermi levels align. The energies of the traps shown in Figure 7.5 can be found by the same process as in Chapter 5 (that is, by identifying crossing points as corresponding to $F_d(V_g)$ peaks and using the fm-AFM MIS model to determine the corresponding trap energy). The F_d peaks in Figure 7.5 occur where $V_g \sim -2$ V, which corresponds to crossing points $V_C = -2.5$ V (top of the oscillation) and $V_C = -1.5$ V (bottom of the oscillation), as demonstrated in Figure 7.6. This analysis gives a trap energy approximately 0.4 eV above the valence band edge, which is similar to DFT calculations which find interstitial defect levels ~ 0.5 eV above the valence band edge[12].

The multilayer MoSe₂ “background” (i.e. regions apparently devoid of point defects, such as the circled region in Figure 7.5) appear to have comparatively low F_d . This means that in this system, there is very little loss associated with charge re-organization where defects are not present (as compared to Si/SiO₂, where in Chapter 4 it was shown that there is bias-dependent loss even far from any defect). The patches on the top MoSe₂ layer (recall the V channel of the KPFM image in Figure 7.1b) do not appear in Figure 7.5 except for a very slight darkening in the spot regions. This is consistent with a slightly larger potential drop across the insulator, which is expected if these patches are PMMA residue as previously

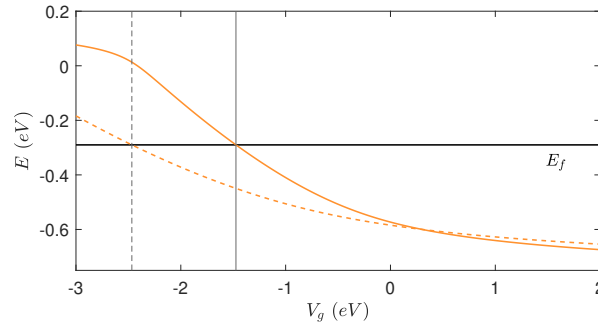


Figure 7.6: **MoSe₂ crossing points.** Bias-dependent energy at the bottom (solid) and top (dashed) of the cantilever oscillation, for a donor-like trap with energy 0.4 eV above the valence band edge. E_f is the sample Fermi energy.

proposed. The circled region in Figure 7.5 does not correspond to one of these patches.

Another notable observation regarding Figure 7.5 is that, as was the case for Si/SiO₂, multipass imaging at positive bias was largely spatially uniform, with no apparent spots or rings. This indicates that acceptor-like states do not play a large role in loss in this system, possibly (like Si/SiO₂) because only acceptor-like states very close to the conduction band edge can be measured with this technique: Very large biases would need to be applied in order to achieve sufficient band bending in the accumulation regime to access deeper-level states. Finally, metallic edge states have been reported in similar systems[21], but such edge states, if present, are not detectable here. The background of Figure 7.5 (~ 39 mV) corresponds to the intrinsic cantilever damping, so any increase in mobility with respect to this background falls below the measurement sensitivity.

7.3.2 Pentacene edges

Figure 7.7 shows a grid spectroscopy image of the pentacene islands in Figure 7.2 (note that the multipass image is slightly “zoomed in” and rotated with respect to the KPFM image). The inverted contrast in Δf of the pentacene island with the KBr background is simply because their $\Delta f(V_g)$ parabolas intersect one another (see Figure 7.8). The thin pentacene island is located directly to the left of the thick island, as seen in Figure 7.2. However, Figure 7.7 shows that this island is indistinguishable from the background. This is consistent with Figure 7.4, and the associated discussion.

F_d is non-uniform over the pentacene island. In particular, F_d appears to be largest at the island edges, and is zero (within measurement uncertainty) at the center of the island (see Figures 7.7 and 7.8). In pentacene, mid-gap states arise due to structural imperfections in the islands[22, 23]. This, combined with the fact that this sample was prepared entirely in UHV without ever being exposed to air, means that structural defects, rather than contaminants, are a likely origin of the loss seen here. This is consistent with the observation in Section 7.1 that island edges are slightly taller, corresponding to structural variations at the island edges. This is consistent with the observation that the pentacene mobility decreases when the grain size decreases, and is consistent with theoretical predictions that mobility decreases at grain boundaries and stacking faults in pentacene films[24].

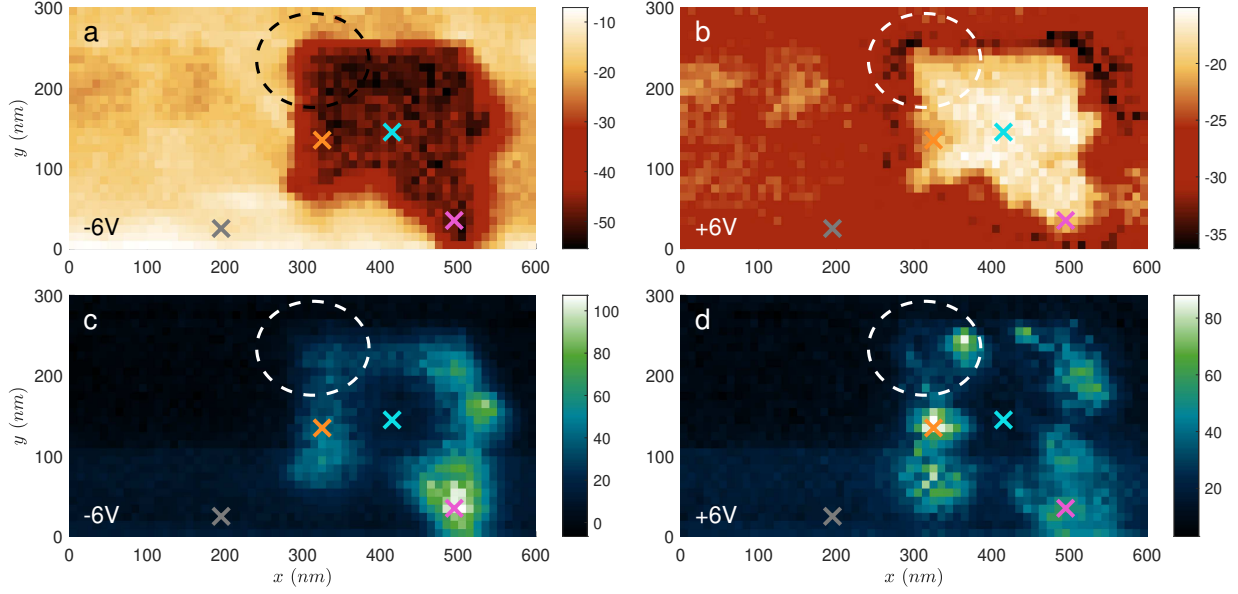


Figure 7.7: **Spatial variability of pentacene loss.** Grid spectroscopy of a small pentacene island, showing Δf (top) and F_d (bottom) at variable V_g (indicated). At each pixel a full spectrum was acquired; the images shown here are “slices” at the V_g indicated. The crosses correspond to the spectra shown in Figure 7.8. The dashed circle is shown to guide the eye to one corner of the pentacene island for easier comparison between the four images. The monolayer is actually also in this image, directly to the left of the dotted circle, as seen in Figure 7.2, but is indistinguishable from the KBr background in both channels.

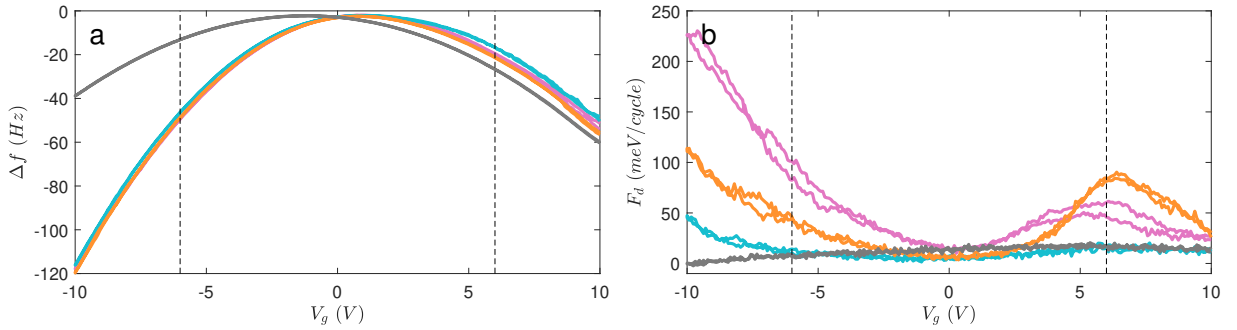


Figure 7.8: **Spatially variable pentacene bias spectra.** (a) $\Delta f(V_g)$ and (b) $F_d(V_g)$ spectra corresponding to the crosses in Figure 7.7. (Grey is the KBr background, orange and pink are at the island edges, and blue is at the island center.)*

* Figure 7.8 provides another opportunity for an fm-AFM methodology check. Nonlinearities in the drive piezo frequency response (i.e. frequency-dependent phase shifts) can lead to coupling between Δf and F_d [25], which are artefacts that are easy to mistake for signals. In Figure 7.8, the blue, orange, and pink curves nearly perfectly overlap in Δf , and yet have very different $F_d(V_g)$ spectra. If the $F_d(V_g)$ non-constancy were due to this artefact, the $F_d(V_g)$ curves would also all overlap with each other; therefore, this can be ruled out as the origin of the signals measured here. Similar checks were done for all of the F_d measurements shown in this work, and this known transfer function artefact cannot explain these measurements.

Figures 7.4 and 7.8 show that pentacene has an F_d peak at positive bias. This, like Si/SiO₂ and MoSe₂, means that loss corresponds to a defect that is spatially localized in x and y – that is, though loss is more prevalent near the island edges, it does not appear to correspond to uniform scattering at the boundary. Based on the discussion of Chapter 5, peaks in a $F_d(V_g)$ spectrum correspond to rings when imaged at variable V_g with the z-controller off. They are hard to see in the grid spectroscopy in Figure 7.7, due to the relatively coarse spacing of the grid, but they are clearly visible in the multipass image in Figure 7.9. This demonstrates that the defects associated with loss in this sample – while being concentrated at the island edges – are still confined to points in x and y .

Since the defect in this sample manifests at positive bias, it is likely acceptor-like, which explains the unintentional p-doping of this sample. (p-type doping is much more commonly found in organic systems[26]. This method, however, presents an opportunity to identify dopant “type”, meaning that if an n-type dopant were found, it could be straightforwardly imaged, opening up the possibility to intentionally increase the density of the desired dopants.)

There is a decrease in F_d at V_g larger than the peak position (see Figure 7.8), which was also the case for Si/SiO₂ and MoSe₂. However, in the pentacene sample, F_d also increased significantly at negative bias (i.e. in the accumulation corresponding to the majority carrier, in this case holes), which was not the case for the other two samples. In pentacene, hole mobility is found to sharply decrease as grain size decreases[27]. This has been attributed to an appreciable decrease in the local density of valence band (highest occupied molecular orbital) states in the pentacene molecules along a stacking fault[24]. This interpretation would also explain why the island edges have the largest F_d at negative bias in Figure 7.7.

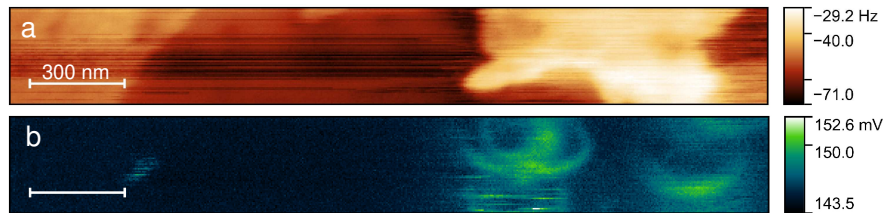


Figure 7.9: **Pentacene rings.** Multipass image of part of the thin and thick islands shown in Figure 7.2. For both Δf (a) and F_d (b), $V_g = +7$ V.

7.4 Random telegraph noise

In Chapter 6, RTN was demonstrated in the SiO_2 system, and was found to be spatially heterogeneous and bias-dependent. The MoSe_2 and pentacene systems also exhibited RTN – these findings are discussed below.

7.4.1 MoSe_2 noise

Figure 7.10 shows a multipass image of part of the top layer of the MoSe_2 island. This image is strikingly similar to the noise measured for the Si/SiO_2 sample in Figure 6.1 in Chapter 6: There is two-state noise in F_d which is localized at particular sites; the noise appears to correspond with a peak in the $F_d(V_g)$ spectrum (i.e. a ring); and there are dramatically variable fluctuation timescales. The scan speed for Figure 7.10 was 60 nm/s , meaning that the noise timescales seen here range from $\text{mHz} - \text{Hz}$. Figure 7.11 shows a bias spectrum measured above a noisy MoSe_2 ring center. There is a significant increase in noise in the Δf and F_d channels between $V_g = -2 : -0.5 \text{ V}$.

While there are many similarities between the RTN in the SiO_2 and MoSe_2 samples, this does not necessarily mean that their noise mechanism is the same. In Si/SiO_2 , hydrogen activation/passivation of an IT was a proposed noise mechanism, but it is not necessarily the case that noise in MoSe_2 is also related to hydrogen. In Section 7.3, it was proposed that interstitial atoms could be the defects measured here; in that case, this measured noise corresponds to the same kind of trap. Much like the SiO_2 , it seems unlikely that this noise arises simply due to tunneling between the MoSe_2 charge density and the trap state (that is,

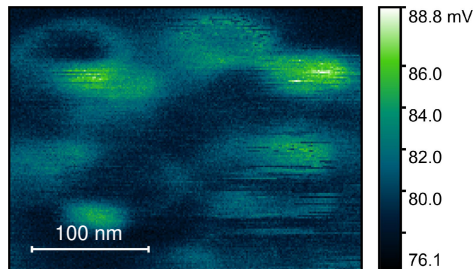


Figure 7.10: **Noise rings in MoSe_2 .** Multipass image showing F_d for $V_g = -1.5 \text{ V}$.

presuming the trap is an interstitial defect, as described in the previous section) because the tunneling barrier would be too small to give the long RTS timescales observed here. These slow timescales could correspond to slow conformational changes which are associated with the charge state of the trap, or with tunneling between the trap and the Si/SiO₂ substrate[28] (the top layer height with respect to the substrate is ~ 10 nm, recall Figure 7.1).

Figure 7.11 shows that for MoSe₂, fluctuations in Δf are correlated with fluctuations in F_d – that is, there is a state “0” which corresponds to one V_g and less loss, and there is a state “1” which corresponds to a different V_g and more loss. This agrees with other experimental observations in TMDC field-effect transistors that defects suppress mobility and give rise to 1/f noise[28, 29]. In monolayer TMDCs grown by chemical vapor deposition, there is an apparent increase in 1/f noise at grain boundaries[28, 30]; that does not appear to be the case here. (There may be a higher concentration of defects at grain boundaries, which would be consistent with the findings presented here, but Figure 7.5 shows that scattering at the island edges themselves is not a significant effect.)

