Antidot Superlattices in InAs–GaSb Double Heterostructures: Transport Studies

A thesis submitted for the degree of
Doctor of Philosophy

Antidot Superlattices in InAs–GaSb Double Heterostructures: Transport Studies

Tim Oliver Stadelmann

Trinity Term 2006

University College, Oxford

A thesis submitted for the degree of Doctor of Philosophy

Antidot Superlattices in InAs–GaSb Double Heterostructures: Transport Studies

Tim Oliver Stadelmann

Trinity Term 2006
University College, Oxford

Abstract

In InAs–GaSb heterostructures, the InAs conduction band lies below the GaSb valence band. Mobile sheets of interacting electrons and holes are formed by charge transfer in the InAs and GaSb layers, respectively. The sheet densities are sensitive to the Fermi level pinning at a nearby surface

Electronic transport properties of carrier sheets in InAs–GaSb double heterostructures are studied experimentally in the presence of a magnetic field at low temperatures. Specifically, the effect of periodic potential modulations on the magnetotransport properties is investigated.

Local anodic oxidation with an atomic force microscope is developed as a means for creating such potential modulations. The formation of periodic oxide structures with lattice constants down to 75 nm is reported; the sensitivity to the probe condition and the shallow modification depth are noted as important limitations. The creation of an intermediate aluminium mask for further etching is shown as a strategy for achieving deeper structures.

Potential modulations with periods down to 200 nm fabricated by direct oxidation of shallow double heterostructures exhibit no significant effect on the magnetoresistance. Etched antidot lattices with periods from 400 to 750 nm reveal commensurability features. The results are compared to theoretical single carrier predictions, and the prominent peaks are attributed to electron-like magnetic breakdown orbits. Additional features including the behaviour in a tilted field are discussed in the light of the electron–hole interaction.

Supplementary magnetotransport studies on structures containing two InAs layers show that the two wells behave independently for well distances above 100 Å. For smaller barrier widths, there are circumstantial signs for a change in the hole levels in the barrier.

Acknowledgements

This thesis is dedicated to my parents, Elke and Hans-Joachim Stadelmann, whose support and encouragement have been vital to my success.

With a sense of profound gratitude I should moreover like to express my thanks to:

The Department of Physics of the University of Oxford and the Clarendon Laboratory for extending to me the use of their facilities.

My supervisor, Robin Nicholas, whose optimism and enthusiasm has been a great inspiration, for his many fascinating ideas and his support in circumnavigating the pitfalls of academic research.

Nigel Mason and Philip Shields for providing me with epitaxial layers grown in the MOVPE reactor, which are at the heart of the research presented here.

Beata Kardynał and Javed Rehman for setting the stage for the work on antidot patterns.

Geb Jones for his invaluable help with electron beam exposures and Michael Pepper's research group in the Cavendish Laboratory in Cambridge for providing the facilities for this work.

My research colleagues in Oxford, Cattleya Petchsingh, Victoria Broadley, Kei Takashina, Rizwan Chughtai, Gavin Conibeer, Praphaphan Wipatawit, Christopher Bumby, Joseph Wiltshire, Lance Lee, James Doig, Ian Mortimer, Benjaporn Nedniyom, and Adrian Nish for making my time in Oxford a pleasant and memorable experience. Special thanks are due to Russell Deacon for his help with printing the thesis.

Finally, I wish to acknowledge the financial backing I received from the Engineering and Physical Sciences Research Council under award number 00309184 and the support of the staff of the Clarendon Laboratory, without whom this work would not have been possible.

Contents

List of Figures

List of Tables

1 Introduction

> They wanted facts. Facts! They demanded facts from him, as if facts could explain anything.
>
> *(Joseph Conrad, Lord Jim)*

1.1 Motivation

The effect of artificial periodic potential modulations on quasi-two-dimensional carrier systems has raised considerable interest since the realization of this idea in physical devices has been demonstrated by Weiss *et al.* in 1989 [1]. The modulations introduce an additional length scale, the modulation period a, into the system, which is already characterized by several intrinsic dimensions. In the simple case of a nearly free electron gas with effective mass m^* and areal density n_e, these are the Fermi wave length $\lambda_F = 2\pi/k_F$, where $k_F = \sqrt{2\pi n_e}$ is the wave number at the Fermi energy E_F, on the one hand and the mean free path $\ell_f = v_F\tau$, where τ is the relaxation time[1] and $v_F = \sqrt{2E_F/m^*} = \hbar k_F/m^*$ the Fermi velocity, on the other hand. In a magnetic field B_z perpendicular to the plane of the electron sheet, the magnetic length $\ell_m(B_z) = \sqrt{\hbar/eB_z}$ and the cyclotron radius $R_c(B_z) = \ell_m^2 k_F = v_F m^*/eB_z$ also become important. Similar considerations apply to more realistic models.

In high-mobility quasi-two-dimensional electron gases (2DEGs) that appear in real semiconductor structures, the Fermi wave length is typically measured in tens of nanometres,

[1] In general, several distinguishable scattering mechanisms can be relevant.

while the mean free path at liquid helium temperatures can be several microns long. The diameter of the cyclotron orbits of electrons moving at the Fermi velocity lies in this range for moderate magnetic fields of several tenths of a Tesla to several Teslas, which are readily accessible to experiments. The magnetic length ℓ_m becomes comparable to a at even smaller fields.

Although the lower end of the scale still poses a significant technological challenge, the creation of superlattices with lattice constants of this size has now become feasible, opening up the possibility of studying physical phenomena that result from the commensurability of ℓ_m and R_c with the artificially imposed period a. In contrast, while every metallic solid has a periodic potential set up by the atomic nuclei, commensurability effects are not observable since the small inter-atomic distances of a few Ångström and large electron densities mean that a magnetic field of tens of thousands of Teslas would be required to achieve a ℓ_m or R_c comparable to the lattice spacing [2].

The original motivation for the interest in such systems centred on the commensurability of a and the magnetic length ℓ_m and was sparked by Hofstadter's 1976 paper [2]. He pointed out that contemporary theoretical work on the problem implied that the energy spectrum of the electron gas forms a fractal structure, which will be revisited in the next section. Since observation depends on resolving the internal structure of Landau levels, and hence on small thermal and impurity broadening as well as on a strong magnetic field, which implies small a, experimental traces of this intriguing spectrum have been hard to come by and significant progress has only recently been made [3].

After its discovery by Weiss *et al.* [1, 4], the alteration of the transport properties caused by the commensurability of the lattice period and the cyclotron radius has instead been at the centre of attention. Although a quantum mechanical treatment is required to fully explain the observed behaviour, the effect is classical in its origin and results from the distortion of ballistic electron trajectories by the superlattice potential. Depending on the relation between a and R_c (and hence B_z), electronic conductivity is enhanced or suppressed; for sufficiently strong modulations—or antidots—a large proportion of the electron orbits be-

comes chaotic. These phenomena have mostly been investigated in 2DEGs formed in GaAs–$Al_xGa_{1-x}As$ heterostructures and are now quite well understood, although theoretical approaches to strongly modulated systems have largely been limited to numerical simulations. Chaotic electron dynamics in periodic superlattices are still a focus of current research, comprising topics such as directional transport in arrays with broken symmetry connected to an external energy source [5–7].

In InAs–GaSb double heterostructures (DHETs), a 2DEG forms in the InAs layer as a result of charge transfer. Our research group in Oxford has developed metal-organic vapour phase epitaxy (MOVPE) techniques to grow such devices with consistently high concentrations of mobile holes approaching that of the electrons [8, 9]. The motivation behind the research leading to the present thesis has been to study the expression of the semiclassical commensurability phenomena in a more complex system consisting of quasi-two-dimensional gases of both electrons and electron holes. The opportunity to test the generality of previous observations by looking at a different material system and the absence of a depletion layer at the free surface of InAs—which opens up the possibility of smaller effective antidot diameters [10]—have been additional considerations.