The fluctuations in Δf seen here can be related to changes in the dopant density of the capacitor[5]. The charge states for a donor-like trap are neutral or positive, so when the trap

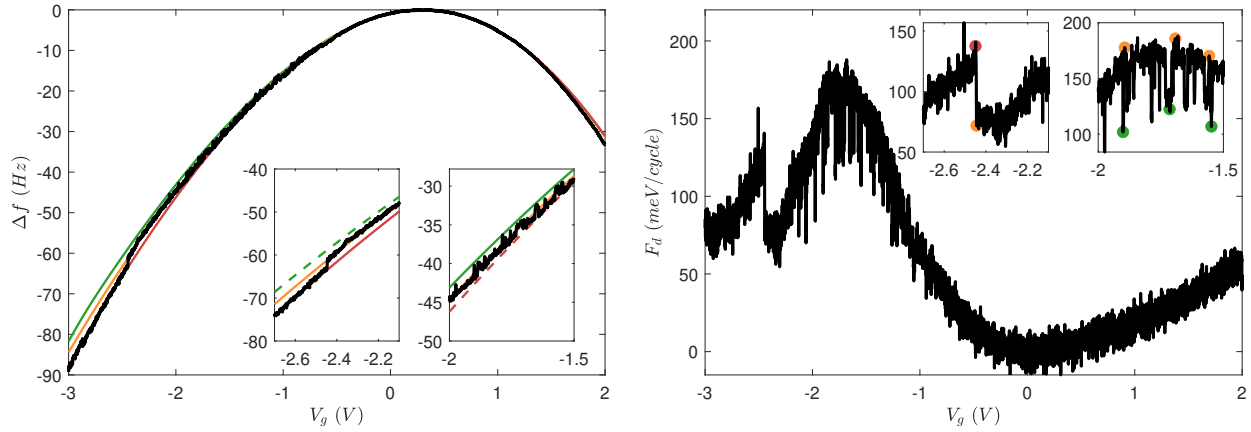


Figure 7.11: **Correlated fluctuations in MoSe₂.** Bias spectra measured (black) on the top layer of the MoSe₂ island in Figure 7.1. Three MIS $\Delta f(V_g)$ models (colour) are shown, which differ only in their n-type dopant concentration. They are: Low (green, $7.59 \times 10^{17} / \text{cm}^3$), medium (orange, $8.71 \times 10^{17} / \text{cm}^3$), and high (red, $9.91 \times 10^{17} / \text{cm}^3$). The insets show regions of interest (with the same x-axes for the the Δf and F_d insets). Coloured dots highlight that higher dopant densities in the Δf spectrum correspond to increased loss measured in the F_d spectrum.

accepts a hole from the valence band at negative bias, this corresponds to an effective increase in the n-type dopant concentration of the capacitor. Figure 7.11 shows a bias spectrum with three MIS models which differ only in their n-type dopant concentration. The change in dopant concentration ($\sim 1.12 \times 10^{17} / \text{cm}^3$ from low to medium and $\sim 1.2 \times 10^{17} / \text{cm}^3$ from medium to high) corresponds to approximately one electron within the 1000 nm^3 probing volume of the fm-AFM tip (the sampling area is approximately the ring area, where the ring radius $\sim 30 \text{ nm}$, and the depth is the sample thickness, $\sim 10 \text{ nm}$).

As the effective dopant density increases, there is a corresponding increase in F_d (see the coloured dots in Figure 7.11). This is consistent with an increase in Coulomb scattering, which is expected when the trap becomes charged. At large negative bias, there was very rarely another jump; in Figure 7.11, this corresponds to the system switching between a singly positive and doubly positive charge state (i.e. the trap accepts two holes from the valence band); and there is a another increase in F_d according to the increased Coulomb potential which gives rise to scattering.

7.4.2 Pentacene noise

Figure 7.12 shows RTSs measured above a pentacene island at variable V_g . The RTS rates are bias dependent, and maximized at a bias corresponding to a peak in F_d , as was the case for both Si/SiO₂ and MoSe₂. For the RTSs shown in Figure 7.12, Δf fluctuations were not correlated with fluctuations in F_d above measurement noise, which indicates that if there are correlated mobility fluctuations, they are smaller than measurement noise.

There was, however, a significant increase in the noise amplitude in both Δf and F_d at large positive bias. (These time traces are not shown, but the noise amplitude increase can be seen by comparing the uncertainties in the points in Figure 7.12 at large positive bias to those at large negative bias). The character of the noise at large negative bias was consistent with drift, but at large positive bias, there appeared to be RTN which was presumably due to many RTSs (like those shown in Figure 7.12) being sampled at once. Individual RTSs did not exhibit fluctuations in F_d , as mentioned previously, but the increase in the standard deviation of F_d at large positive bias suggests that this RTN is related to mobility, and that – when sampled in large enough numbers – the effect that these tiny mobility fluctuations

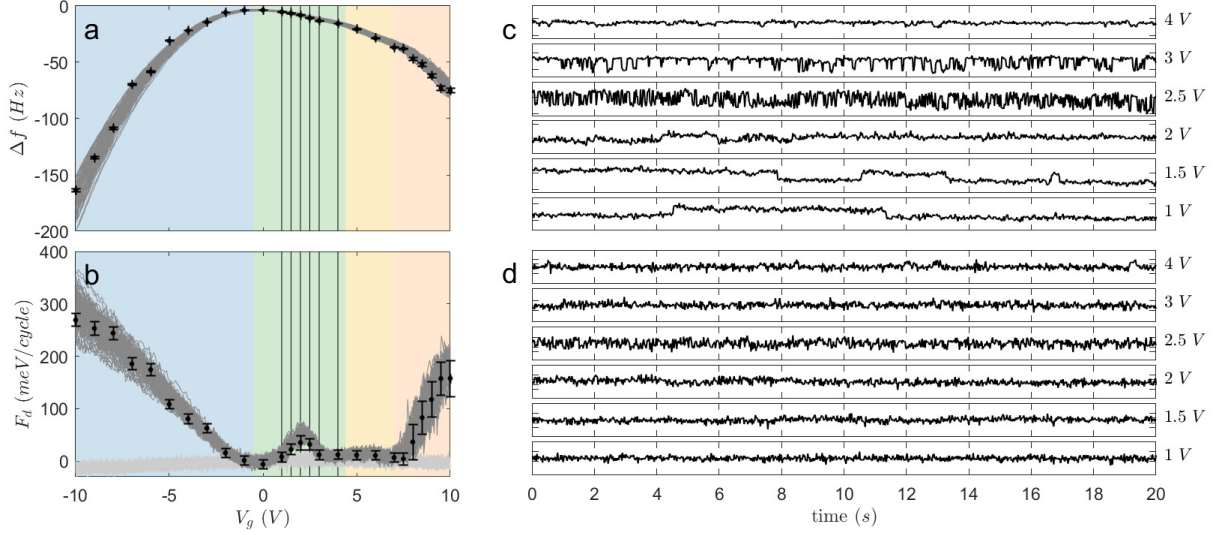


Figure 7.12: **Pentacene noise.** Bias spectra (a,b) and measured time traces (c,d) at variable V_g (indicated). The points and error bars in (a,b) correspond to the mean and standard deviation of 20 s time traces measured at each V_g . The biases corresponding to the time traces in (c,d) are indicated by vertical lines.

have on the total mobility becomes appreciable.

A possible noise mechanism is a slow modification of the structural organization of the pentacene island edges. Slow noise associated with adding and removing molecules from island edges on the order of $mHz - Hz$ has been measured for C60[8]. It has also been observed that a decrease in grain boundary size corresponds to an increase in $1/f$ noise in pentacene MOSFETs, and the noise was attributed to disorder-originated traps – specifically, acceptor-like traps at grain boundaries[31], which is precisely the kind of trap measured here. Nanoscale dewetting and growth of the islands of this sample was measured over hours and days (as in [32]), which is further evidence that the island edges are dynamic – the molecular gas phase of pentacene on the substrate leads to the islands changing shape over time.

Figure 7.13 shows transient noise of the thick pentacene island. Such transients were commonly observed for this sample at both negative and positive bias, though at positive bias, the transient state manifested as a series of jumps without RTN, whereas at negative bias (as in Figure 7.13), there are a series of jumps which have associated varying RTS rates. In fact, this is what made it particularly challenging to develop a more statistical understanding of RTN in this system: The noise characteristics changed often – about every

~ 100 s or less (as compared to the SiO_2 results shown in Chapter 6, where noise time traces were measured for > 500 s with overall constant RTS rates and amplitudes, even though the tip drifted by ~ 1 nm) which made it effectively impossible to sample the noise for long periods of time.

Similar transient RTN has also been measured in Si/SiO_2 [33] (though without such state-dependent rates) and was associated with deep level filling under strong inversion. In this case, if the structural origin of this noise proposed above is correct, than this transiency can be attributed to the island slowly changing shape over time (as more pentacene molecules are added to the edges, for example, or as the angle of the edge molecules with respect to the surface normal changes). An enlightening experiment would be to measure the change in the shape of a pentacene island as a function of time under strong inversion. Given the results in Figure 7.13, it seems likely that the island shape would not change uniformly, but rather would pass through a series of stages, where in some stages the edge molecules move a lot (i.e. the fast RTN stages in Figure 7.13) and in others the island conformation appears approximately static (i.e. slow RTN). In this experiment, no variability in F_d as a function of time was observed; but a correlation is expected, so perhaps with higher sensitivity in this channel, correlations between the island structure and loss could be measured directly.

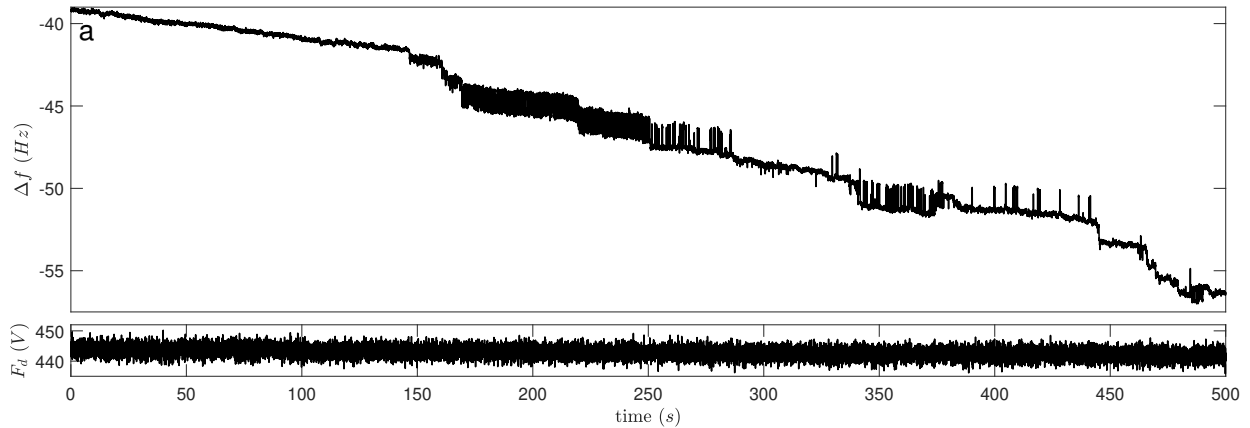


Figure 7.13: **Transient noise in pentacene.** Δf (a) and F_d (b) measured for 500 s with $V_g = +10$ V above the thick pentacene island in Figure 7.2.

References

- [1] McEuen, P. L. “Single-wall carbon nanotubes”. *Physics World* 13.6 (**2000**). DOI: 10.1088/2058-7058/13/6/26.
- [2] Arca, F., Mendez, J., Ortiz, M., and Ariza, M. “Strain-tuning of transport gaps and semiconductor-to-conductor phase transition in twinned graphene”. *Acta Materialia* 234 (**2022**). DOI: 10.1016/j.actamat.2022.117987.
- [3] Eftekhari, A. “Molybdenum diselenide (MoSe_2) for energy storage, catalysis, and optoelectronics”. *Applied Materials Today* 8 (**2017**). DOI: 10.1016/j.apmt.2017.01.006.
- [4] Cowie, M., Plougmann, R., Benkirane, Y., Schué, L., Schumacher, Z., and Grütter, P. “How high is a MoSe_2 monolayer?” *Nanotechnology* 33.12 (**2021**). DOI: 10.1088/1361-6528/ac40bd.
- [5] Cowie, M., Plougmann, R., Schumacher, Z., and Grütter, P. “Single-dopant band bending fluctuations in MoSe_2 measured with electrostatic force microscopy”. *Physical Review Materials* 6.10 (**2022**). DOI: 10.1103/physrevmaterials.6.104002.
- [6] Neff, J. L., Milde, P., León, C. P., Kundrat, M. D., Eng, L. M., Jacob, C. R., and Hoffmann-Vogel, R. “Epitaxial growth of pentacene on alkali halide surfaces studied by Kelvin probe force microscopy”. *ACS Nano* 8.4 (**2014**). DOI: 10.1021/nn404257v.
- [7] Seo, S. and Evans, P. G. “Molecular structure of extended defects in monolayer-scale pentacene thin films”. *Journal of Applied Physics* 106.10 (**2009**). DOI: 10.1063/1.3262618.
- [8] Burke, S. A., Mativetsky, J. M., Hoffmann, R., and Grütter, P. “Nucleation and sub-monolayer growth of C_{60} on KBr ”. *Physical Review Letters* 94.9 (**2005**). DOI: 10.1103/physrevlett.94.096102.
- [9] Sze, S. M. and Ng, K. K. *Physics of semiconductor devices*. 3rd ed. John Wiley & Sons, Inc, **2007**. Chap. 1,4. ISBN: 9780470068328. DOI: 10.1002/0470068329.

-
- [10] Vliet, C. M. van and Vasilopoulos, P. “Electrical and optical bandgap narrowing due to heavy doping in silicon”. *physica status solidi (a)* 94.2 (**1986**). DOI: 10.1002/pssa.2210940226.
- [11] King, P. D. C., Veal, T. D., McConville, C. F., Zúñiga-Pérez, J., Muñoz-Sanjosé, V., Hopkinson, M., Rienks, E. D. L., Jensen, M. F., and Hofmann, P. “Surface band-gap narrowing in quantized electron accumulation layers”. *Physical Review Letters* 104.25 (**2010**). DOI: 10.1103/physrevlett.104.256803.
- [12] Kim, S. et al. “Interstitial Mo-assisted photovoltaic effect in multilayer MoSe₂ phototransistors”. *Advanced Materials* 30.12 (**2018**). DOI: 10.1002/adma.201705542.
- [13] Tongay, S., Zhou, J., Ataca, C., Lo, K., Matthews, T. S., Li, J., Grossman, J. C., and Wu, J. “Thermally driven crossover from indirect toward direct bandgap in 2D semiconductors: MoSe₂ versus MoS₂”. *Nano Letters* 12.11 (**2012**). DOI: 10.1021/nl302584w.
- [14] Lee, H., Ahn, J., Im, S., Kim, J., and Choi, W. “High-responsivity multilayer MoSe₂ phototransistors with fast response time”. *Scientific Reports* 8.1 (**2018**). DOI: 10.1038/s41598-018-29942-1.
- [15] Kang, Y., Jeon, D., and Kim, T. “Local mapping of the thickness-dependent dielectric constant of MoS₂”. *The Journal of Physical Chemistry C* 125.6 (**2021**). DOI: 10.1021/acs.jpcc.0c11198.
- [16] Zhang, Q., Zhang, S., Sperling, B. A., and Nguyen, N. V. “Band offset and electron affinity of monolayer MoSe₂ by internal photoemission”. *Journal of Electronic Materials* 48.10 (**2019**). DOI: 10.1007/s11664-019-07396-z.
- [17] Han, W., Yoshida, H., Ueno, N., and Kera, S. “Electron affinity of pentacene thin film studied by radiation-damage free inverse photoemission spectroscopy”. *Applied Physics Letters* 103.12 (**2013**). DOI: 10.1063/1.4821445.
- [18] Ha, S. D., Qi, Y., and Kahn, A. “Relative permittivity and Hubbard U of pentacene extracted from scanning tunneling microscopy studies of p-doped films”. *Chemical Physics Letters* 495.4-6 (**2010**). DOI: 10.1016/j.cplett.2010.06.085.