1.2 Hofstadter's Butterfly

In the presence of a periodic potential $V(\mathbf{r}) = V(\mathbf{r}+\mathbf{R})$, where $\mathbf{R}$ is a lattice vector, the wave function $\varphi(\mathbf{r})$ of an electron in a magnetic field $\mathbf{B}$ obeys the time-independent Schrödinger equation

$$\hat{H}\varphi(\mathbf{r}) = \frac{[\hat{\mathbf{p}} + e\mathbf{A}(\mathbf{r})]^2}{2m}\varphi(\mathbf{r}) + V(\mathbf{r})\varphi(\mathbf{r}) = E\varphi(\mathbf{r}), \tag{1.1}$$

where $\hat{\mathbf{p}} = -i\hbar\,\nabla_{\mathbf{r}}$ is the momentum operator and $\mathbf{A}(\mathbf{r})$ the magnetic vector potential with $\mathbf{B} = \nabla_{\mathbf{r}} \times \mathbf{A}$. For $\mathbf{A}(\mathbf{r}) = \mathbf{0}$, the eigenstates of $\hat{H}$ are of the Bloch form $\varphi_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r}) \exp(i\mathbf{k}\cdot\mathbf{r})$, where $\mathbf{k}$ is the crystal momentum and $u_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r}+\mathbf{R})$; their eigenvalues are the Bloch bands $E_n(\mathbf{k})$ with band index n.

The energy spectrum of Eq. (1.1) may be obtained by solving the Schrödinger-like equation

$$\hat{\hat{H}}\bar{\varphi}(\mathbf{r}) = E\bar{\varphi}(\mathbf{r}), \tag{1.2}$$

where the *effective Hamiltonian* $\hat{\hat{H}}$ to first order in the magnetic field is $E_n(\hat{\boldsymbol{\kappa}})$ with $\hbar\hat{\boldsymbol{\kappa}} = \hat{\mathbf{p}} + e\mathbf{A}(\mathbf{r})$ [11]. For a two-dimensional potential with $\mathbf{R} = a(n_x, n_y)$, where $n_x, n_y \in \mathbb{Z}^+$, the Bloch bands can be approximated by $E_n(k_x, k_y) = E_n^{(0)} + E_n^{(1)}(\cos k_x a + \cos k_y a)$, where $E_n^{(0)}$ and $E_n^{(1)}$ are empirical parameters.[2] Eq. (1.2) then becomes

$$E_n^{(0)}\bar{\varphi}(x,y) + \frac{E_n^{(1)}}{2}\Big[\bar{\varphi}(x+a,y) + \bar{\varphi}(x-a,y) + e^{-ieB_z x/\hbar}\bar{\varphi}(x,y+a) + e^{ieB_z x/\hbar}\bar{\varphi}(x,y-a)\Big] = E\bar{\varphi}(x,y) \tag{1.3}$$

for a single band n, where the Landau gauge $\nabla_{\mathbf{r}} \cdot \mathbf{A}(\mathbf{r}) = 0$ with $\mathbf{A}(\mathbf{r}) = (0, B_z x, 0)$ has been used. By defining $n \overset{\text{def}}{=} x/a$, $\nu \overset{\text{def}}{=} k_y a$, and $\varepsilon \overset{\text{def}}{=} 2(E - E_n^{(0)})/E_n^{(1)}$, making the *Ansatz* $\bar{\varphi}(x,y) = \exp(i\nu y/a)g_n$, and introducing the dimensionless parameter $\alpha \overset{\text{def}}{=} eB_z a^2/(2\pi\hbar)$, one can simplify Eq. (1.3) to Harper's equation [2, 13]:

$$g_{n+1} + g_{n-1} + 2\cos(2\pi n\alpha - \nu)g_n = \varepsilon g_n. \tag{1.4}$$

The reduced magnetic field $\alpha = (B_z a^2)\big/(h/e) = \Phi/\Phi_0^{(D)}$ is the ratio of the magnetic flux per unit cell $\Phi = B_z a^2$ to the Dirac flux quantum[3] $\Phi_0^{(D)} = h/e$; it can also be expressed in terms of the magnetic length ℓ_{m} as $\alpha = (a/\ell_{\mathrm{m}})^2\big/2\pi$.

The eigenvalue spectrum ε_α of the difference equation (1.4) is the set of all ε for a given α such that for some value of ν there is a g_n (and hence $\bar{\varphi}(x)$ and $\varphi(x)$) that is bounded for all n. It has the peculiar property that it depends on the *rationality* of α, *i.e.*, whether $\alpha = p/q$ for some prime numbers p and q. In this case, there are q distinct energy bands, whereas for irrational α the spectrum consists of infinitely many isolated values. In his famous 1976 paper [2], Hofstadter points out that the union of all ε_α (see Fig. 1.1) forms a self-similar fractal set [14], which has since become known as Hofstadter's butterfly. Its graph is periodic

[2] See, *e.g.*, Kittel's book [12] for a derivation of $E_n(\mathbf{k})$ in the Kronig-Penney framework, from which this semiempirical 'Harper model' may be justified.

[3] Not to be confused with the superconducting flux quantum $\Phi_0^{(s)} = h/2e$.

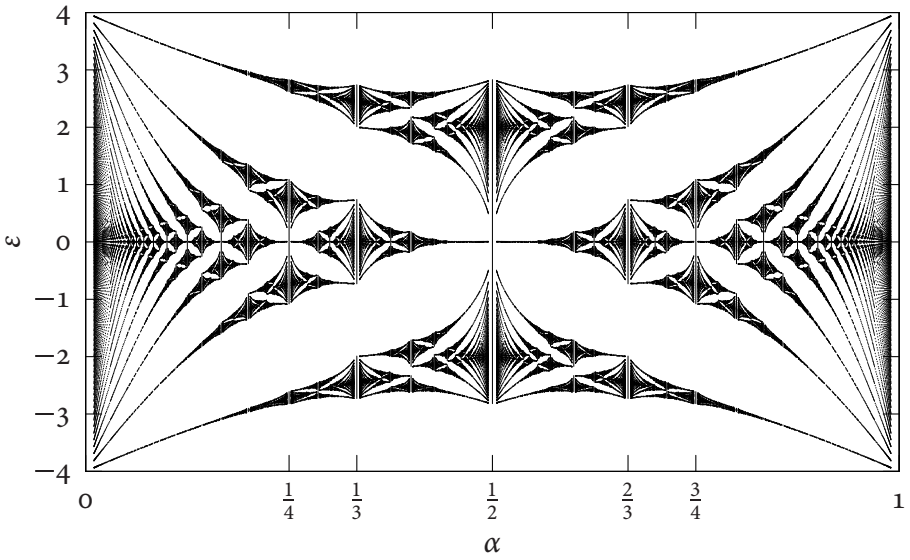


Figure 1.1: Hofstadter's butterfly

in α with period 1—so that the pattern repeats for each additional flux quantum per unit cell—and the unit interval $[0, 1)$ exhibits reflection symmetry in the lines $\varepsilon_\alpha = 0$ and $\alpha = \frac{1}{2}$. At low α, the bands group into clusters which can be identified with the familiar Landau fan that is predicted in the absence of a periodic potential. As is characteristic of fractals arising from mathematical descriptions of physical systems, the unphysical property of being discontinuous everywhere is a formal consequence of the idealized theoretical representation. If a finite uncertainty in B_z is introduced [2], or if a finite size system is considered [15], continuity is recovered, and the scale at which the fractal is resolved depends on the particular model.

Interestingly, Harper's equation (1.4) *also* emerges from first-order perturbation theory in the complementary limit of a weak potential modulation of the form $V(x, y) = V_0(\cos Kx + \cos Ky)$, with $K \stackrel{\text{def}}{=} 2\pi/a$, perturbing a Landau quantized system [16, 17].[4] In this case, the parameters $\alpha \stackrel{\text{def}}{=} \Phi_0^{(D)}/\Phi$ and $\nu \stackrel{\text{def}}{=} -k_y \ell_{\text{m}}^2/(2\pi a)$ have a different meaning—so that the energy spectrum is periodic in $1/B_z$—and Hofstadter's butterfly describes the internal structure of a single Landau level. Hofstadter [2] points out that such a mapping is consistent with

[4] Landau quantization is discussed in Sec. 5.6.2

the scaling properties of the complete pattern. However, the range of validity is different for both approaches and the exact reappearance of the same equation is an artefact of the peculiar choices for $E(k_x, k_y)$ and $V(x, y)$.

1.3 Development

The initial work on the InAs–GaSb DHETs was done using the electron beam lithography (EBL) facilities available to our group in Oxford. These consisted of a converted scanning electron microscope (SEM) with a resolution not much better than 0·5 µm, and it became quickly clear that the lithographical performance of this setup was severely limiting the experiments that could be accomplished. This was especially true as the high density of the required antidot patterns implied a significant contrast reduction due to exposure from secondary electrons. Additionally, the poor selectivity of the applicable etch chemistries for the semiconductor materials over EBL resists meant that the success rate at the intended etch depths was low and could only be improved by switching to a substantially more complicated technique using an intermediate mask.

As the upgrade of the EBL equipment accessible to me in Oxford or through our collaborators was delayed, and a new atomic force microscope (AFM) had become available to our group that could be modified for such work, I began to investigate the option of patterning the DHETs by local anodic oxidation (LAO) using this instrument. While only shallow surface modifications were possible with this approach, preliminary experiments had shown that removal of only a part of the thin GaSb layer forming the uppermost part of the DHETs was enough to impose a significant potential on the carrier gases. Moreover, the technique promised a much improved resolution and greater flexibility in device fabrication.

After LAO performance, which is mostly limited by poor reproducibility, had been optimized to the point where patterned DHETs as originally envisaged could be created, it became apparent that the initial results had been misleading. Under a wide range of conditions, the modulation potential set up by the surface corrugation was insufficient to prompt a meas-

urable effect. In response to this, I modified the technique to allow for deeper modifications by patterning an intermediate etch mask instead of the semiconductor surface. In the meantime, improved EBL facilities had become available in the Cavendish Laboratory in Cambridge and thanks to the help of GEB JONES these could be used to create very similar samples, which formed the basis of much of the work reported in Chapter 7.

Throughout the evolution of this research I worked closely with NIGEL MASON and PHILIP SHIELDS on the growth of InAs–GaSb DHETS by MOVPE in the now-defunct atmospheric pressure reactor in the Clarendon Laboratory. In the context of this work, I undertook measurements on a large number of DHETS both for growth assessment and to improve our understanding of the fundamental properties of these systems, including for the first time a systematic investigation of the magnetotransport properties of structures containing two InAs wells.

1.4 Thesis Structure

The remainder of this thesis can roughly be divided in two parts: Chapters 2, 3, and 4 focus on the engineering aspects, especially the work on lithographical techniques using a scanning probe microscope, while Chapters 5, 6, and 7 are concerned with the physics of InAs–GaSb double heterostructures, paying special attention to lateral potential modulations in the presence of a perpendicular magnetic field. Each part starts with a chapter that sets the scene by providing the essential background information.

In the case of the first part, this is Chapter 2, which describes the operation and the limitations of the atomic force microscope and outlines how the special capabilities of this instrument can be applied to surface manipulation. As the issues are quite general, and to allow for the straightforward comparison with similar lithographic approaches, I shall approach the topic from the point of view of generic scanning probe microscopes and introduce the AFM as an important application of these ideas. Chapter 3 builds on this basis in discussing my work on the modification of GaSb, InAs, and Al surfaces by local anodization with an

AFM. How this technique may be integrated in a fabrication scheme for antidot samples or similar devices is explained in Chapter 4, which also covers the details of the experimental procedures used in the present work.

The second part starts with a an overview of the basic physics of InAs–GaSb DHETs in Chapter 5. In the same chapter, the growth of these structures in our research group in Oxford is surveyed, and the individual DHETs studied in the remainder of the thesis are introduced. Chapter 6 reports studies of a series of substrates containing double InAs wells, relating them to previous work on superlattices. Lateral antidot lattices modulating InAs–GaSb DHETs are finally investigated in Chapter 7, exploring the possibility of a contribution to the observed commensurability effects from the presence of a mobile hole layer.

Chapter 8 concludes the thesis with a summary of the most important results and attempts to assess the lessons learnt in the conduct of this research. What follows is a series of appendices containing material that was excluded from the above chapters so as not to distract from the main exposition, but which I consider none the less important for substantiating some of the claims made in the main text.

Bibliography

[1] D. Weiss, K. von Klitzing, K. Ploog, and G. Weimann. Magnetoresistance oscillations in a two-dimensional electron gas induced by a submicrometer periodic potential. *Europhysics Letters*, 8(2):179–184, January 1989.

[2] D. R. Hofstadter. Energy levels and wave functions of Bloch electrons in rational and irrational magnetic fields. *Physical Review B*, 14(6):2239–2249, September 1976.

[3] C. Albrecht, J. H. Smet, K. von Klitzing, D. Weiss, V. Umansky, and H. Schweizer. Evidence of Hofstadter's fractal energy spectrum in the quantized Hall conductance. *Physical Review Letters*, 86(1):147–150, January 2002.

[4] D. Weiss, M. L. Roukes, A. Menschig, P. Grambow, K. von Klitzing, and G. Weimann. Electron pinball and commensurate orbits in a periodic array of scatterers. *Physical Review Letters*, 66(21):2790–2793, May 1991.

[5] A. Lorke, S. Wimmer, B. Jager, J. P. Kotthaus, W. Wegscheider, and M. Bichler. Far-infrared and transport properties of antidot arrays with broken symmetry. *Physica B*, 249–251:312–316, 1998.

[6] A. D. Chepelianskii and D. L. Shepeliansky. Directing transport by polarized radiation in presence of chaos and dissipation. *Physical Review B*, 71:05208, February 2005.

[7] G. G. Carlo, G. Benenti, G. Casati, and D. L. Shepelyansky. Quantum ratchets in dissipative quantum systems. *Physical Review Letters*, 94:164101, April 2005.

[8] A. Aardvark, N. J. Mason, and P. J. Walker. The growth of antimonides by MOVPE. *Progress in Crystal Growth and Characterization*, 35(2–4):207–241, 1997.

[9] G. R. Booker, M. Daly, P. C. Klipstein, M. Lakrimi, T. F. Kuech, J. Li S. G. Lyapin, N. J. Mason, I. J. Murgatroyd, J.-C. Portal, R. J. Nicholas, D. M. Symons, P. Vicente, and P. J. Walker. Growth of InAs/GaSb strained layer superlattices by MOVPE: III. Use of UV absorption to monitor alkyl stability in the reactor. *Journal of Crystal Growth*, 170:777–782, 1997.

[10] J. Eroms, M. Zitzlsperger, and D. Weiss. Skipping orbits and enhanced resistivity in large-diameter InAs/GaSb antidot lattices. *Physical Review B*, 59(12):7829–7832, March 1999.

[11] W. Kohn. Theory of Bloch electrons in a magnetic field: The effective Hamiltonian. *Physical Review*, 115(6):1460–1478, September 1959.

[12] C. Kittel. *Introduction to Solid State Physics*, chapter 7. John Wiley & Sons, New York, seventh edition, 1996.

[13] P. G. Harper. Single band motion of conduction electrons in a uniform magnetic field. *Proceedings of the Physical Society. Section A*, 68(10):874–878, October 1955.

[14] K. Falconer. *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley & Sons Ltd., Baffins Lane, Chichester, West Sussex, England, 1990.

[15] J. G. Analytis, S. J. Blundell, and A. Ardavan. Landau levels, molecular orbitals, and the Hofstadter butterfly in finite systems. *American Journal of Physics*, 72(5):613–618, May 2004.

[16] D. Langbein. The tight-binding and the nearly-free-electron approach to lattice electrons in external magnetic fields. *Physical Review*, 180(3):633–648, April 1969.

[17] D. Pfannkuche and R. R. Gerhardts. Theory of magnetotransport in two-dimensional electron systems subjected to weak two-dimensional lattice potentials. *Physical Review B*, 46(19):12606–12626, November 1992.

2 Review of Scanning Probe Microscopy Techniques

> Νόμῳ γάρ φησι γλυκὺ [καὶ] νόμῳ πικρόν,
> νόμῳ ψυχρόν, νόμῳ χροιή, ἐτεῇ δὲ ἄτομα καὶ
> κενόν.
>
> *(Democritus, fragment 9)*

2.1 Introduction

2.1.1 Motivation

The atomic force microscope (AFM) plays an important part in the research covered in this thesis. I investigate local anodic oxidation (LAO)—a technology based on the AFM—as one approach for fabricating the antidot patterns studied in Chapter 7. Its adaption to the materials and the patterns at the focus of this work is expounded in Chapter 3. Chapters 4 and 5 also use AFM micrographs for evaluation purposes. To apprehend the challenges of AFM-based lithography and correctly interpret AFM images a thorough understanding of the capabilities and limitations of the instrument is indispensable.