-
- [19] *Nanosensors PPP-NCHPt; Electrical, Tapping Mode AFM Probe*. <https://www.nanoandmore.com/AFM-Probe-PPP-NCHPt>. Accessed: 2023-07-07.
 - [20] Yu, Z., Ong, Z.-Y., Li, S., Xu, J.-B., Zhang, G., Zhang, Y.-W., Shi, Y., and Wang, X. “Analyzing the carrier mobility in transition-metal dichalcogenide MoS₂ field-effect transistors”. *Advanced Functional Materials* 27.19 (2017). DOI: 10.1002/adfm.201604093.
 - [21] Zhang, C., Johnson, A., Hsu, C.-L., Li, L.-J., and Shih, C.-K. “Direct imaging of band profile in single layer MoS₂ on graphite: Quasiparticle energy gap, metallic edge states, and edge band bending”. *Nano Letters* 14.5 (2014). DOI: 10.1021/nl501133c.
 - [22] Sueyoshi, T., Fukagawa, H., Ono, M., Kera, S., and Ueno, N. “Low-density band-gap states in pentacene thin films probed with ultrahigh-sensitivity ultraviolet photoelectron spectroscopy”. *Applied Physics Letters* 95.18 (2009). DOI: 10.1063/1.3258351.
 - [23] Jurchescu, O. D., Baas, J., and Palstra, T. T. M. “Effect of impurities on the mobility of single crystal pentacene”. *Applied Physics Letters* 84.16 (2004). DOI: 10.1063/1.1704874.
 - [24] Watanabe, S. and Nakayama, T. “Electronic structures and impurity segregation around extended defects in pentacene films: first-principles study”. *Japanese Journal of Applied Physics* 60.SB (2021). DOI: 10.35848/1347-4065/abdf71.
 - [25] Labuda, A., Miyahara, Y., Cockins, L., and Grütter, P. H. “Decoupling conservative and dissipative forces in frequency modulation atomic force microscopy”. *Physical Review B* 84.12 (2011). DOI: 10.1103/physrevb.84.125433.
 - [26] Anthony, J. E., Facchetti, A., Heeney, M., Marder, S. R., and Zhan, X. “n-Type organic semiconductors in organic electronics”. *Advanced Materials* 22.34 (2010). DOI: 10.1002/adma.200903628.
 - [27] Günther, A. A., Widmer, J., Kasemann, D., and Leo, K. “Hole mobility in thermally evaporated pentacene: Morphological and directional dependence”. *Applied Physics Letters* 106.23 (2015). DOI: 10.1063/1.4922422.

- [28] Hsieh, K., Kochat, V., Zhang, X., Gong, Y., Tiwary, C. S., Ajayan, P. M., and Ghosh, A. “Effect of carrier localization on electrical transport and noise at individual grain boundaries in monolayer MoS₂”. *Nano Letters* 17.9 (**2017**). DOI: 10.1021/acs.nanolett.7b02099.
- [29] Karnatak, P., Paul, T., Islam, S., and Ghosh, A. “1/f noise in van der Waals materials and hybrids”. *Advances in Physics: X* 2.2 (**2017**). DOI: 10.1080/23746149.2017.1314192.
- [30] Kim, J.-K., Song, Y., Kim, T.-Y., Cho, K., Pak, J., Choi, B. Y., Shin, J., Chung, S., and Lee, T. “Analysis of noise generation and electric conduction at grain boundaries in CVD-grown MoS₂ field effect transistors”. *Nanotechnology* 28.47 (**2017**). DOI: 10.1088/1361-6528/aa9236.
- [31] Kang, H., Jagannathan, L., and Subramanian, V. “Measurement, analysis, and modeling of 1/f noise in pentacene thin film transistors”. *Applied Physics Letters* 99.6 (**2011**). DOI: 10.1063/1.3622651.
- [32] Topple, J. M., Burke, S. A., Ji, W., Fostner, S., Tekiel, A., and Grütter, P. “Tailoring the morphology and dewetting of an organic thin film”. *The Journal of Physical Chemistry C* 115.1 (**2010**). DOI: 10.1021/jp107644u.
- [33] Kirton, M. J., Uren, M. J., Collins, S., Schulz, M., Karmann, A., and Scheffer, K. “Individual defects at the Si:SiO₂ interface”. *Semiconductor Science and Technology* 4.12 (**1989**). DOI: 10.1088/0268-1242/4/12/013.

Chapter 8

Conclusions

Energy dissipation in materials can occur by a variety of mechanisms. For the frequencies explored here, Ohmic losses dominate, which means that energy is dissipated by scattering of carriers in the depletion width or at the semiconductor-insulator interface as the cantilever oscillates. If the number of carriers moving in the depletion width increases (which is bias-dependent and dopant density-dependent, as shown in Chapter 4), energy dissipation increases. Additionally, if there are local increases in scattering (for example, at a defect site, as shown in Chapter 5), dissipation also increases. Finally, if there are fluctuations in the defect state, there can be corresponding fluctuations (noise) in scattering (as shown in Chapter 6).

For all of the methodology and results presented in this thesis, a bias is applied to a semiconductor surface via an oscillating top gate. The relevance, therefore, of this system for semiconductor devices in general – where geometries tend to be fixed – may not be immediately obvious; however, the effect of the oscillating cantilever is entirely identical to applying an AC gate bias. The fm-AFM MIS force is related to Q_S , the charge of the capacitor, and if the capacitor charging and discharging is instantaneous, this means that the tip-sample force $F_{ts}(t)$ and the tip position $z_{ins}(t)$ are in-phase. In terms of a capacitive circuit, this means that the AC voltage (z_{ins}) and current ($\partial Q_S/\partial t$) are 90 degrees out of phase, with the current leading the voltage. If energy is dissipated due to resistive losses in

the sample as the cantilever oscillates, this is equivalently like adding a resistor in series to the circuit: This “equivalent series resistance” (ESR) introduces a reduction of the phase angle between $z_{ins}(t)$ and $\partial Q_S/\partial t$ (i.e. $F_{ts}(t)$ slightly lags $z_{ins}(t)$, so there is an out-of-phase tip-sample force contribution).

In this “circuit language”, MIS fm-AFM is entirely analogous to bias-dependent admittance spectroscopy[1–3], in which the capacitance and conductance of MIS devices are measured upon application of an AC gate bias. The conductance signal, which is essentially a measurement of the ESR, is associated with energy loss, and therefore is comparable to F_d measurements in MIS fm-AFM. Admittance spectroscopy is used to characterize interfacial states which, as seen in Chapter 5, lead to significant losses which manifest in a bias-dependent peak in the conductance (F_d) signal. In addition to this bias dependence[4–6], other dependencies can be explored, principally the temperature dependence[3, 7], frequency dependence[3, 4, 6], and illumination-dependence[8, 9]. These experiments can be done to measure various characteristics of interfacial traps, such as the trap energy, as was done in Chapter 5, characteristic capture and emission rates, and trap capture cross sections.

The results presented here show the first bias-dependent admittance spectroscopy measurements of individual defects at a semiconductor surface; exploring the other parameters listed above (temperature, frequency, and illumination) with fm-AFM are each interesting future directions for this work. Each experiment would lead to sample-specific insights regarding the role that defects play at semiconductor surfaces.

Measuring the temperature dependence is a relatively straightforward next experiment. In particular, measuring the temperature dependence of the various random telegraph fluctuations shown in this thesis would allow the activation energies and temperature-dependent RTS timescales to be determined, which would provide further insights regarding the noise mechanisms. If the noise mechanisms are known, they may be able to be suppressed. For example, if the hydrogen activation/passivation mechanism proposed in Chapter 6 is a dominant noise source at the Si/SiO₂ interface, this could have particular consequences for atomic-scale hydrogen lithography: In hydrogen lithography, surfaces are exposed to significant hydrogen concentrations, so they may have high interstitial hydrogen concentrations and be particularly susceptible to noise if they are not appropriately annealed post-patterning.

Measuring the frequency dependence poses more challenges for MIS fm-AFM. Specifically, the AC frequency is fixed to the natural cantilever resonance frequency. Exploring frequency space, therefore, is not straightforward: It can be accomplished by exchanging cantilevers, potentially exciting higher cantilever harmonics (though this approach will have limited sensitivity because the higher modes are stiffer); or by applying a bias at a variable AC frequency and measuring the phase difference between the applied signal and the measured one using a lock-in amplifier. The latter approach does not make use of the natural signal enhancement by the cantilever transfer function, so may also be limited in sensitivity, though there may be workarounds such as changing the cantilever temperature (which changes its effective Q [10]) or heterodyning multiple AC frequencies[11]. It should be noted, however, that if the cantilever is still oscillating, then the combined effect of applying two or more frequencies (the cantilever oscillation plus an AC bias) will need to be treated with care. In particular, the derivation of Chapter 4, which assumes a single interfacial polarization relaxation time constant, would need to be revisited.

In the future, if it were possible to hold the cantilever static and vary the AC bias from radio to terahertz frequencies, it would be fascinating to observe the spatial heterogeneity of every loss mechanism in the material. That is, in the current setup with a 300 kHz cantilever, charge reorganization timescales can only be measured into the ns regime, due to the finite sensitivity of the PLL as discussed at the end of Chapter 2. For example, at high (GHz) frequencies, loss peaks and rings observed in Chapter 5 are expected to decrease, because the timescales for trap state relaxations were found to be $\sim 1 - 100\text{ ns}$ ($1 - 0.01\text{ GHz}$). Perhaps at high enough excitation frequencies, phonons could be excited directly, allowing for a detailed, spatially resolved understanding of where phonon excitations take place. In this case, energy dissipation would be due to resonant excitation rather than Ohmic loss. Perhaps STM is a potential avenue to explore some ranges of this frequency dependence, but how an STM signal should be interpreted in the context of loss is not obvious.

Preliminary experiments exploring the illumination dependence of loss peaks and noise were performed for pentacene. It was found that the surface photovoltage (SPV) had a peak corresponding to the loss peak in Figure 7.3f. The SPV is a measure of the difference in the surface potential when the light (in this case, a 780 nm beam) is on versus off. (This

measurement was performed by optically chopping the beam at 109 Hz and using a lock-in amplifier to demodulate Δf at the chopper frequency, enabling a simultaneous measurement of the “on” and “off” signals[12]). This preliminary finding indicates that there is a large change in the surface charge density for V_g between the trap state crossing points, perhaps signifying that there is a significant change in the population of trap states due to photo-excited carriers. A future experiment should simultaneously demodulate the F_d channel at the chopper frequency, to observe whether illumination leads to a corresponding increase or decrease in defect-mediated loss.

As an aside, it should also be noted that the surface SPV peak which was observed in pentacene is expected to be nonlinear with light intensity, due to the nonlinearity of band bending. It is therefore expected that in similar systems, variations in the illumination intensity will lead to nonlinearities in the Δf response. Consider, for example, an experiment where a fs pulsed beam is split into two “legs” which are subsequently collinearly aligned and delayed with respect to one another in time. The resulting beam will have the form of a field autocorrelation function, with fringes corresponding to sub- fs delay times centered around zero delay. When directed onto the sample in the fm-AFM tip-sample junction, Δf will asymmetrically follow the intensity variations of the beam, which could resemble an interferometric autocorrelation function and therefore be mistaken for ultrafast (fs) temporal dynamics in the sample. In such an experiment scheme, if nonlinear optical sample responses are to be measured by fm-AFM, the excitation should be sub-band gap to eliminate photo-excited carriers, the F_d channel should be monitored and found to be independent of illumination, and there should be very little bias dependence in the “autocorrelation” response (unless it can be definitively attributed to the Pockels effect, or another similar bias-dependent optically induced nonlinearity). These are the kinds of nuances that need to be considered when measuring semiconducting samples with fm-AFM.

Typically in fm-AFM experiments, semiconductor surfaces are approximated as metallic, in that the derivation of the frequency shift equation assumes instantaneous charging of a capacitor with an infinite density of states. In some fm-AFM and STM experiments, the tip-sample junction is modelled as a MIS capacitor[13–20], though in the case of fm-AFM the effect of the oscillating cantilever tends to be neglected. On the other hand, there is a

recognition that the F_d channel is related to energy dissipation in the sample[21–23], though without the associated understanding that comes from describing the interface as a MIS capacitor.

This work showed that semiconductors have nanoscale heterogeneities in their dielectric dispersion and noise. This means, for example, that two atoms spaced less than 1 *nm* apart in a semiconducting lattice could have significantly different responses to a time-varying field, even if they were positioned near a gate mere nanometers away. Within those few nanometers, the two atoms could each also be susceptible to different random telegraph noise inherent to spatially localized defects in the substrate. These considerations are of practical concern for the continued development of nano- and atomic-scale semiconductor devices, quantum sensors, and quantum computers.

References

- [1] Nicollian, E. H. and Goetzberger, A. “The Si-SiO₂ interface - Electrical properties as determined by the metal-insulator-silicon conductance technique”. *Bell System Technical Journal* 46.6 (**1967**). DOI: 10.1002/j.1538-7305.1967.tb01727.x.
- [2] Engström, O. *The MOS system*. Cambridge University Press, **2014**. ISBN: 9780511794490. DOI: 10.1017/CBO9780511794490.
- [3] Piscator, J., Raeissi, B., and Engström, O. “Multiparameter admittance spectroscopy for metal-oxide-semiconductor systems”. *Journal of Applied Physics* 106.5 (**2009**). DOI: 10.1063/1.3213384.
- [4] Raeissi, B., Piscator, J., and Engstrom, O. “Multiparameter admittance spectroscopy as a diagnostic tool for interface states at oxide/semiconductor interfaces”. *IEEE Transactions on Electron Devices* 57.7 (**2010**). DOI: 10.1109/ted.2010.2049064.
- [5] Raeissi, B. et al. “High-k-oxide/silicon interfaces characterized by capacitance frequency spectroscopy”. *Solid-State Electronics* 52.9 (**2008**). DOI: 10.1016/j.sse.2008.04.005.

-
- [6] Li, J. V. and Wong, M. H. “Investigation of electric field effect on defects in GaAsN by admittance spectroscopy”. *Thin Solid Films* 758 (2022). DOI: 10.1016/j.tsf.2022.139422.
- [7] Li, J. V. “Defect characterization using raw admittance spectroscopy”. *The Journal of Physical Chemistry C* 125.5 (2021). DOI: 10.1021/acs.jpcc.0c10853.
- [8] Poon, T. C. and Card, H. C. “Admittance measurements of Si-SiO₂ interface states under optical illumination”. *Journal of Applied Physics* 51.11 (1980). DOI: 10.1063/1.327552.
- [9] Blank, A., Suhareva, N., and Zuev, N. “Admittance spectra of silicon photocells: From dark mode to weak illuminate mode”. *Sensors and Actuators A: Physical* 334 (2022). DOI: 10.1016/j.sna.2021.113336.
- [10] Iervolino, E., Riccio, M., Santagata, F., Wei, J., Herwaarden, A. van, Irace, A., Breglio, G., and Sarro, P. “Resonance frequency of locally heated cantilever beams”. *Sensors and Actuators A: Physical* 190 (2013). DOI: 10.1016/j.sna.2012.10.008.
- [11] Fukuzawa, R. and Takahashi, T. “Development of dual bias modulation electrostatic force microscopy for variable frequency measurements of capacitance”. *Review of Scientific Instruments* 91.2 (2020). DOI: 10.1063/1.5127219.
- [12] Schumacher, Z., Miyahara, Y., Spielhofer, A., and Grutter, P. “Measurement of surface photovoltage by atomic force microscopy under pulsed illumination”. *Physical Review Applied* 5.4 (2016). DOI: 10.1103/physrevapplied.5.044018.
- [13] Cowie, M., Plougmann, R., Schumacher, Z., and Grütter, P. “Single-dopant band bending fluctuations in MoSe₂ measured with electrostatic force microscopy”. *Physical Review Materials* 6.10 (2022). DOI: 10.1103/physrevmaterials.6.104002.
- [14] Feenstra, R. M., Dong, Y., Semtsiv, M. P., and Masselink, W. T. “Influence of tip-induced band bending on tunnelling spectra of semiconductor surfaces”. *Nanotechnology* 18.4 (2006). DOI: 10.1088/0957-4484/18/4/044015.