Collecting the background information in this chapter allows me to concentrate on the subject matter later on. The reader will hopefully find it a convenient place to look up detailed information; those who are intimately familiar with the theory and application of scanning probe microscopes (SPMs) are free to skip it on first reading. Although I have stressed the common features of all SPMs wherever possible, the focus is by necessity on the AFM. I have also taken the opportunity to round off the account by briefly covering topics such as the near

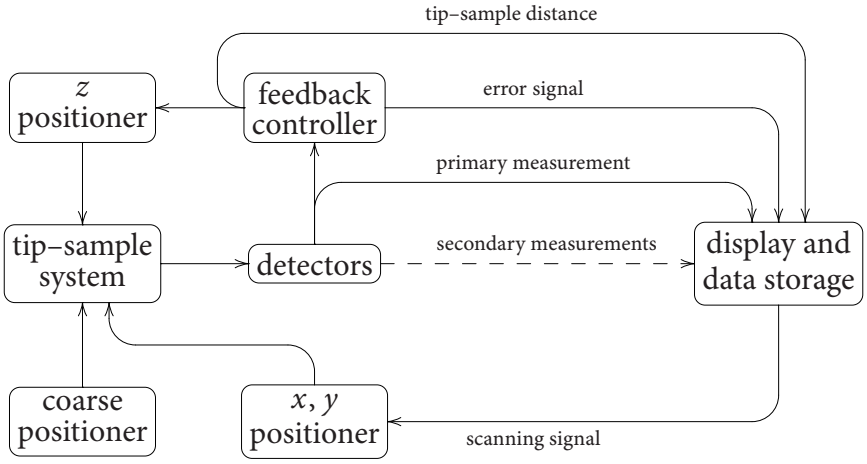


Figure 2.1: Schematic diagram of an SPM

field scanning optical microscope (NFSOM, Sec. 2.2.3) and the question of atomic resolution in scanning tunnelling microscopes (STMs) and AFMs (Secs. 2.2.1 and 2.4.6). These issues are not directly pertinent to this thesis but help to illustrate the driving forces behind the development of the SPM and the wide applicability of the principle. As such this chapter may be read as a general introduction to the subject.

2.1.2 A Simple Idea

The fundamental principle of all scanning probe microscopes is the use of the interaction between a sharp tip and the surface of a sample to measure its *local* physical properties. Fig. 2.1 provides a schematic view of the interactions between the fundamental components of a generalized SPM. A map of the specimen is build by sweeping the tip across its surface scan line by scan line with a two-dimensional actuator or scanner (*cf.* Sec. 2.3). The scanner should ideally be able to control the relative position of the tip to within the resolution limit imposed by the interaction; for atomic resolution this implies a precision of 1 Å or better. During this scanning process, the tip–sample interaction can be recorded directly, or, more commonly, a feedback loop keeps one parameter at a set point by varying the tip–sample distance. The correction to the distance is then used to form an image; at the same time, other surface properties can be measured.

In addition to the scanner, an SPM typically requires a mechanism for coarse positioning to bring the sample within the range of movement provided by the scanner, and to move the probe to different areas of the sample [1]. The accuracy of the positioner must be high enough to overlap the range of motion of the scanner—typically this translates to a resolution better than 1 µm in the z-direction, and several µm in the x, y-plane. The required range of movement depends on the size of the instrument and the sample and may vary from several mm to several cm. On the timescale of the measurement, the stability of the positioner must generally be within the ultimate resolution of the instrument.[1]

2.1.3 The Development of the Scanning Probe Microscope

The Stylus Profilometer

The idea of using a scanning probe to visualize the roughness of a surface is actually quite old. As early as 1929, SCHMALZ [2] developed an instrument that had much in common with the modern AFM: the stylus profilometer. A probe is lightly pressed against the surface by a leaf spring and moved across it; a light beam is reflected off the probe and its projection on a photographic emulsion exposes a magnified profile of the surface, using the optical lever technique (*cf.* Sec 2.4.3). The fundamental difference between these instruments and modern AFMs is the attainable resolution, which is limited by the relatively blunt stylus, the scanning and detection mechanism, and thermal and acoustical noise.

The Topographiner

The STM, which started off the development of SPMs, has its roots in the 'topographiner' advanced by YOUNG in 1971 [3, 4]. This non-contact profiler uses the current between a conducting tip and sample to sense the proximity of the surface. It already used a feedback circuit to keep the working distance constant; the use of piezoelectric positioners is another

[1] Drift in the z-direction may be corrected by the main feedback loop depending on the operating mode. Drift in the x, y-plane will lead to systematic distortion of the image. Noise may be reduced by mechanically decoupling the scanner and the sample from the coarse positioner.

feature it shares with most modern SPMs. Unlike the STM, which places the tip close to the sample and uses direct tunnelling, it operates in the Fowler-Nordheim field emission regime (*cf.* Sec 2.2.1). Because of this and insufficient isolation from external noise it only achieves a resolution comparable to that of optical microscopes [5].

Tunnelling Experiments

YOUNG already used his topographiner to perform spectroscopic experiments in the direct tunnelling regime and demonstrated the strong dependence of the current on the distance, but could not achieve stable imaging under these conditions [3]. Similarly, the work by GERD BINNIG and HEINRICH ROHRER, which should lead to the development of the STM, was originally centred around local spectroscopy of thin films. The idea was to use vacuum tunnelling as a means to probe the surface properties [5].

The First Scanning Tunnelling Microscope

The fundamental achievement of BINNIG and ROHRER, which was honoured with the Nobel prize in Physics in 1986,[2] was to realize that the exponential distance dependence of the tunnel current would enable true atomic resolution and to put the pieces of the puzzle together in building an microscope, the STM, that would make this vision reality [5, 6]. Unlike its predecessor, it could produce images in the direct tunnelling regime and had an improved vibration-isolation system, which in the first prototype used magnetic levitation of a superconducting lead bowl [5].

Further Developments

Since the SPM was popularized by the work of BINNIG and ROHRER in the early 1980s, the principle has been applied to a wide range of problems. This includes the scanning force microscope (SFM) invented by GERD BINNIG, CALVIN QUATE, and CHRISTOPH GERBER in

[2] Together with ERNST RUSKA, who was awarded the other half of the prize for the invention of the electron microscope.

1985 [7] (Sec. 2.4), the dynamic force microscope (DFM), which evolved as a refinement of the SFM following the work of YVES MARTIN and KUMAR WICKRAMASINGHE [8] (Sec. 2.4.5), and the various approaches to NFSOM [9,10] (Sec. 2.2.3). The SFM in particular has become an enabling technology for several measurement (Sec. 2.2.2) and sample modification (Sec. 2.5) applications at the nanometre scale and below.

2.2 Applications of the SPM design

2.2.1 The Scanning Tunnelling Microscope

Mechanism

The STM measures the tunnel current through the gap between a sharp tip and a conducting sample surface while the tip is scanned across the surface. Although the current can be recorded directly, it is more common to keep it constant and build an image from the z-axis correction signal [5, 6, 11, 12]. The STM can operate in air or any non-conducting fluid, but optimal resolution may require ultra high vacuum (UHV) and low temperatures.

In the low voltage limit the tunnel current between two metal surfaces with average work function $\overline{\varphi}$ at distance z is of the form

$$I_t \propto \sqrt{\overline{\varphi}} V \exp\left(-2\sqrt{2m_e/\hbar^2}\sqrt{\overline{\varphi}} z\right), \tag{2.1}$$

where V is the bias voltage and m_e the electron mass [3, 5, 6, 13]. As only electrons close to the Fermi energy can participate in conduction, the magnitude of the current also depends on the density of states at the Fermi level in both tip and sample.

Because of the exponential dependence on the distance, the experiment is very sensitive to variations in the topography of the sample: a step of one atomic diameter causes the tunnel current to change by 2 to 3 orders of magnitude if one assumes an average work function of order 1 eV [6]. Even more importantly, the exponential decay implies that only a few atoms at the apex of the tip significantly contribute to the current, so that the effective resolution of

the instrument is much better than the sharpness of the tip suggests. It is dominated by the size of the microasperities or microtips at the front of the macroscopic tip. The macroscopic radius of curvature only matters in so far as it determines together with the surface roughness the likelihood of exactly one microtip coming within critical distance of the surface [5, 6].

If the sample work function and the density of states are constant, the error signal from the feedback loop represents the topography of the sample under investigation. In practice these parameters can and will change as the tip scans across the surface, and the data is convolved with information about its electrical properties [6]. In order to correctly interpret the resulting images at atomic resolution a detailed theoretical model of the tip–sample interaction is needed [11, 12, 14].