- [15] Johnson, J. P., Zheng, N., and Williams, C. C. “Atomic scale imaging and spectroscopy of individual electron trap states using force detected dynamic tunnelling”. *Nanotechnology* 20.5 (**2009**). DOI: 10.1088/0957-4484/20/5/055701.
- [16] Johnson, J. P., Winslow, D. W., and Williams, C. C. “Measurement of depth and energy of buried trap states in dielectric films by single electron tunneling force spectroscopy”. *Applied Physics Letters* 98.5 (**2011**). DOI: 10.1063/1.3549150.
- [17] Bernardo, I. D., Blyth, J., Watson, L., Xing, K., Chen, Y.-H., Chen, S.-Y., Edmonds, M. T., and Fuhrer, M. S. “Defects, band bending and ionization rings in MoS₂”. *Journal of Physics: Condensed Matter* 34.17 (**2022**). DOI: 10.1088/1361-648x/ac4f1d.
- [18] Wang, R. and Williams, C. C. “Dynamic tunneling force microscopy for characterizing electronic trap states in non-conductive surfaces”. *Review of Scientific Instruments* 86.9 (**2015**). DOI: 10.1063/1.4931065.
- [19] Winslow, D. and Williams, C. “Local density of trap states in SiO₂ and Si₃N₄ films studied by single electron tunneling force spectroscopy”. *Journal of Applied Physics* 110.11 (**2011**). DOI: 10.1063/1.3662145.
- [20] Sadewasser, S. and Glatzel, T., eds. *Kelvin probe force microscopy: From single charge detection to device characterization*. Springer International Publishing, **2018**. ISBN: 9783319756868.
- [21] Denk, W. and Pohl, D. W. “Local electrical dissipation imaged by scanning force microscopy”. *Applied Physics Letters* 59.17 (**1991**). DOI: 10.1063/1.106088.
- [22] Suzuki, K., Kobayashi, K., Labuda, A., Matsushige, K., and Yamada, H. “Accurate formula for dissipative interaction in frequency modulation atomic force microscopy”. *Applied Physics Letters* 105.23 (**2014**). DOI: 10.1063/1.4903484.
- [23] Arai, T., Kura, D., Inamura, R., and Tomitori, M. “Resistivity change in Joule heat energy dissipation detected by noncontact atomic force microscopy using a silicon tip terminated with/without atomic hydrogen”. *Japanese Journal of Applied Physics* 57.8S1 (**2018**). DOI: 10.7567/jjap.57.08nb04.

Appendix A

Experimental details

A.1 UHV System

The microscope used for the fm-AFM measurements in this work is a room temperature commercial and customized JEOL JSPM-4500A ultra-high vacuum (UHV) system, shown in Figure A.1. It contains three connected chambers: A loadlock, for moving tips and samples in and out of vacuum; a preparation chamber, where various sample treatments, (such as salt cleaving, annealing, and molecular evaporation for fabrication of the pentacene

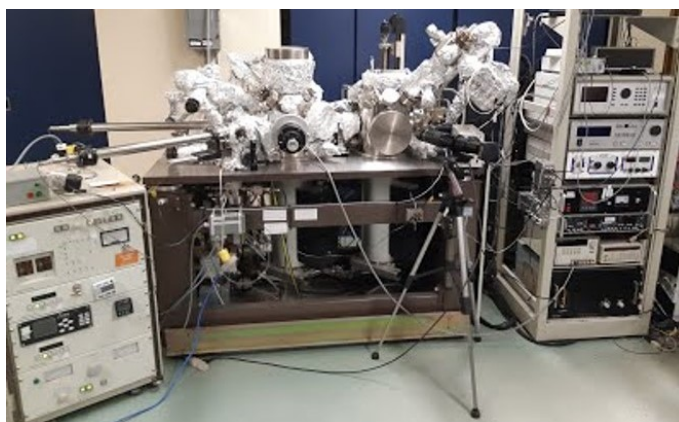


Figure A.1: **UHV system.** To the left of the UHV system is the vacuum control unit, for operating pneumatic valves, pumps, and gauges. The loadlock is a small chamber next to the vacuum control unit; it connects to the prep chamber, which in turn is connected to the measurement chamber. To the right of the UHV system is an electronics rack for controlling various UHV components (e.g. the molecular evaporator and quartz microbalance, annealing stage, the fm-AFM laser diode, and the sample motors).

sample) are carried out; and a measurement chamber, where all fm-AFM measurements take place. A Nanonis scanning probe microscopy (SPM) control system was used for all fm-AFM measurements. For optics experiments, Zurich Instruments UHV and HF2 lock-in amplifiers were also used.

The base pressure of the JEOL preparation and measurement chambers is between $1 \times 10^{-11} - 5 \times 10^{-10}$ mbar. This is achieved using a combination of pumps: A turbomolecular pump, where spinning blades transfer momentum to gas molecules in one direction through the pump; ion pumps, in which gaseous species are ionized and are accelerated via an the pump’s electric field into a trap; and a titanium sublimation pump (TSP), where titanium is sublimated onto the interior walls of the UHV chamber, where it chemically reacts with gaseous species. An important aspect of achieving these low pressures is the system “bake”, or heating of the system after every time it is vented. By heating the system, any species that coated the walls of the system while it was exposed to air (e.g. a thin water layer) evaporate quickly (over the course of days). If the system were not baked, these species would slowly desorb and need to be pumped over time, leading to higher pressures. Figure A.2 shows a typical “pumpdown curve” measured during a JEOL system bake.

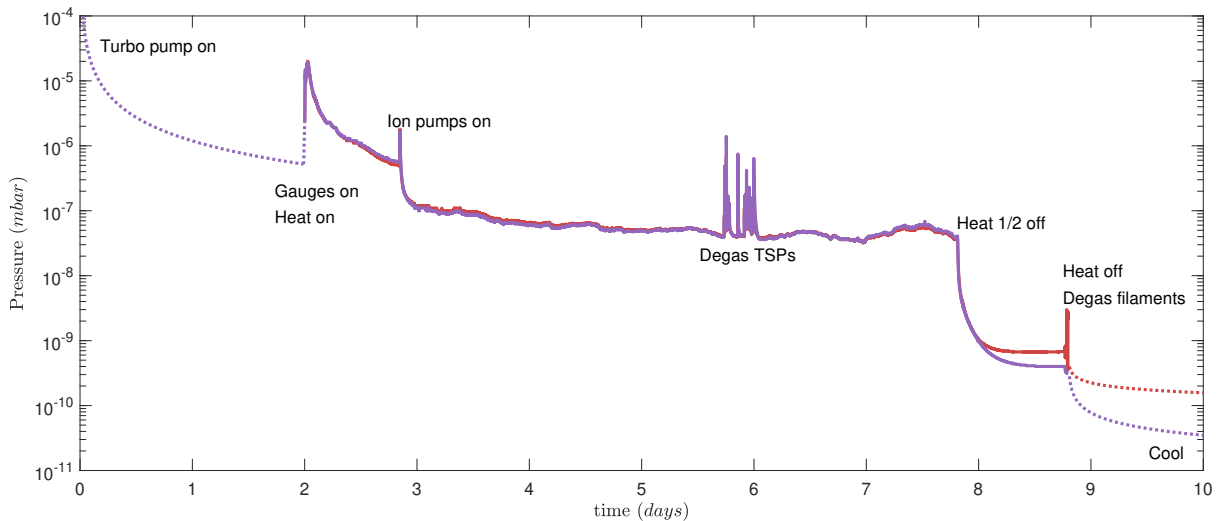


Figure A.2: UHV system pumpdown curve. Typical pumpdown curve for the measurement (red) and preparation (purple) chambers. The solid lines are data; the dashed lines demonstrate typical trends where data was not recorded. (The JEOL system measurement chamber has one o-ring seal. This flange has a higher permeability than the rest of the (copper) flanges in the UHV chambers, and consequently the base pressure of the measurement chamber is higher than that of the preparation chamber.)

A.2 KPFM imaging

Section 3.2.3 explains the KPFM operating principle, where electrostatic forces (F_{el}) are nulled by applying $V_g = V_{FB}$. If $V_g \neq V_{FB}$, the measured “topography” during a scan (i.e. the z channel) is a convolution of electrostatic and topographic information about the sample. Figure A.3, for example, shows two fm-AFM images of the MoSe₂ island introduced in Figure 7.1. Contrary to Figure 7.1, for the images shown here, the KPFM controller was turned off. Figure A.3a shows that at $V_g = 0$ V, it appears that the bright patches and edges are much taller than the top island layer; however, given Figure 7.1 it is clear that the true height difference between them is small and the large contrast in Figure A.3a is purely due to their different V_{FB} . Figure A.3b, similarly, shows another example of a potentially misleading “topography” image: At $V_g = -5$ V, the top layer appears to be ~ 15 nm tall as compared to the substrate, which is much taller than the ~ 9 nm which is known given Figure A.3. Furthermore, the first and second steps of the island (to the left of the top layer) even appear to be lower than the substrate. This is because at $V_g = -5$ V, F_{el} of the first and second layers happens to be smaller than F_{el} for the substrate, which makes the first and second layers have an apparent decrease in height.

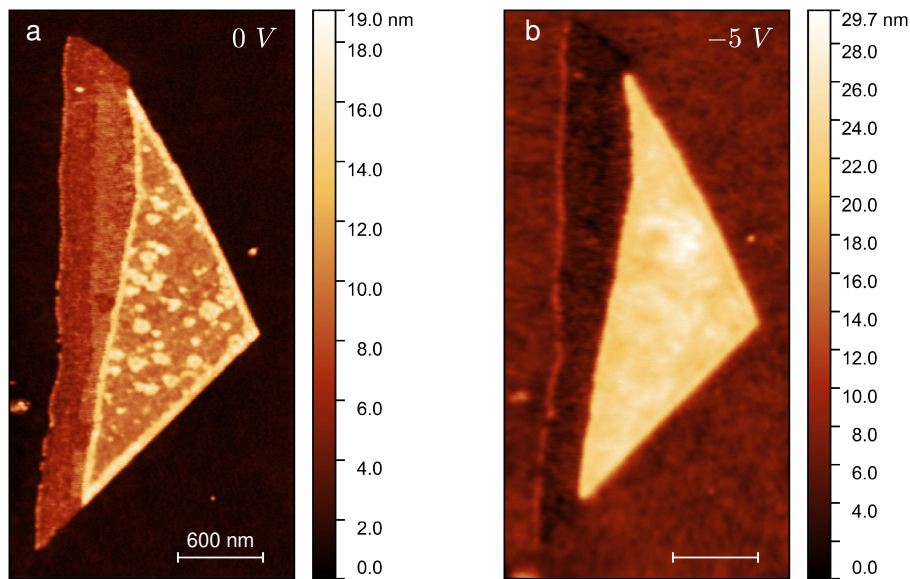


Figure A.3: **Bias-dependent “topography” of MoSe₂.** MoSe₂ island imaged without KPFM feedback (as compared to with KPFM feedback, Figure 7.1) at variable V_g (indicated).

A.3 Height measurements with AFM

A common way to measure sample “heights” (i.e. signal, often thickness, with respect to the substrate) in AFM is with a line profile. Figure A.4, for example, demonstrates this approach for an image of a pentacene island on KBr, for both the z and V_{FB} channels of a KPFM image. This method suffices for approximations of the height, but there is not an obvious way to measure the height uncertainty using this approach. For example, Figure A.4c could be reported as 0.8 nm taking the z of the substrate immediately before the island step and the z of the island immediately after the step; it would be 0.92 nm taking the minimum substrate z and the maximum island z ; and it would be 0.71 nm taking the maximum substrate z and the minimum island z . The uncertainty might be estimated by looking at variations in these values or fitting lines to each step and measuring standard deviations of the data to that line. This appendix demonstrates a simple and more robust approach for measuring sample heights with AFM.

The first thing to note is that, if sample thickness is the desired measurement, all forces which are not due to the sample height (principally electrostatics) should be nulled. This is most easily accomplished by KPFM, as explained in Appendix A.2.

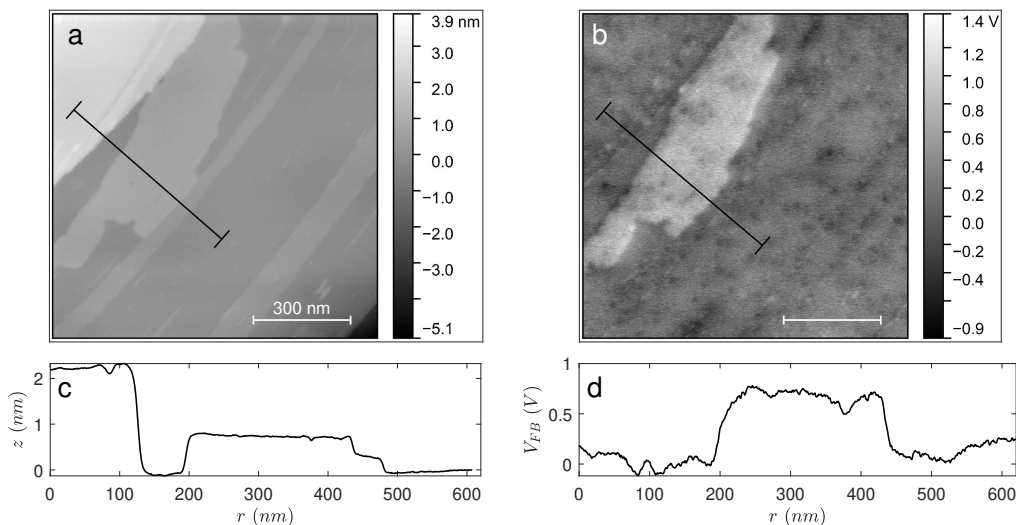


Figure A.4: **Line traces measured with AFM.** KPFM image of a pentacene island on KBr. (a) z channel, (b) V_{FB} channel, and (c-d) line traces corresponding to the black lines in (a-b).

Second, it is essential to make sure that the heating of the tip-sample junction is approximately constant during the experiment. If the cantilever is heated or cooled due to, for example, variations in room temperature or illumination, its natural (“free”) resonance frequency changes[1]. This change can be appreciable: Empirically, drift associated with slight room temperature variability for the cantilevers (and mounting apparatus) used in this work can be $\sim 1\text{ Hz}$, the cantilever resonance frequency can shift by $> 10\text{ Hz}$ if it is illuminated with a lamp, and $> 100\text{ Hz}$ if it is illuminated with a laser. (This is why it is important to wait, after changing the illumination of the cantilever, before approaching the sample to perform an experiment.) In a typical imaging experiment with a $\sim -3\text{ Hz}$ setpoint, drift $\sim 1\text{ Hz}$ will cause the z -controller to move the tip slightly ($\sim nm$) away from the sample surface, which manifests in a change in “height” which does not actually correspond to the sample. In general, room temperature fluctuations are very slow, so this effect only becomes relevant for long imaging times. The images shown in this work (particularly the grid and multipass images) were measured over $\sim 8\text{ h}$: This is the timescale over which temperature variability might impact the measurement. If such drift occurs, it will manifest as a band (horizontally, if the sample is scanned top to bottom or vice versa) in the z channel and possibly the F_d channel. Correcting for this kind of drift is not straightforward, and if it occurs, the image should be measured again in a shorter amount of time to allow for a reliable quantitative interpretation of the image during the analysis stage.