The Tunnel Current

In the planar approximation, the exponential behaviour of the tunnel current in the low voltage limit can be understood by solving the one-dimensional Schrödinger equation for a barrier of finite height and considering the transmittivity for electrons at the Fermi level [13].[3] Eq. (2.1) is actually derived from Simmons' more general expression for the tunnel effect,

$$J_t = \frac{e}{\hbar(2\pi\beta z)^2}\left[\overline{\varphi}e^{-A\sqrt{\overline{\varphi}}} - (\overline{\varphi}+eV)\,e^{-A\sqrt{\overline{\varphi}-eV}}\right], \tag{2.2}$$

where J_t is the tunnel current density, $A = 2\beta z\sqrt{2m_e/\hbar^2}$, and $\beta \approx 1$ for small eV [3, 15]. The approximation is valid for $eV \ll \overline{\varphi}$, while the high voltage limit corresponds to the Fowler-Nordheim field emission regime. These calculations played an important role in the initial development of the STM [5, 6], but they cannot accurately predict contrast on an atomic scale.

In 1983 Tersoff and Hamann [14] introduced an approach based on first-order perturbation theory that takes into account the electronic structure of the sample; the tip is approximated by a single *s*-type wave function. It predicts a tunnel current proportional to the sample density of states at the Fermi energy evaluated at the centre of the tip orbital and is still used

[3] Gasiorowicz introduces the STM in the context of cold emission. This may be misleading, as this instrument operates in the direct tunnelling regime, which is precisely what differentiates it from Young's topographiner.

extensively for simulating STM images [12]. A different perturbation theory approach, which is based on the work of BARDEEN, allows for a more realistic tip model and uses transfer matrices to calculate the current between the two systems. Other methods of note for calculating the tunnel current on the basis of realistic molecular models include: Scattering theory using, *e.g.*, the Landauer-Büttiker formula; the non-equilibrium Green function technique based on the KELDYSH formalism, which allows for different chemical potentials in the tip and the sample and has attracted some interest in recent years [11, 12].

While these methods give realistic results for semiconductor surfaces, they severely underestimate the contrast for free-electron-like metals. Dynamic models, which take the deformation of the surface and tip crystal structures into account and consider excited electronic states, give better agreement with experiment at the expense of higher computational cost [11].

True Atomic Resolution

One of the main motivations behind the development of the STM is its ability to achieve true atomic resolution [5]. A measurement of a periodic structure with the correct symmetry and period may represent the envelope convolution of a complicated probe with the surface (SPMs with a blunt probe) or averaging over many layers (transmission electron microscope, TEM). It does not demonstrate the independent observation of individual atoms. Only the imaging of non-periodic structures at the atomic scale, such as defects or adatoms, allows to claim true atomic resolution. The capability of the STM in this regard was already demonstrated in 1982 by BINNIG *et al.* with a prototype operating in UHV by imaging step lines on $CaIrSn_4$ and Si and fully accepted in 1985 when other groups succeeded in obtaining similar results [5].

Today, the best laboratory built STMs have a vertical resolution better than 1 pm, about 1/200 of an atomic diameter [12]. The effective resolution of the instrument is then limited by the configuration of the tip. Unfortunately, the tip geometry and composition is in general not known. While the apex shape can be determined with high accuracy by field ion micro-

scopy [12], there is no practical way to determine the species and arrangement of the crucial atoms at the very end of the probe. Moreover, this configuration may change during normal operation as the tip is deformed and picks up atoms from the surface. In fact, it is normal practice to repeatedly bring the tip into contact with the sample until a tip configuration is created that has only one significant microtip with orbitals giving optimal contrast [12]. As a result, our understanding of the mechanisms leading to contrast at the atomic scale is still limited, and the interpretation of images is often ambiguous and may require careful comparison with simulations assuming different tips.

2.2.2 The Scanning Force Microscope

Principle

An SFM uses the deflection or resonant frequency of a cantilever or leaf spring to measure the force between a probe attached to its end—usually a sharp tip—and the sample under investigation. In analogy to the tunnel current in the STM, the force can be recorded directly or kept constant by means of a feedback loop. The operation of the SFM is discussed in more detail in Section 2.4.

The Atomic Force Microscope

The archetypal SFM is the AFM invented by BINNIG, QUATE, and GERBER in 1985 [7], which uses the repulsive force between a sample and a sharp tip pressed against it and measures the sample topography. Unlike the STM, it is not limited to conducting samples.

Other Forces and Functionalization

The operational principle of the SFM is quite general and there are many forces that can be measured. An AFM may also use the attractive force felt by a tip close to the sample surface and the lateral force that results from friction as the tip is scanned across the surface. Moreover, the tip can be functionalized so that magnetic (magnetic force microscope, MFM),

electrostatic, or chemical forces can be detected. Since the SFM provides the capability of scanning a probe across a surface with great accuracy without placing many constraints on the properties of the tip or the sample it also forms the basis for other measurement and lithography techniques. This includes local capacitance (scanning capacitance microscope, SCM) and conductivity measurements as well as mechanical and electrochemical (see Chapter 3) sample modification methods.

2.2.3 The Near Field Scanning Optical Microscope

Conception

The NFSOM provides a way of probing the optical properties of a surface at length scales smaller than the optical wave length [9,10,16,17]. It avoids the diffraction limit is by using a light source or detector whose size and distance from the sample is much less than the wave length so that the interaction is dominated by the near field. The resolution is then limited by the effective aperture size.

The idea of utilizing the near field zone of an aperture in combination with scanning to form an image with sub-wave-length resolution was already proposed by SYNGE in 1928 [17] and accomplished with microwaves by ASH *et al.* in 1972 [9,17]. Inspired by the contemporary work on STM, POHL *et al.* built the first first working NFSOM using wave lengths in the visible range from 1982 to 1983.

Requirements

The main challenges in realizing a NFSOM are to achieve enough brightness with a small aperture to allow for reliable detection and to bring the aperture close enough to the surface of the sample [17]. While the former has become feasible in the 1980s because of advances in laser and detector technology, the latter was made possible by placing the aperture at the apex of a sharp dielectric tip. A probe of this kind can be manufactured by etching a material

such as quartz to a sharp point, covering the resulting tip with a metal layer, and opening a hole at the apex by mechanical or chemical means [9].

Distance Control

Various approaches can be used to control the distance between the probe and the sample. It is possible to simply let the tip touch the sample or a constant distance may be approximately maintained by controlling the tip position using capacitive or interferometric sensing [9]. However, techniques based on STM [9] or DFM [17] have proved most useful as they allow for maintaining a constant gap at the nanometre scale and allow for simultaneous measurement of the surface topography.

Operating Modes

Similar to conventional optical microscopes, NFSOMs can be operated either in transmission or reflection mode. Although it is possible to use the tip for both emission and detection in reflection mode, it is difficult to achieve a useful signal-to-noise ratio with this setup and generally the scanning probe is used either for illumination or measurement: In illumination mode, the tip forms the light source and the detector uses conventional optics, while in collection mode the sample is illuminated broadly and the light coupled into the probe through the aperture is measured [17].

The Photon Scanning Tunnelling Microscope

A variant of collection mode NFSOM is the photon scanning tunnelling microscope (PSTM) introduced by REDDICK in 1988 [10]. Here, an optical evanescent wave is set up at the surface of the sample by total internal reflection. If an optical probe is brought close to the surface, photons tunnel into the probe and can be detected. The field intensity outside the sample and the tunnel current decrease exponentially with increasing distance:

$$I \propto \exp\left(-2kz\sqrt{\sin\theta - (n_t/n_i)^2}\right), \tag{2.3}$$

where k is the wave number, z the distance, θ the incident angle, and n_t and n_i the refractive indices of the gap and the sample. The tunnel signal may be kept constant by a feedback loop that adjusts the distance of the tip to the sample. Operation is then analogous to that of an electron STM, providing a measurement of the surface topography convoluted with the local *optical* properties of the sample [10, 17].

Interpretation of Images

Similar to the situation with STM, it is difficult to formulate simple rules for contrast formation in the near field. In particular, intuitions from conventional far field optics do not carry over directly, and the behaviour depends on the properties of the sample [17]. Correct interpretation of the data requires a detailed theoretical model of the interaction between the tip and the sample surface [16].