Third, there are intrinsic nonlinearities in the AFM imaging mechanism which should be recognized. During an AFM experiment, the tip is moved over the sample surface (or the sample is moved under a stationary tip, as is the case in the JEOL) by piezoelectric materials (i.e. materials that expand or contract when a voltage is applied to them). Piezos tend to be hysteretic, which mostly manifests as differences in the forward (trace) and backward (retrace) images, and they tend to have a nonlinear (e.g. “bowed”) response, which manifests as an approximately parabolic background of the z signal in the x and y directions. Additionally, there might be an angle between the sample and the tip, which manifests as an approximately linear background in one direction. It is best to minimize these artefacts where possible (e.g. empirically, there is less piezo hysteresis if the scanner has already been active for some time, or the time per line can be longer to avoid hysteresis; smaller scan areas

lead to less piezo bowing; and many AFM controllers include “tilt correction” which account for the sample angle as the image is scanned). It is challenging to completely eliminate all of these effects, but since they do not correspond to actual changes in the tip-sample separation, they can be reliably corrected for during image analysis.

Fourth, once the image is measured it should be analyzed. The first step is to perform background subtractions for the image. To first order, this is accomplished by subtracting a parabolic background in both directions and a linear background along one direction, to account for piezo bowing and a tilt angle as described previously. These effects, however, might not be *exactly* parabolic and linear, so for a robust measurement, more care is needed. The basic principle is that some features of the sample are “known” to be flat. For example, in Figure A.4a, large areas of the KBr substrate – which is known to be atomically flat – are visible. A histogram of this substrate (see Figure A.5) should therefore have a single peak, and the width of the peak should correspond to the intrinsic measurement (white) noise. If the peak width is greater than the measurement noise, the sample is either rougher than the measurement sensitivity or it exhibits additional noise (both of which are “real” sample physics), or the background subtraction is not done appropriately. Which of the above cases is true depends on the sample/substrate. Background subtraction is an iterative process.

Finally, with the background subtraction complete, the heights can be measured. If the noise everywhere in the image is Gaussian, then histograms of the z channel should now be

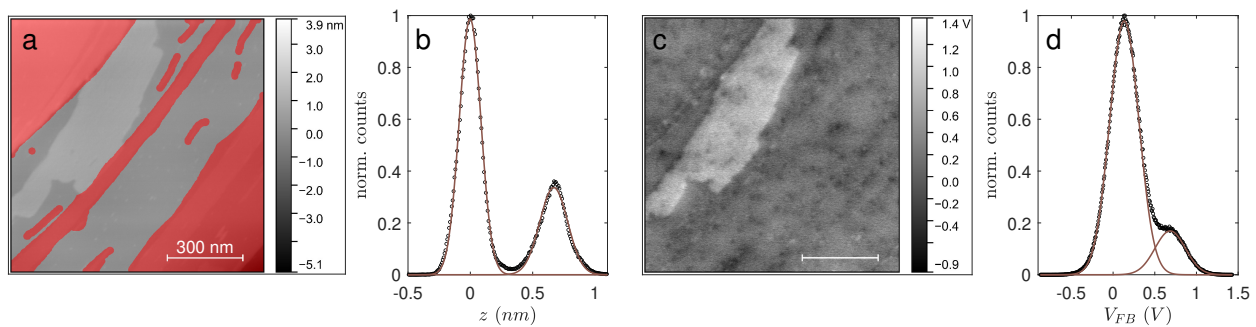


Figure A.5: Masking and fits of an AFM image. Masked images and histograms (points) and Gaussian fits (red) for the z (a-b) and V_{FB} (c-d) channels of a pentacene island on KBr. The fits in (b) correspond to the un-masked (non-red) regions in (a). The fits in (d) correspond to the entire area in (c). The R-square values for all Gaussian fits (four in total) are 1. The masking and background subtraction for all images with z channels displayed in this work are done using Gwyddion SPM software[2].

fitted with Gaussian peaks. The sample height is:

$$h_{sam} = \left[\bar{z}_{sam} - \bar{z}_{sub} \right] \pm \left[\sqrt{\delta z_{sam}^2 + \delta z_{sub}^2} \right] \quad (\text{A.1})$$

where $\bar{z}$ are the means of the Gaussian fits for the sample and substrate, and δz are the full-width half-maximum of each fit.

Note that the background subtraction described above should not be done for the V_{FB} (V_{CPD}) channel in the KPFM image. The “correctable” artefacts described above (e.g. piezo bowing) should not manifest in this channel; if a background subtraction appears to be required, the first kind of drift (the kind which does correspond to a change in tip-sample separation, e.g. heating) has occurred, which cannot be accounted for with the simple procedure outlined here). The un-corrected data in the V_{FB} channel should be Gaussian (if there is no significant non-Gaussian noise due to sample physics, and if the CPD on the substrate and sample is indeed uniform), and the width corresponds to the intrinsic sensitivity of the measurement.

As a final note, it can be difficult to exactly level the entire image. In Figure A.4, for example, the bottom right corner is not level (that is, a histogram the large terrace in this region is not Gaussian). In this sample, the island height was the desired measurement, so it was necessary to level the island and its surrounding substrate. The image can be masked (as in Figure A.5) to only evaluate the desired regions.

With the methodology outlined here, the substrate has $z = 0.00 \pm 0.09 \text{ nm}$ and the island has $z = 0.67 \pm 0.11 \text{ nm}$, meaning that the island height is $z = 0.67 \pm 0.14 \text{ nm}$. The substrate has $V_{FB} = 130 \pm 20 \text{ mV}$ and the island has $V_{FB} = 680 \pm 20 \text{ mV}$. The R-square values for all of the Gaussian fits shown are 1. The uncertainty for z corresponds to the base noise of this measurement, but for V_{FB} is higher than the measurement noise ($\sim 10 \text{ mV}$), indicating sample variability. (This corresponds to the “pock-mark” features in Figure A.5.)

The height found using this method is close to the values estimated using the line profile, but it has a robust uncertainty. It can also be noted, perhaps, that the images used for the line profiles had backgrounds subtracted by the methodology described above; with a different background subtraction, the line profile estimates and uncertainties could vary significantly.

A.4 Samples

Photos of the three samples measured in this work are shown in Figure A.6. The pentacene sample was uniformly evaporated onto a surface of KBr to achieve 0.3 monolayer coverage. The pentacene initially assembled in monolayer or bilayer islands $\sim 300 \text{ nm}^2$, which corresponded to about one island per the $4 \mu\text{m}^2$ scan area; locating pentacene islands was therefore very easy. (Though note that the pentacene islands generally became taller over time, which decreased the island density.)

The Si/SiO₂ and MoSe₂ samples, on the other hand, were much harder to find, because they were not uniformly distributed on the substrate, but rather were located in one specific location. The viewing angle of the tip-sample junction in the JEOL system is highly oblique, which makes it difficult to determine the precise position of the tip relative to the substrate by eye. The samples, therefore, needed to be located by AFM imaging. The samples were $\sim 2 \times 2 \mu\text{m}^2$ and the AFM scan area is $\sim 4 \times 4 \mu\text{m}$, which are both tiny compared to the $\sim 5 \text{ mm}^2$ substrate areas. Thankfully, both samples had markers which helped with this navigation. (The markers – a series of patterned crosses in the Si/SiO₂ sample and a series of gold numbers and corners in the MoSe₂ sample – are faintly visible in Figure A.6.) Essentially, the (generally featureless) substrate was scanned, the tip was retracted, the motors were moved in one direction by $\sim 4 \mu\text{m}$, and the substrate was scanned again (see Figure A.7). This was done repeatedly until a marker was found, which was used to determine

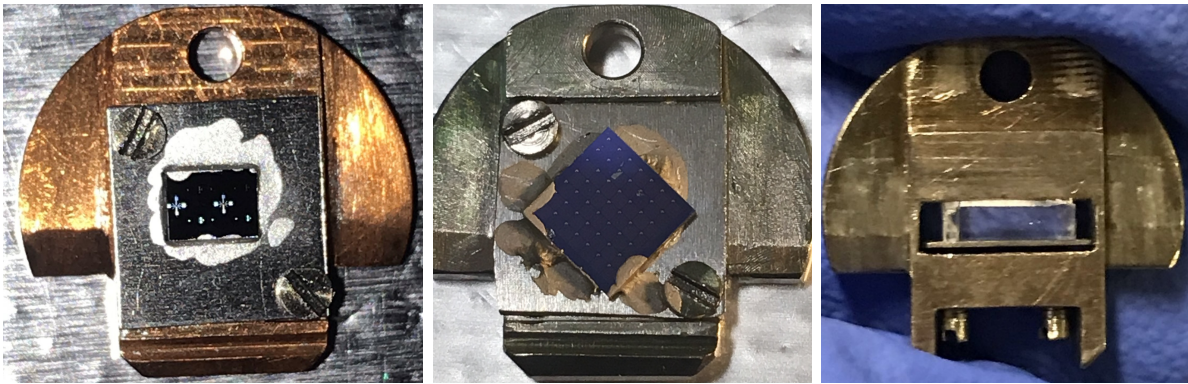


Figure A.6: **Samples.** Photos of the samples as mounted during experiments: Si/SiO₂ (left); MoSe₂ (middle); and pentacene (post-cleave in UHV and after molecular deposition, right). These images are $\sim 2 \times 2 \text{ cm}^2$.

the position of the tip on the substrate and then find the sample.

In the future, locating patterned or exfoliated samples with this approach would be significantly faster if the samples could be fabricated on optimized patterned substrates. The substrate markers would have the following features: They would be spaced every $\sim 20 - 50 \mu m$ so that one could be found within 5 – 10 approaches when moving along one direction; they would be easily distinguished from the substrate (to allow for fast scanning); they would be short (so that the tip would not crash into one while fast scanning) – perhaps in this sense buried markers with a very different V_{FB} (V_{CPD}) than the substrate would be ideal; they would be small (it must be possible to identify a marker with one or two $4 \times 4 \mu m^2$ scan areas); they would be asymmetric in all four directions (i.e. a two should not look like the upside down version of a five; an eight should have a different top and bottom); and they would be informative (one marker should be able to exactly reveal the sample location).

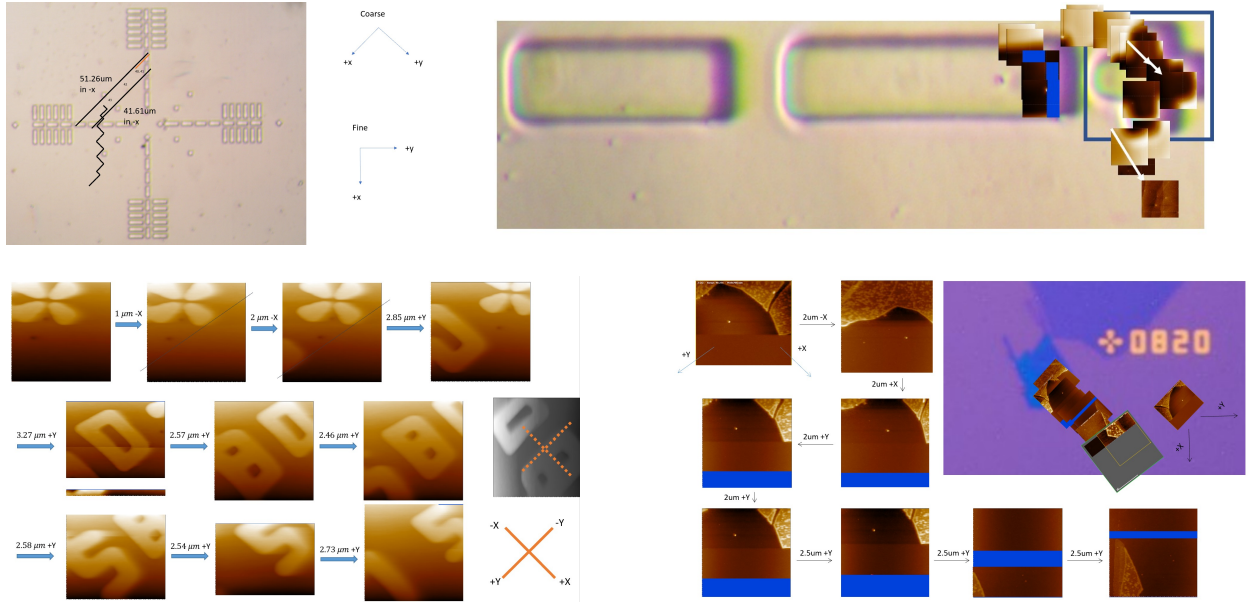


Figure A.7: **Sample navigation.** Images from the sample navigation process for the Si/SiO₂ (top) and MoSe₂ (bottom) samples. The Si/SiO₂ sample was approached in a region believed to be in the lower left quadrant of the cross (visible in Figure A.6). The tip was then moved along a zig-zag line (top left) until a marker (top right) was found. The marker was identified by imaging, and then the sample was located. For the MoSe₂ sample, the tip was moved along one direction until numbers were found. The numbers were read by a series of AFM images (bottom left). The sample was then located nearby (bottom right). All AFM images are $4 \times 4 \mu m^2$. The cross in the top right image is $\sim 15 \times 15 \mu m^2$ and $\sim 300 nm$ tall. The numbers in the bottom images are $\sim 2 \times 2 \mu m^2$, and $\sim 50 nm$ tall.

Appendix B

Model & fitting details

B.1 Distance dependencies

A rigorous assessment of the MIS model fit acknowledges the full distance-dependent force $F_{ts}(z)$. This is done by comparing the MIS model to experimental spectra at varying tip lift (z_{TL}) above the closest approach position $z_{ins,c}$, shown in Figure B.1. Modelled results are in black and experimental spectra are coloured, where each colour represents a different $z_{ins,c}$ (i.e. a different z_{TL}). Ten sweeps are shown for each z_{TL} . Figure B.1a shows that there is good agreement between the model and the data within the measurement uncertainty, which is captured by the spread of the sweeps. Figure B.1b-e show these results at four arbitrary biases. No fits are shown for $F_d(V_g)$ (Figure B.1(f-j)) F_d is bias-dependent, as was discussed in Chapter 4. Figure B.1k shows the relative difference between the data and the model, which is less than 10%. Finally, Figure B.1i shows the difference of each Δf spectrum (data and model) with its parabolic fit. This figure is intended to demonstrate the non-parabolicity of both the data and the model. There are differences in the shapes of these curves, but overall they have a similar trend and order of magnitude. The data in Figure B.1 were used to determine the relative contributions of the tip (r, z_{ins}) and cantilever ($a_c, z_{ins,c}$) in the MIS model. If the tip contribution is too small with respect to the cantilever, $\Delta f(V_g)$ is approximately constant as a function of z_{TL} ; if the tip contribution is too big, the shape of $\Delta f(V_g)$ changes too rapidly with z_{TL} . The geometries found with this method are consistent with those specified by the tip manufacturer, as discussed in Chapter 7.