2.3 The Scanner

2.3.1 Design

Apart from the detection mechanism, the three-dimensional scanner is the most important component of an SPM. It must be able to control the relative position of the sample and the tip at the intended resolution of the instrument. Its capacity to respond quickly to changes in the set point and hence follow the surface topography limits the scanning rate at which the maximal resolution can be realized. For lithography applications such as the local anodic oxidation discussed in Chapter 3, the ability of the scanner to accurately return to a specific position after covering a potentially large area is essential to ensure that individual parts of large patterns can fit together. The mechanical resonant frequencies of the scanner are also important, since they determine the acceptance of external acoustic noise and restrict the feasible scanning frequencies. To ensure that all resonances occur at high frequencies, the entire scanning assembly must be mechanically stiff.

Most commercial and research SPMs use scanners based on piezoelectric actuators made from ferroelectric materials such as lead zirconate titanate (PZT, $Pb(Zr_xTi_{1-x})O_3$) [1]. These positioners are not affected by backlash or discrete step sizes and offer theoretically unlimited resolution. That the deformation is indeed continuous even for polycrystalline piezoceramics has ultimately only been established by the original STM experiments [5]. In practice, resolution is limited by noise and the finite accuracy of the control electronics. The design of piezoelectric and electromagnetic scanners is explained in detail in Appendix A.

2.3.2 Error Correction Strategies

A piezoelectric actuator in general and an SPM scanner in particular may be operated either in open or closed loop configuration. In open loop mode a mathematical model or a set of calibration data is used to derive the required driving signal from the desired position, while in closed loop mode the actual position of the scanner is measured and used as the input of a feedback loop.

In the open loop configuration, the software can correct for deviations either by processing the finished image (off-line) or by modifying the driving signal of the scanner (on-line). On-line operation ensures uniform resolution over the entire scanning area. Software using a sophisticated theoretical model of the scanner can compensate for most of the nonlinearity and hysteresis, but does not correct for drift and aging [18, 19]. Performance depends on the the accuracy of the model and the repeatability of the physical scanner behaviour.

Closed loop operation requires an accurate measurement of the scanner position. This is routinely achieved using capacitive [18–21] or interferometric [20, 22] sensors. In this mode, performance is generally determined by the accuracy of the sensors [22] and the response time and stability of the feedback loop. An increasing number of SPMs use hardware feedback because it allows the scanner to return to a precise point on the surface, a property that is particularly important for lithography and metrology applications.

The hysteresis of a piezoelectric positioner can also be reduced by using the charge instead of the voltage as the controlling parameter [20, 23]. This approach may be combined with other error compensation strategies, but reduces the effective resolution of the actuator [18].

For scanner designs that do not rely on the linearity of the piezoelectric effect, which changes its sign if the electric field is reversed, electrostrictive[4] materials are a viable alternative [20]. Electrostriction is a property of all dielectrics, which deform in proportion to the square of the external electric field. Because of its strong electrostrictive response, the material most widely used for actuators is lead magnesium niobate (PMN, $Pb(Mg_{1/3}Nb_{2/3})O_3$) [24]. PMN exhibits less creep and hysteresis than PZT, but has a stronger temperature dependence and smaller range for similarly sized actuators. In practice, its use in SPM scanners is limited to applications where control of the hysteresis is crucial, such as force-distance measurements [20].

2.4 A Closer Look at Scanning Force Microscopy

2.4.1 Overview

In this section I shall explain the operation of scanning force microscopes in more detail. Naturally, special attention is given to the atomic force microscope, and the various ways of functionalizing the tip of an SFM or using it as an enabling technology for other microscopy methods are not covered.

SFMs use the deflection or resonant frequency of a cantilever as a measure of the tip–sample interaction. The deflection of the probe must be measured with sufficient accuracy to achieve the desired resolution, and Sec. 2.4.3 compares various approaches to this problem. The dynamic force microscope expounded in Sec. 2.4.5 vibrates its probe, and the frequency response is determined; otherwise, the interaction force is recorded directly in a static force measurement.

[4] From Latin stringere, 'to draw tight'

In Sec. 2.4.6 I shall briefly touch on the capability of AFM for atomic resolution imaging. While not germane to the experiments reported in this thesis, this serves to illustrate the versatility of the technique and puts it in the context of the development of SPMs in general. Finally, in Sec. 2.4.8 the origins of artefacts in AFM images are discussed, which will allow the reader to better understand the AFM images presented in this thesis and follow the discussion in Chapter 3.

2.4.2 The Probe

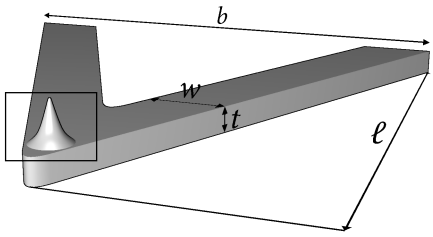
As shown in Fig. 2.2, an SFM probe consists of a sensing tip attached to the end of a flexible cantilever. Nowadays, a large range of probes for different applications is commercially available. The tip is characterized by its shape as well as its electrical, chemical, and mechanical properties. It is manufactured from a crystalline material by mechanical crushing, or, more commonly, chemical etching. The tip angle determines the ability of the probe to follow rough surfaces exhibiting features with high aspect ratios. The tip radius limits the resolution of measurements using long-range forces. Silicon tips with tip angles of approximately 10° and radii of curvature of 20 nm are readily available. Even sharper tips are possible for specialized applications. For electrochemical applications, electronic measurements, and DFM measurements in the presence of a bias voltage, tips must be conducting. This often means heavily doped semiconductors or metal coatings, but for conduction measurements solid metal tips may be required. For mechanical surface modification, diamond tips can be used.

The elastic deformation of the cantilever bearing the tip is used to measure the tip–sample interaction. The deflection z is approximately proportional to the applied force F_{ts}:

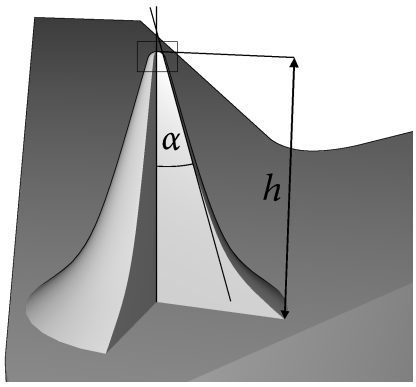
$$F_{\mathrm{ts}} = kz, \tag{2.4}$$

where k is the stiffness of the cantilever. For a rectangular beam,

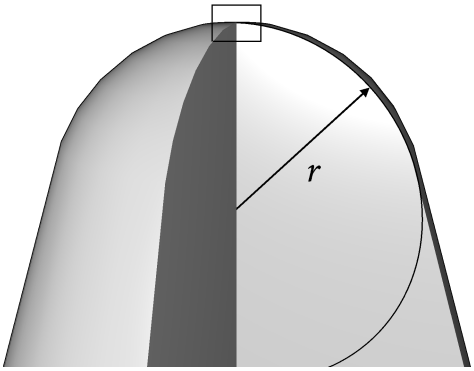
$$k = \frac{Et^3 w}{4(\ell - \Delta\ell)^3}, \tag{2.5}$$



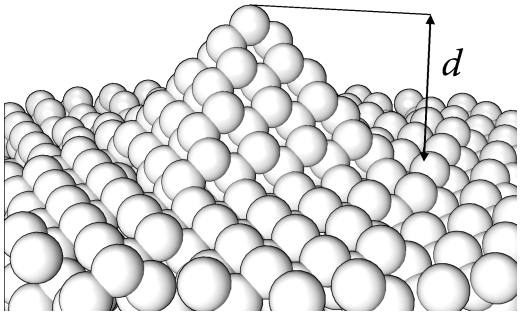
(a) V-shaped cantilever with width w, length ℓ, and thickness t



(b) Tip with height h and angle α



(c) Apex with radius of curvature r



(d) Microtip with height d

Figure 2.2: An atomic force microscopy probe at different magnifications. The thin box shows the location of the image at the next magnification step.

where E is Young's modulus and $\Delta\ell$ the distance of the tip from the end of the cantilever. In practice, a V-shaped cantilever is often used in an attempt to increase the lateral and torsional stiffness, although Sader [25] has shown that this line of reasoning is in fact incorrect and V-shaped cantilevers are *more* susceptible to lateral forces. For such a cantilever,

$$k = \frac{Et^3 w}{2(\ell - \Delta\ell)^3}\left(1 + \frac{4w^3}{b^3}\right)^{-1}. \tag{2.6}$$

The stiffness of the cantilever determines the sensitivity and the resonant frequency. According to Hooke's law (2.4), a stiffer cantilever deflects less for the same force, and therefore has a larger range and reduced sensitivity. The force that can be measured in the repuls-

ive regime is limited by the sample's threshold for inelastic deformation. For typical AFM applications, a cantilever is chosen that is compliant enough to allow for easy detection of forces significantly below this limit. The movement of the cantilever in air or vacuum can be approximated by that of a point mass on a massless spring; the resonant frequency is then

$$\omega_0 = \sqrt{\frac{k}{m^*}}, \tag{2.7}$$

where m^* is the effective mass. This frequency governs the susceptibility of the cantilever to external noise and dictates the approximate frequency at which it must be vibrated in a DFM.