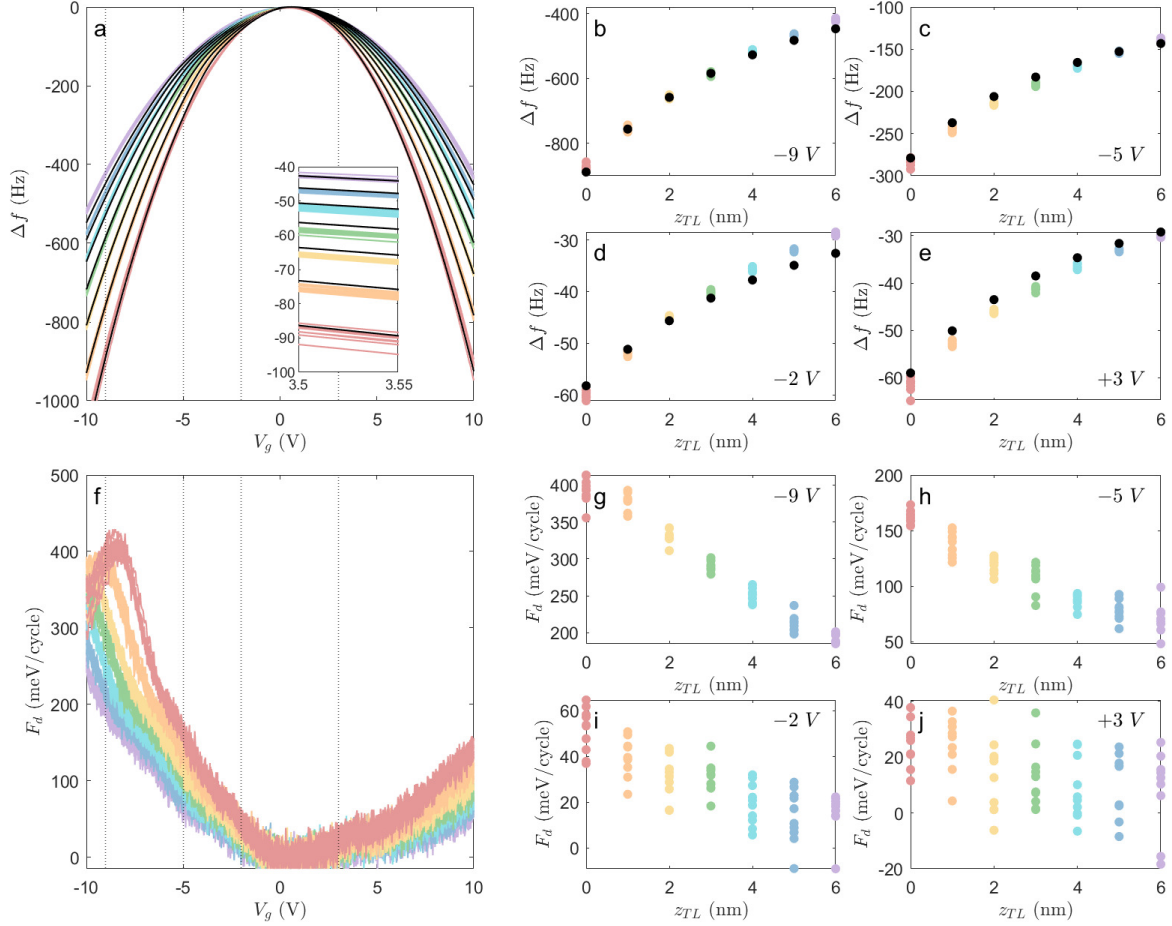


Figure B.1: **Comparison of Si measurements and the MIS model.** Experimental bias sweeps for $z_{TL} = 0$ nm (red), 1 nm (orange), 2 nm (yellow), 3 nm (green), 4 nm (teal), 5 nm (blue), and 6 nm (purple). Ten spectra are shown overlaid for each z_{TL} . Corresponding modelled results are shown in black. (a) Shows the full Δf spectrum, and (b-e) show cuts at $V_g = -9$ V, -5 V, -2 V, and $+3$ V, respectively. Simultaneously measured F_d spectra are shown in (f-j). [Figure from [3].]

B.2 Si/SiO₂ fit assessment

Figure B.2 demonstrates the sensitivity of MIS sample parameters for the Si/SiO₂ surface. (Figures B.3 and B.4 in Appendices B.3 and B.4 show similar assessments for MoSe₂ and pentacene.)

An experimental $\Delta f(V_g)$ bias spectrum is shown, along with the model used for this work (where the only variable throughout this work was z_{ins} , according to the tip lift set during the the experiment.) The greyscale image shows $\Delta f(V_g)$ for a large parameter space covering “reasonable” sample parameter values, for E_g , ϵ , χ , N_d , α , z_{ins} , m_e , m_h , and T . Below the greyscale images, “cuts” in this parameter space are shown as bias spectra.

Some parameters are much more sensitive than others. In all samples, variations in the non-parabolicity (corresponding to the “steepness” of the $V_S(V_g)$ curve) is primarily affected by E_g , ϵ , and $N_{d(a)}$. χ simply introduces a lateral offset. z_{ins} and α lead to vertical stretching and compression of the $\Delta f(V_g)$ curve. m_e , m_h , and T are largely insensitive parameters. There is a discontinuity in the E_g parameter space for MoSe₂. This appears to be an artefact of the numerical solver and currently its origin is unknown. Overall, there is good agreement between the experimental data and the model, and all of the values used in modelling these results are “realistic”, as discussed in Chapter 7.

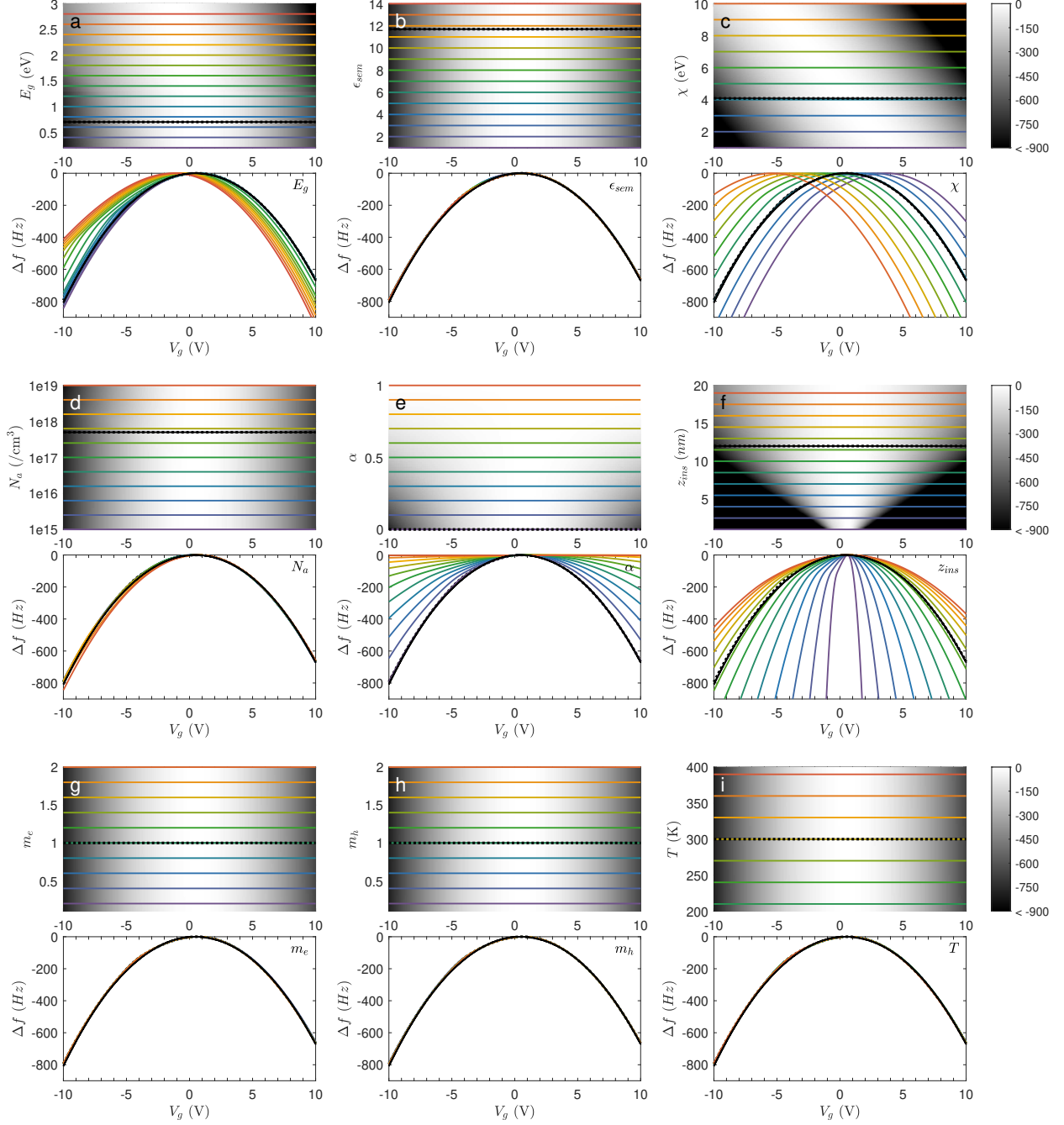


Figure B.2: **Si/SiO₂ MIS model sensitivities.** Δf bias spectra for varying MIS parameters: (a-i) E_g , ϵ , χ , N_d , α , z_{ins} , m_e , m_h , and T . The grey colour scale in the top images shows Δf . The bottom images show “slices” of $\Delta f(V_g)$ for the parameters with the corresponding colour in the top images. The dashed black line corresponds to the parameter used for modelling the sample; the solid black line shows a measured Si/SiO₂ spectrum.

B.3 MoSe₂ fit assessment

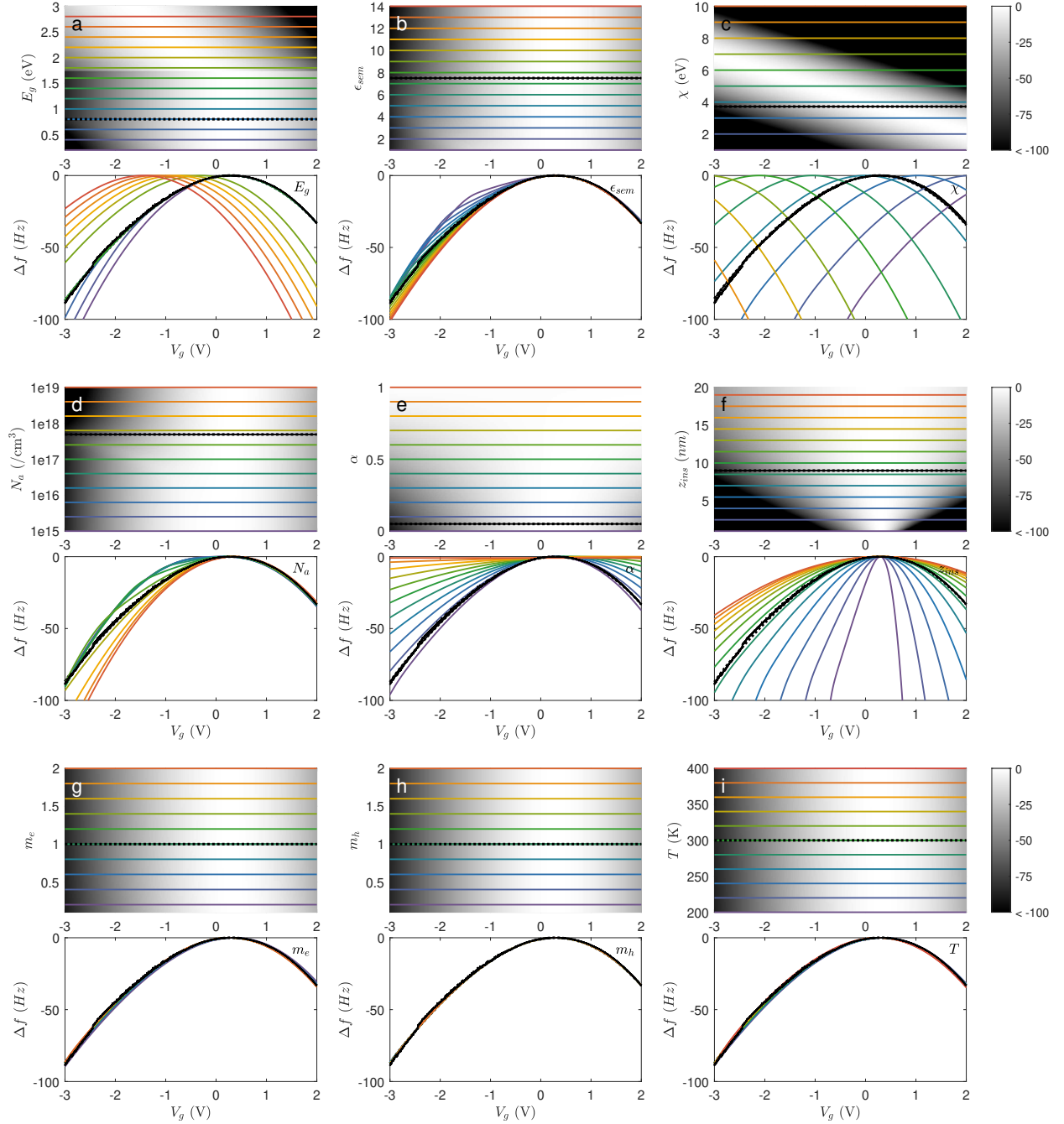


Figure B.3: **MoSe₂ MIS model sensitivities.** Δf bias spectra for varying MIS parameters: (a-i) E_g , ϵ , χ , N_d , α , z_{ins} , m_e , m_h , and T . The grey colour scale in the top images shows Δf . The bottom images show “slices” of $\Delta f(V_g)$ for the parameters with the corresponding colour in the top images. The dashed black line corresponds to the parameter used for modelling the sample; the solid black line shows a measured MoSe₂ spectrum.