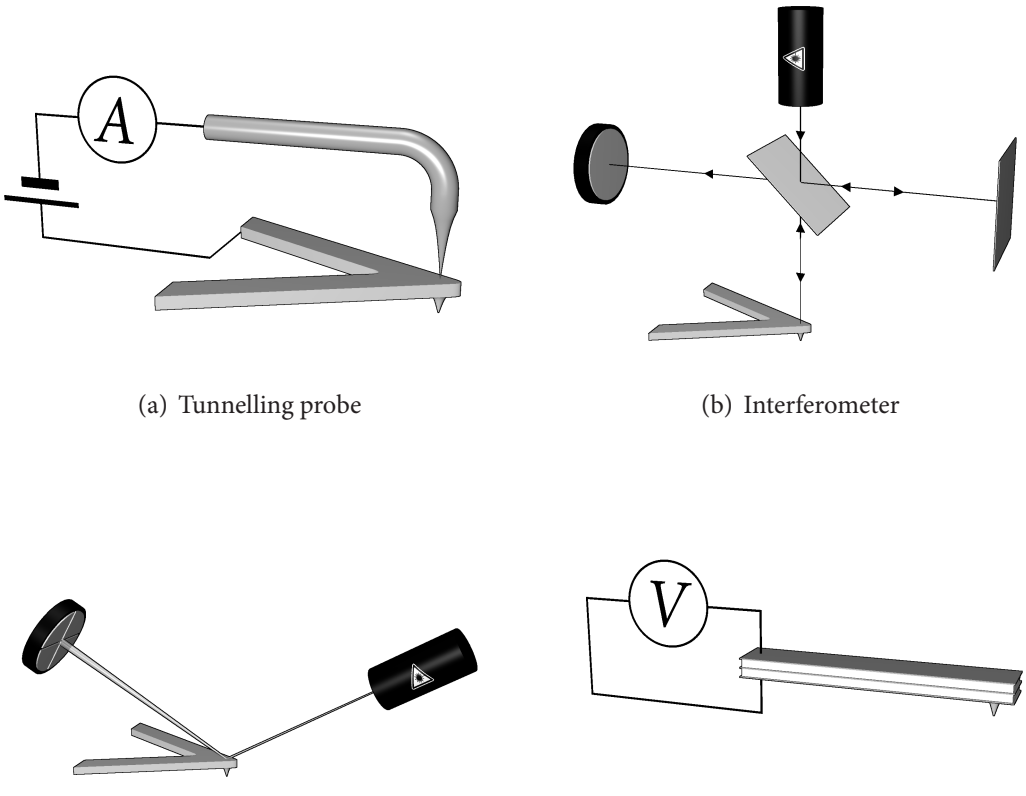
2.4.3 Detection Strategies

Tunnelling Probe

The original atomic force microscope proposed by BINNIG, QUATE, and GERBER [7] uses the tip of an STM to measure the deflection of a conducting cantilever bearing the AFM tip as shown in Fig. 2.3(a). This method can use separate feedback loops for the AFM and STM parts of the experiment. Because of its complexity, limitation to conducting cantilevers, and sensitivity to contamination it is of little relevance nowadays.

Interferometer

The interferometer is a standard instrument for measuring small changes in position: Light is sent along two different paths, one of which depends on the displacement to be measured. The light is then allowed to interfere, and the relative phase, which depends on the path difference, can be determined. The technique can be applied straightforwardly to the SFM as shown in Fig. 2.3(b) and has the advantage of providing an intrinsic calibration of the deflection via the wave length [8, 26].



(a) Tunnelling probe

(b) Interferometer

(c) Optical lever

(d) Piezoelectric cantilever

Figure 2.3: Detection strategies for SFM cantilever deflection

Optical Lever

Today, most SFMs use the inexpensive optical lever technique illustrated in Fig. 2.3(c) to measure cantilever deflection. A collimated beam of light is reflected off the cantilever and projected on a position-sensitive photodetector (PSD). A detector with two segments allows the detection of the movement along one axis only, while one with four segments will detect a shift along the perpendicular direction as well. The bending of the cantilever changes the angle of incidence, so that the deflected beam falls onto a different vertical position on the PSD; a lateral shift is caused by torsion in the presence of a frictional force. In both cases the shift of the spot depends on the distance between the cantilever and the PSD; given sufficient stability, even small tilts can be detected easily [27].

This method was known long before the SFM existed, being used in early profilometers which are conceptually similar to AFMs but operate at much lower resolution. Instead of a PSD, these instruments used a moving strip of photographic film, on which a magnified representation of the sample profile would be traced by the reflected beam.

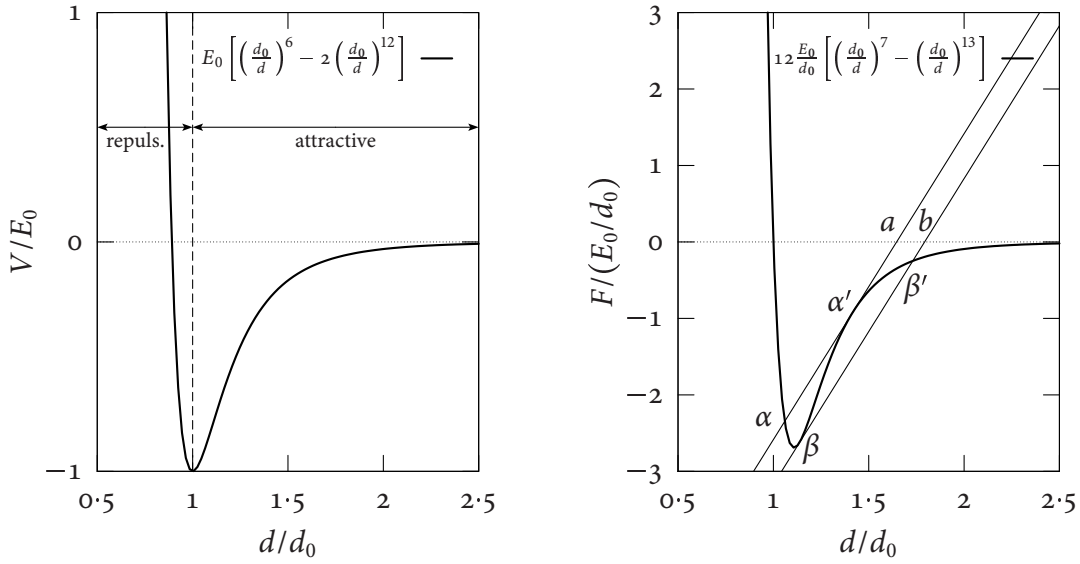
Piezoelectric Cantilever

Fig. 2.3(d) shows how the deflection can be measured directly by using a cantilever built from a piezoelectric unimorph or bimorph sensor (see Section A.1.4). As no external detection mechanism is required, the complexity of the instrument is reduced. However, fabricating and changing the cantilevers is more difficult and expensive. Piezoelectric detection is particularly interesting for applications where an external detection mechanism is difficult to implement because of limited space or accessibility [28, 29].

2.4.4 Tip–Sample Interaction in the Atomic Force Microscope

The Problem

In contrast to the STM, there is in general no useful approximation such as Eq. (2.1) for the force between the tip of an AFM and the sample. Instead, the potential has contributions from different interactions. Some of these contributions decay more slowly than the tunnel current, so that the macroscopic shape of the tip and the medium play an important role and a purely local model may be insufficient to predict the force on the tip. Even so, short-rage forces between the atoms at the tip apex and the sample may still enable atomic resolution. Especially when operating in the repulsive regime, the interaction force can be high enough that the elastic deformation of the tip and the sample must also be taken into account [20]. For topographic measurements at the scale of several nanometres and above, these complications can often be ignored, but understanding contrast at the atomic scale and between different materials depends on a realistic model of these forces.