B.4 Pentacene fit assessment

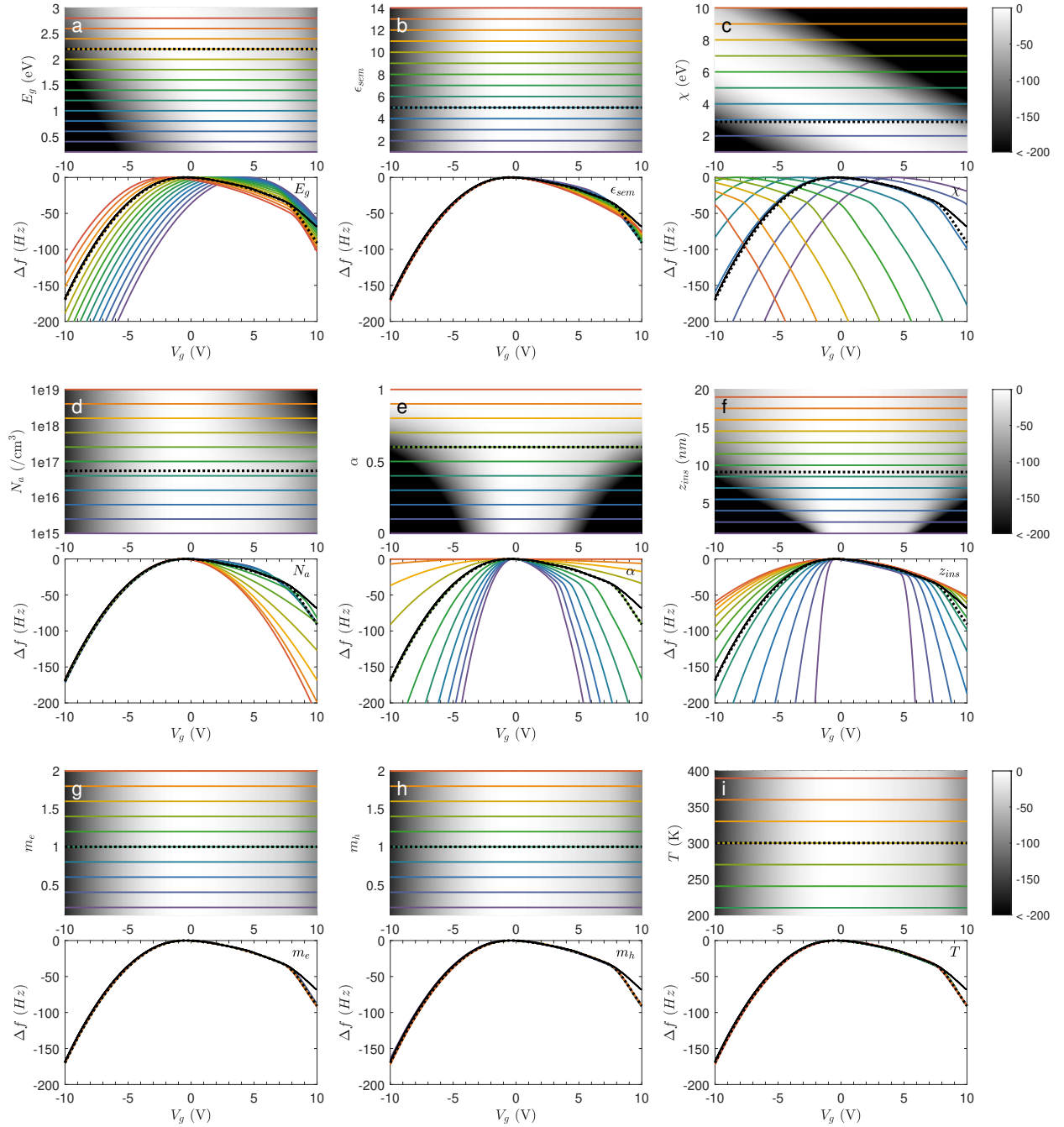


Figure B.4: **Pentacene MIS model sensitivities.** Δf bias spectra for varying MIS parameters: (a-i) E_g , ϵ , χ , N_a , α , z_{ins} , m_e , m_h , and T . The grey colour scale in the top images shows Δf . The bottom images show “slices” of $\Delta f(V_g)$ for the parameters with the corresponding colour in the top images. The dashed black line corresponds to the parameter used for modelling the sample; the solid black line shows a measured pentacene spectrum.

B.5 RTS timescales

The RTS timescales in Chapter 6 correspond to the amount of time the system spends in the “up” and “down” states before switching. This Appendix describes the RTS timescale measurement procedure.

In Figure B.5 the shaded regions correspond to the standard deviation of the Gaussian fits (b,d). The first data point in the time series is identified as being in the “up” state (0 for Δf ; 1 for F_d) or the “down” state. Then, starting with $t = 0$ and moving forward in the time series, every data point is compared to the point before. If the state is “up”, it is defined to flip to “down” when the value is less than the highest value in the shaded region of the “down” peak. If the state is “down”, it flips to “up” when the value is greater than the lowest value in the shaded region of the “up” peak.

Figure B.5c,f show the values at which the state flipped. In the F_d analysis, where the noise amplitude is larger than the peak width, the measured flip values are approximately normally distributed in each state. In the Δf analysis, however, the measured flip values are not normally distributed, indicating an overestimation of flips. This manifests as shorter measured RTS timescales for Δf (see the Figure B.5 caption). The arrows in (a,d) point to mismatches in the Δf and F_d flips, but overall the correlation between the two channels is high ($\Phi = 0.8$), and the RTS timescales are close within measurement uncertainty.

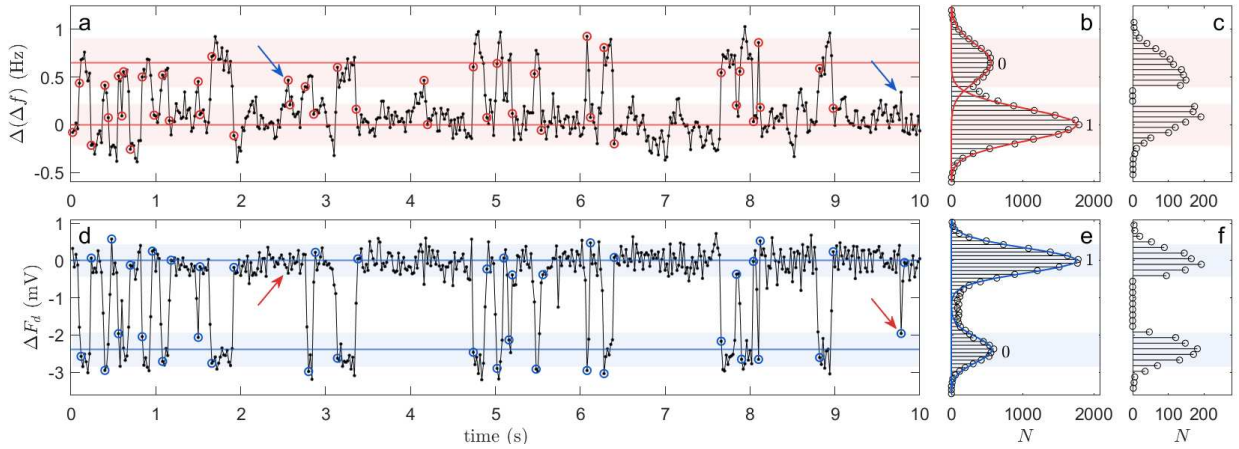


Figure B.5: **RTS timescales measurement.** RTS of the Si/SiO₂ sample measured at $V_g = -6.5$ V. The timescales found using the methodology described above are: $\tau_0^{\Delta f} = 0.093 \pm 0.007$ s, $\tau_1^{\Delta f} = 0.28 \pm 0.014$ s, $\tau_0^{F_d} = 0.10 \pm 0.01$ s, $\tau_1^{F_d} = 0.33 \pm 0.02$ s.

Appendix C

Derivations

C.1 The classical limit of Fermi-Dirac

This Appendix shows when it is appropriate to use the Maxwell Boltzmann distribution in lieu of Fermi Dirac. The Fermi-Dirac ($f_{FD}(E)$) and Maxwell-Boltzmann ($f_{MB}(E)$) distribution functions are given below:

$$f_{FD}(E) = \frac{1}{1 + e^{(E-E_f)/k_B T}} \quad f_{MB}(E) = \frac{1}{e^{(E-E_f)/k_B T}} \quad (\text{C.1})$$

In the limit $\exp\left(\frac{E-E_f}{k_B T}\right) \gg 1$ (i.e. $E - E_f \gg k_B T$), $f_{FD}(E) \approx f_{MB}(E)$. This limit is sometimes called the “high-temperature limit”, where what is meant by “temperature” in this statement is the particle energy E : At high energy E the density of states is large, approaching the classical (non-quantized) case. More specific is to show where $f_{FD}(E)$ and $f_{MB}(E)$ agree within 5%[4]:

$$\begin{aligned} \frac{f_{MB}(E) - f_{FD}(E)}{f_{FD}(E)} &= 0.05 \\ e^{-(E-E_f)/k_B T} &= 0.05 \end{aligned} \quad (\text{C.2})$$

$$E - E_f \approx 3k_B T$$

In the case of an electron in the lowest energy of the conduction band of an intrinsic semiconductor with $E_g = 1 \text{ eV}$ (i.e. $E - E_f \sim 0.5 \text{ eV}$) at room temperature (300 K, 0.025 eV), $f_{FD}(E)$ and $f_{MB}(E)$ agree within $\sim 10^{-7}\%$.

C.2 Bulk carrier densities

This Appendix shows the derivation of Equation 2.6.

Equation 2.6 is the solution of Equation 2.5, given the density of states $g(E)$ (Equation 2.4) and approximating the Fermi-Dirac distribution function ($f_{FD}(E)$) $\approx$ the Maxwell-Boltzmann distribution function ($f_{MB}(E)$) distribution. For electrons, this gives:

$$\begin{aligned}
 n &= \int_{E_C}^{\infty} g_c(E) f_{MB}(E) dE \\
 &= \int_{E_C}^{\infty} \frac{1}{2\pi^2} \left(\frac{2m_n}{\hbar^2} \right)^{3/2} \sqrt{E - E_C} \left[\exp \left(\frac{E_f - E}{k_B T} \right) \right] dE \\
 &= \frac{1}{2\pi^2} \left(\frac{2m_n}{\hbar^2} \right)^{3/2} \int_0^{\infty} \sqrt{k_B T x} \left[\exp \left(\frac{E_f - E_C}{k_B T} - x \right) \right] k_B T dx \\
 &= \frac{1}{2\pi^2} \left(\frac{2m_n k_B T}{\hbar^2} \right)^{3/2} \left[\exp \left(\frac{E_f - E_C}{k_B T} \right) \right] \int_0^{\infty} \sqrt{x} \exp(-x) dx \\
 &= \frac{1}{2\pi^2} \left(\frac{2m_n k_B T}{\hbar^2} \right)^{3/2} \left(\frac{\sqrt{\pi}}{2} \right) \left[\exp \left(\frac{E_f - E_c}{k_B T} \right) \right] \\
 &= \frac{1}{\sqrt{2}} \left(\frac{m_n k_B T}{\pi \hbar^2} \right)^{3/2} \left[\exp \left(\frac{E_f - E_c}{k_B T} \right) \right] \\
 &= N_c \exp \left(\frac{E_f - E_c}{k_B T} \right)
 \end{aligned}$$

where as an intermediate step on line 2 I have made the substitution $x = \frac{E - E_c}{k_B T}$. N_c is the effective density of states in the conduction band. A similar derivation can be done to find N_v , the effective density of states in the valence band, to give:

$$N_c = \frac{1}{\sqrt{2}} \left(\frac{m_n k_B T}{\pi \hbar^2} \right)^{3/2} \quad (\text{C.3a})$$

$$N_v = \frac{1}{\sqrt{2}} \left(\frac{m_p k_B T}{\pi \hbar^2} \right)^{3/2} \quad (\text{C.3b})$$

C.3 Bulk intrinsic energy and density

This Appendix shows the derivation of Equations 2.7 and 2.8.

For an intrinsic semiconductor, it is necessarily true that $n = p$, and by definition we set $n_i = n = p$ and $E_i = E_f$. So now we can derive the intrinsic level, starting from Equation 2.6 and recalling the expressions for N_v and N_c derived in Appendix C.2:

$$\begin{aligned}
 N_c \exp\left(\frac{E_i - E_c}{k_B T}\right) &= N_v \exp\left(\frac{E_v - E_i}{k_B T}\right) \\
 \exp\left(\frac{2E_i - E_c - E_v}{k_B T}\right) &= \frac{N_v}{N_c} \\
 E_i &= \frac{E_c + E_v}{2} + \frac{k_B T}{2} \ln\left(\frac{N_v}{N_c}\right) \\
 E_i &= E_{midgap} + \frac{k_B T}{2} \ln\left(\frac{m_p}{m_n}\right)^{3/2}
 \end{aligned} \tag{C.4}$$

where E_{midgap} is the energy level exactly mid-gap. Therefore, we can see that E_i only lies at mid-gap if $m_n = m_p$ or if $T = 0$ K.

Given that $E_g = E_c - E_v$, and given Equation 2.6, for an intrinsic semiconductor where $E_f = E_i$:

$$\begin{aligned}
 n &= N_c \exp\left(\frac{E_i - E_c}{k_B T}\right) \\
 &= N_c \exp\left(\frac{E_i - (E_g + E_v)}{k_B T}\right) \\
 &= N_v \exp\left(\frac{-(E_c - E_i) - (E_g)}{k_B T}\right) \\
 &= N_c \frac{N_v}{p} \exp\left(\frac{-E_g}{k_B T}\right) \\
 n \times p &= N_c N_v \exp\left(\frac{-E_g}{k_B T}\right) \\
 n_i &= \sqrt{N_c N_v} \exp\left(\frac{-E_g/2}{k_B T}\right)
 \end{aligned} \tag{C.5}$$

C.4 Bulk extrinsic carrier densities

This Appendix shows the derivation of Equation 2.10.

Generically, $n_i^2 = n \times p$. Therefore, $p = \frac{n_i^2}{n}$, and Equation 2.9 can be rewritten as:

$$\begin{aligned} 0 &= p - n + N_d - N_a \\ 0 &= \frac{n_i^2}{n} - n + N_d - N_a \\ n &= \frac{(N_d - N_a)}{2} \pm \sqrt{\left(\frac{N_d - N_a}{2}\right)^2 + n_i^2} \end{aligned} \quad (\text{C.6})$$

It is necessarily true that when $N_d = N_a$, $n = n_i$, so the positive sign corresponds to the physical solution. Then, the solution is positive when $N_d > N_a$, which corresponds to an n-type semiconductor. For a p-type semiconductor, the above yields a negative n , which is not physical, so the above equation does not hold.

For a p-type semiconductor, the above derivation is repeated starting with $n = \frac{n_i^2}{p}$, giving:

$$p = \frac{(N_a - N_d)}{2} \pm \sqrt{\left(\frac{N_a - N_d}{2}\right)^2 + n_i^2} \quad (\text{C.7})$$

for which, again, taking the positive sign and demanding $N_a > N_d$ corresponds to the physical solution. Charge neutrality demands that:

$$0 = p - n + N_d - N_a \quad (\text{C.8})$$

Therefore, the free electron and hole concentrations of an n-type semiconductor are:

$$n_n = \frac{N_d - N_a}{2} + \sqrt{\left(\frac{N_d - N_a}{2}\right)^2 + n_i^2} \quad ; \quad p_n = \frac{n_i^2}{n} \quad (\text{C.9})$$

and for a p-type semiconductor,

$$p_p = \frac{N_d - N_a}{2} + \sqrt{\left(\frac{N_d - N_a}{2}\right)^2 + n_i^2} \quad ; \quad n_p = \frac{n_i^2}{p} \quad (\text{C.10})$$

C.5 The MIM capacitor force

This Appendix shows the derivation of Equations 2.15 and 2.16.

In an MIM capacitor, the potential of each plate is spatially uniform, such that $\int_v \rho V \partial v = V \int_v \rho \partial v$. Additionally, the charge densities of each plate are located entirely at the metal surfaces, so that the total charge $Q = \int_v \rho \partial v = \int_a \sigma \partial a$, where a is the plate area. Furthermore, the surface charge distribution is spatially uniform, so that $\int_a \sigma \partial a = \sigma \int_a \partial a = \sigma a = Q$. Therefore, the energy of a capacitor (Equation 2.14) can be simplified as:

$$U_{AB} = \frac{1}{2} \left(\int_A \rho_A V_A \partial v_A + \int_B \rho_B V_B \partial v_B \right) \quad (\text{C.11a})$$

$$= \frac{1}{2} (Q_A V_A + Q_B V_B) \quad (\text{C.11b})$$

Given the requirement for conservation of electric charge (i.e. assuming charge neutrality in the system), $Q_{AB} = Q_A = -Q_B$. Furthermore, the potential difference between the plates is $V_{AB} = V_A - V_B$. Therefore, the above expression can be re-written as:

$$\begin{aligned} U_{AB} &= \frac{1}{2} Q_{AB} V_{AB} \\ &= \frac{1}{2} C_{AB} V_{AB}^2 \end{aligned} \quad (\text{C.12})$$

where, by definition, $C_{AB} = Q_{AB}/V_{AB}$.

Generically, $F_{AB} = -\nabla U_{AB}$. Assuming a one-dimensional system, $F_{AB} = -\partial U_{AB}/\partial z_{AB}$. Therefore, the capacitive force of a one-dimensional MIM capacitor is:

$$F_{AB} = -\frac{1}{2} \left[\frac{\partial C_{AB}}{\partial z_{AB}} V_{AB}^2 + \frac{\partial V_{AB}^2}{\partial z_{AB}} C_{AB} \right] \quad (\text{C.13})$$

According to Equation 2.12, $V_{AB} = V_g - V_{CPD}$ (which is z -independent in a perfectly one-dimensional system), so:

$$F_{AB} = -\frac{1}{2} \frac{\partial C_{AB}}{\partial z_{AB}} (V_g - V_{CPD})^2 \quad (\text{C.14})$$

C.6 The MIS capacitor electric field

This Appendix shows the derivation of Equation 2.19.

The number of carriers near the semiconductor surface ($n_d(z)$ and $p_d(z)$) is z -dependent due to the z -dependent electric field, such that the total charge density in the depletion region $\rho_d(z)$ is:

$$\rho_d(z) = |e| (p_d(z) - n_d(z) - N_a + N_d) \quad (\text{C.15})$$

with

$$n_d(z) = n \exp\left(\frac{|e|V(z)}{k_B T}\right) \quad (\text{C.16a})$$

$$p_d(z) = p \exp\left(\frac{-|e|V(z)}{k_B T}\right) \quad (\text{C.16b})$$

where $V(z)$ is the spatially variable potential and n and p are the bulk carrier densities given by Equation 2.6 or Equation 2.6. Substituting Equations 2.17 and 2.18 into the one-dimensional Poisson equation ($-\frac{\partial^2 V(z)}{\partial z^2} = \frac{\rho_d(z)}{\epsilon}$) gives:

$$\frac{\partial^2 V(z)}{\partial z^2} = \frac{-|e|}{\epsilon} \left[p \exp\left(\frac{-|e|V(z)}{k_B T}\right) - n \exp\left(\frac{|e|V(z)}{k_B T}\right) - N_a + N_d \right] \quad (\text{C.17})$$

In the semiconductor bulk, due to charge neutrality, it is necessarily true that $N_d - N_a = n - p$.

Therefore, the above equation can be written as:

$$\frac{\partial^2 V(z)}{\partial z^2} = \frac{-|e|}{\epsilon} \left[p \left(\exp\left(\frac{-|e|V(z)}{k_B T}\right) - 1 \right) - n \left(\exp\left(\frac{|e|V(z)}{k_B T}\right) - 1 \right) \right] \quad (\text{C.18})$$

Given that the electric field is generically $E(r) = -\nabla V(r)$ (in one dimension, $E(z) = \frac{\partial V(z)}{\partial z}$), integrating the above equation once (with the boundary conditions $V(0) = V_S$ and $\frac{\partial V(\infty)}{\partial z} = 0$) gives the solution for the electric field at any point:

$$\begin{aligned} \int_0^{[\frac{\partial V}{\partial z}]_z} \frac{\partial V}{\partial z} \partial \left(\frac{\partial V}{\partial z} \right) &= \frac{-|e|}{\epsilon} \int_0^{V_z} p \left(\exp\left(\frac{-|e|V}{k_B T}\right) - 1 \right) - n \left(\exp\left(\frac{|e|V}{k_B T}\right) - 1 \right) \partial V \\ E_z^2(z) &= \frac{k_B T}{\epsilon} \left[p \left(\exp\left(\frac{-|e|V_z(z)}{k_B T}\right) + \frac{|e|V_z(z)}{k_B T} - 1 \right) + n \left(\exp\left(\frac{|e|V_z(z)}{k_B T}\right) - \frac{|e|V_z(z)}{k_B T} - 1 \right) \right] \end{aligned} \quad (\text{C.19})$$

C.7 The MIS capacitor energy

This Appendix shows the derivation of Equation 2.24.

The generic expression for the energy of a capacitor is given in Equation C.11. If the first capacitor plate (A) is metallic, ρ_A and V_A are both spatially constant. If the second plate (B) is semiconducting, ρ_B and V_B are both spatially variable, such that in one dimension:

$$U = \frac{Q_M V_M}{2} + \frac{1}{2} \int_0^\infty \rho(z) V(z) \partial z \quad (\text{C.20})$$

where $Q_M = \rho_A v_A$ and V_M are the total metallic charge and potential. Poisson's equation in one dimension is:

$$\frac{\partial^2 V}{\partial z^2} = \frac{-\rho(z)}{\epsilon} \quad (\text{C.21})$$

and given that $\frac{\partial V(\infty)}{\partial z} = 0$, Equation C.21 can be integrated once spatially to give:

$$Q_S = -\epsilon \frac{\partial V(0)}{\partial z} \quad (\text{C.22})$$

where $-Q_M = Q_S = \int_0^\infty \rho(z) \partial z$ is the total charge in the semiconductor. The second integral of Equation C.20 can now be solved by integration by parts, setting $u = V(z)$, $\partial u = \partial V$, $v = \frac{\partial V}{\partial z}$, and $\partial v = \frac{\partial^2 V}{\partial z^2} \partial z$ (for $\int_a^b u \partial v = u(a)v(a) - u(b)v(b) - \int_a^b v \partial u$):

$$\begin{aligned} \frac{1}{2} \int_0^\infty \rho(z) V(z) \partial z &= -\epsilon \int_0^\infty \frac{\partial^2 V}{\partial z^2} V(z) \partial z \\ &= -\frac{\epsilon}{2} \left(\left[V(z) \frac{\partial V}{\partial z} \right]_0^\infty - \int_0^\infty \frac{\partial V}{\partial z} \partial V \right) \\ &= \frac{\epsilon}{2} \left(V_S Q_S + \int_{V_S}^0 \frac{\partial V(z)}{\partial z} \partial V \right) \end{aligned}$$

where the integration bounds are set given $V(z=0) = V_S$ and $V(z \rightarrow \infty) = 0$. Therefore, the energy of a capacitor with one metallic plate and one semiconducting plate is[5]:

$$U = \frac{Q_M V_M}{2} + \frac{Q_S V_S}{2} + \frac{\epsilon}{2} \int_{V_S}^0 \frac{\partial V(z)}{\partial z} \partial V \quad (\text{C.23})$$

C.8 The MIS capacitor force

This Appendix shows the derivation of Equation 2.25.

In the derivation of Equation 2.24, no statements were made about V_M except that it is spatially constant. So, V_M can be defined as $V_M = V_g - V_{CPD}$, such that:

$$U = \frac{Q_M(V_g - V_{CPD})}{2} + \frac{Q_S V_S}{2} + \frac{\epsilon}{2} \int_{V_S}^0 \frac{\partial V(z)}{\partial z} \partial V \quad (\text{C.24})$$

Equation 2.21, so $V_g = V_{CPD} + V_S + V_{ins}$. Additionally, $Q_M = -Q_S$, $V_{ins} = -Q_S/C_{ins}$, and $C_{ins} = \epsilon_o/z$, so Equation 2.24 can be written as:

$$U = \frac{Q_S^2 z}{2\epsilon_o} + \frac{\epsilon}{2} \int_{V_S}^0 \frac{\partial V(z)}{\partial z} \partial V \quad (\text{C.25})$$

The electrostatic force can be found according to $\vec{F} = -\vec{\nabla}U$ (in one dimension, $\vec{F} = -\frac{\partial U_z}{\partial z}$). Q_S and V_S do not depend on z , so:

$$F = -\frac{Q_S^2}{2\epsilon_o} \quad (\text{C.26})$$

This derivation differs slightly from that presented in Hudlet (1995)[5] in that it includes V_{CPD} . However, because V_{CPD} is z -independent, the end result is identical.

C.9 The damped, driven free harmonic oscillator

The equation of motion for a linearly damped, driven, free (no external force) harmonic oscillator is:

$$m\ddot{z} + \xi\dot{z} + kz = F_{drive} \quad (\text{C.27})$$

(though note that it can be set up in different ways, see e.g. Section 3.1 for a discussion of the fm-AFM equation of motion.) If the drive is periodic:

$$F_{drive}(t) = \text{Re} [F_d e^{-i\omega t}] \quad (\text{C.28a})$$

$$= |F_d| \cos(\omega t) \quad (\text{C.28b})$$

the position is periodic:

$$z(t) = \text{Re} [A e^{i(-\omega t + \phi)}] \quad (\text{C.29a})$$

$$= |A| \cos(\omega t - \phi) \quad (\text{C.29b})$$

$$= A_{in} \cos(\omega t) - A_{out} \sin(\omega t) \quad (\text{C.29c})$$

where ϕ is the angle in the complex plane and $A_{in} = |A| \cos(\phi)$ and $A_{out} = |A| \sin(\phi)$. Combining Equation C.27 with $F_{drive}(t)$ and $z(t)$ gives:

$$A = \frac{F_d e^{-i\phi}}{k - m\omega^2 - i\omega\xi} \quad (\text{C.30})$$

According to Equation C.29, $z(t)$ can be fully expressed either in terms of $|A|$ and ϕ (Equation C.29b) or A_{in} and A_{out} (Equation C.29c). These solutions are:

$$|A| = \frac{F_d}{\sqrt{(k - m\omega^2)^2 + (\omega\xi)^2}} \quad ; \quad \phi = \tan^{-1} \left(\frac{\omega\xi}{k - m\omega^2} \right) \quad (\text{C.31a})$$

$$A_{in} = \frac{F_d(k - m\omega^2)}{(k - m\omega^2)^2 + (\omega\xi)^2} \quad ; \quad A_{out} = \frac{F_d(\omega\xi)}{(k - m\omega^2)^2 + (\omega\xi)^2} \quad (\text{C.31b})$$

When $\omega = \sqrt{\frac{k}{m}} := \omega_o = 2\pi f_o$ (that is, at the resonance frequency), A is maximized and $\phi = 90^\circ$ (see Figure 3.2 in the main text), or equivalently $A_{in} = 0$ and A_{out} is maximized.

C.10 fm-ac-EFM

This Appendix shows the derivation of Equation 3.13.

The position of the cantilever in fm-AFM can be described as:

$$z(t) = z_o + A \cos(\omega t) := z_o + A \cos(\omega_o t + \theta) := z_o + A \cos(\phi) \quad (\text{C.32})$$

where the instantaneous frequency $\omega = \partial\phi/\partial t$. Given $\Delta\omega = \omega - \omega_o$, and given Equation 3.10:

$$\begin{aligned} \phi &= \int \omega \partial t = \int [\omega_o + \Delta\omega] \partial t \\ &= \int \left[\omega_o + \frac{\omega_o}{4k} \frac{\partial^2 C}{\partial z^2} ((V_{DC} - V_{CPD}) + V_{AC} \sin(\omega_{AC} t))^2 \right] \partial t \\ &= \omega_o t + \frac{\omega_o}{4k} \frac{\partial^2 C}{\partial z^2} \left[t \left((V_{DC} - V_{CPD})^2 + \frac{V_{AC}^2}{2} \right) \dots \right. \\ &\quad \left. - \frac{1}{\omega_{AC}} \left(2V_{AC}(V_{DC} - V_{CPD}) \cos(\omega_{AC} t) + \frac{V_{AC}^2}{4} \sin(2\omega_{AC} t) \right) \right] \end{aligned} \quad (\text{C.33})$$

which can be expressed in terms of its frequency components, where $\phi = \phi_{DC} + \phi_{1\omega_{AC}} + \phi_{2\omega_{AC}}$:

$$\phi_{DC} = \omega_o t + \frac{\omega_o}{4k} \frac{\partial^2 C}{\partial z^2} \left((V_{DC} - V_{CPD})^2 + \frac{V_{AC}^2}{2} \right) t \quad (\text{C.34a})$$

$$\phi_{\omega_{AC}} = \frac{\omega_o}{4k} \frac{\partial^2 C}{\partial z^2} \left(\frac{-2V_{AC}(V_{DC} - V_{CPD})}{\omega_{AC}} \right) \cos(\omega_{AC} t) \quad (\text{C.34b})$$

$$\phi_{2\omega_{AC}} = \frac{\omega_o}{4k} \frac{\partial^2 C}{\partial z^2} \left(\frac{-V_{AC}^2}{4\omega_{AC}} \right) \sin(2\omega_{AC} t) \quad (\text{C.34c})$$

Therefore, the instantaneous frequency $\omega = \partial\phi/\partial t = \omega_{DC} + \omega_{1\omega_{AC}} + \omega_{2\omega_{AC}}$ gives rise to an instantaneous frequency shift $\Delta\omega$, as given in Equation 3.13 of the main text.

This derivation is roundabout, since first there is an integral and later a derivative, which undo each other. The final result is found much more quickly by expanding the quadratic in Equation 3.10, but in this expansion the instantaneous phase ϕ is found, which is how Figure 3.5d is drawn ($z = A \sin(\phi)$.)

C.11 Exponential probability distribution

If the switching mechanism of an RTS as function of time is a discrete, random, memoryless process with a characteristic timescale, the number N_τ of $0 \rightarrow 1$ (or $1 \rightarrow 0$) events over a time interval τ is Poisson distributed:

$$P(N = N_\tau) = \frac{(\lambda\tau)^{N_\tau} e^{-\lambda\tau}}{N_\tau!} \quad (\text{C.35})$$

where $P(N = N_\tau)$ is the probability mass function and λ ($\lambda \in [0 \infty]$) is the expected (average) rate of events such that $\langle N_t \rangle = \lambda\tau = \tau/\tau_S$, where τ_S is the characteristic lifetime of the state. If τ_t is the wait time before a transition, it follows that the probability of no transition ($N = 0$, i.e. $\tau_t > \tau$) is:

$$P(N = 0) = P(\tau_t > \tau) = \frac{(\lambda\tau)^0 e^{-\lambda\tau}}{0!} = e^{-\lambda\tau} \quad (\text{C.36})$$

and so the probability that a transition will occur over the interval τ is the cumulative distribution function:

$$P(N > 0) = P(\tau_t < \tau) = 1 - e^{-\lambda\tau} \quad (\text{C.37})$$

Therefore, the probability of a transition occurring at time τ is:

$$\begin{aligned} P(\tau = \tau_t) &= \frac{\partial}{\partial \tau} (1 - e^{-\lambda\tau}) \\ &= \lambda e^{-\lambda\tau} \\ &= \frac{1}{\tau_S} e^{-\tau/\tau_S} \end{aligned} \quad (\text{C.38})$$

which is the exponential probability distribution function.

References

- [1] Iervolino, E., Riccio, M., Santagata, F., Wei, J., Herwaarden, A. van, Irace, A., Breglio, G., and Sarro, P. “Resonance frequency of locally heated cantilever beams”. *Sensors and Actuators A: Physical* 190 (**2013**). DOI: 10.1016/j.sna.2012.10.008.
- [2] *Gwyddion – Free SPM (AFM, SNOM/NSOM, STM, MFM, ...) data analysis software*. <http://gwyddion.net/>. Accessed: 2023-07-28.
- [3] Cowie, M., Stock, T. J. Z., Constantinou, P. C., Curson, N., and Grütter, P. *Spatially resolved dielectric loss at the Si/SiO₂ interface*. **2023**. DOI: 10.48550/ARXIV.2306.13648.
- [4] Neamen, D. A. *Semiconductor physics and devices : Basic principles*. 4th ed. McGraw-Hill, **2012**. ISBN: 9780073529585.
- [5] Hudlet, S., Jean, M. S., Roulet, B., Berger, J., and Guthmann, C. “Electrostatic forces between metallic tip and semiconductor surfaces”. *Journal of Applied Physics* 77.7 (**1995**). DOI: 10.1063/1.358616.