(a) Example potential with attractive and repulsive parts

(b) Interaction force corresponding to the potential in (a). Straight lines represent the elastic force of the cantilever.

Figure 2.4: Features of the tip–sample interaction in an AFM.

Interaction Regimes and Jump-to-Contact

In general, the potential has a strongly repulsive part at small distances, which ultimately results from the Pauli exclusion principle, and an attractive part at larger distances, which results from the exchange and electromagnetic interactions. Fig. 2.4(a) uses the familiar Lennard-Jones potential to illustrate this point. It should, however, be noted here that the Lennard-Jones potential, while it captures the essential features of most realistic potentials, is *not* generally a good model for the tip–sample interaction.

Fig. 2.4(b) shows the interaction force corresponding to the potential in Fig. 2.4(a). The cantilever holding the tip above the sample deforms until the elastic force $F_c = k(d - z)$, where k is the stiffness of the cantilever and z its position above the sample, exactly balances the tip–sample interaction force F_{ts}. If the stiffness k is smaller than the largest slope along the curve F_{ts}, there is a region in the attractive part of the potential where two stable equilibrium deflections $d - z$ correspond to a single cantilever position z. The straight lines

in Fig. 2.4(b) correspond to the elastic force of a cantilever at the two positions $z = a$ and $z = b$ delimiting this range; the two points where the forces balance are marked α (β) and α' (β'); for cantilever positions between a and b, there is a third (metastable) point where $F_c = F_{ts}$. When the probe approaches the surface, the deflection increases continuously until it reaches a, where the deflection increases suddenly and the tip jumps from the position corresponding to α' to that corresponding to α. During retraction the deflection changes continuously until the cantilever reaches b, where the tip jumps from β to β'. Consequently, the interaction force measured as a function of the cantilever position z shows discontinuities and hysteresis; the part of the force-distance curve between α' and β is never sampled. The two discontinuities are known as 'jump-to-contact' and 'jump-off-contact', respectively. They can only be avoided by increasing k, which leads to smaller deflections and a lower signal-to-noise ratio [20, 28].

Important Contributions to the Tip–Sample Interaction

Under ambient conditions, many surfaces including oxidized semiconductor surfaces are covered by a thin layer of adsorbed water [20, 26]. Once the water layers on the tip and the surface touch, a meniscus is formed and the surface tension pulls the tip towards the sample. The influence of the water layer is small before the meniscus forms, but once it has formed, it dominates the interaction, extending far from the sample as the meniscus is stretched. The magnitude of the meniscus force is a complicated function of the distance and depends on the shape of the tip; for large $d \gg r$ it varies approximately as $1/d$ until the meniscus ruptures [20].

At large distances and in the absence of a meniscus force, the interaction is dominated by the van der Waals force. This force results from electromagnetic dipole interactions. Between individual atoms or molecules, all contributions to the van der Waals force vary as $1/d^7$. It is generally attractive, but can become repulsive in dense media. The situation is more complic-

ated for macroscopic bodies. Assuming isotropy and additivity, the force between a surface and a sphere of radius r at a distance $d \ll r$ is approximately proportional to d/r^2 [20, 30].

If there is a voltage difference between the tip and the sample, an attractive Coulomb force pulls them together. For a tip with radius of curvature r, the force varies with distance as $1/d$ if $r \gg d$ and as $1/d^2$ if $r \ll d$ [20]. If the tip or the sample are conducting, image forces have to be considered as well. For topographic measurements at the nanometre scale and above, the force caused by an intentional bias can be used to improve the stability of a dynamic force microscope [26].

Chemical forces result from the interaction of the electron clouds and nuclei at the apex of the tip with those at the sample surface. They are the same forces that are responsible for covalent and hydrogen bonds and cause the repulsion that is measured in contact AFM. The attractive part of the potential decays very rapidly with distance, and it is this behaviour that enables atomic resolution in an AFM, since it means that only the microtip coming closest to the sample contributes significantly to the interaction [11, 12].

2.4.5 Dynamic Force Microscopy

Mechanism

A DFM is an SFM in which the cantilever and tip system is vibrated close to a resonant frequency, usually using a piezoelectric transducer. The dynamic behaviour of the cantilever changes as a result of the tip–sample interaction, and it is this change that is measured and used to form an image. The original AFM by BINNIG *et al.* already included dynamic force capabilities [7], but static force measurements were found to be more reliable. The first working DFM was demonstrated by MARTIN *et al.* [8] in 1987. Compared to static force microscopy, DFM is more suitable for measurements in the attractive region of the tip–sample potential and avoids jump-to-contact. The signal-to-noise ratio can be improved by using narrow-band detection in combination with a standard lock-in technique [28].

The Frequency Shift

The motion of the cantilever can be approximated[5] by that of a point mass on a massless harmonic spring [11, 30, 31]. In the presence of a tip–sample interaction force $F_{\text{ts}}(z)$ and a driving force $F_{\text{ext}}(t) = F_0 \cos(\omega t)$ the equation of motion is

$$m^* \ddot{z} + 2\gamma m^* \dot{z} + k\,(z - z_0) = F_{\text{ts}}(z) + F_{\text{ext}}(t), \tag{2.8}$$

where m^* is the effective mass, γ the damping coefficient, and k the stiffness. If the amplitude of the oscillation is small compared to the distance over which $F_{\text{ts}}(z)$ changes, $F_{\text{ts}}(z) \approx (z - z_0)F'_{\text{ts}}(z_0)$. For realistic long range interaction forces, this approximation is valid if the amplitude A is much smaller than the distance of closest approach D; for general forces, the required A may be arbitrarily small [32, 33]. Eq. (2.8) then becomes

$$m^* \ddot{z} + 2\gamma m^* \dot{z} + \left[k - F'_{\text{ts}}(z_0)\right](z - z_0) = F_{\text{ext}}(t). \tag{2.9}$$

This is just the equation of a forced harmonic oscillator with an effective spring constant $k^* \stackrel{\text{def}}{=} k - F'_{\text{ts}}(z_0)$ and the resonant (angular) frequency is

$$\omega_0 = \sqrt{\frac{k^*}{m^*}} = \sqrt{\frac{k - F'_{\text{ts}}(z_0)}{m^*}}. \tag{2.10}$$

The frequency shift due to the interaction between the tip and the sample is hence

$$\Delta\omega_0 = \sqrt{\frac{k}{m^*}} - \sqrt{\frac{k - F'_{\text{ts}}(z_0)}{m^*}} \approx -\frac{F'_{\text{ts}}(z_0)}{2k}\omega_0 \tag{2.11}$$

for $F'_{\text{ts}}(z_0) \ll k$ [8, 26, 28]. An image taken at constant frequency shift $\Delta\omega_0$ therefore corresponds to a surface of constant force gradient F'_{ts}. In practice, however, the condition $A \ll D$ is *not* fulfilled in many DFM experiments [32, 33].

Schwarz *et al.* [33] suggest a different approach that approximates the true anharmonic potential by two harmonic potentials and is valid for $A \gg D$, provided the decay length λ of the tip–sample interaction is much smaller than A. In this theory, the frequency shift obeys

$$\Delta\omega_0 \propto \frac{V_{\text{ts}}(D)}{\sqrt{\lambda}}, \tag{2.12}$$

[5] This is a reasonable approximation for cantilevers *in vacuo* but increasingly less so as hydrodynamic effects become more important in more viscous media.

where $V_{ts}(D)$ is the tip–sample interaction potential at the point of closest approach and λ the range of the interaction. It can be understood intuitively that the potential at the point of closest approach dominates the effect: The tip has its lowest speed there and experiences the potential for a longer time; since $\lambda \ll A$, the interaction is negligible at the other turning point [11, 33].

To arrive at a more accurate understanding of the tip dynamics, comparison with simulations from a realistic model of the tip–sample interaction is required and the behaviour of the feedback system determining F_{ext} has to be taken into account.

Amplitude Modulation

In amplitude modulation AFM (AM-AFM), the cantilever is excited at a constant frequency. The detection mechanism is used to measure the change in the amplitude response caused by the shift in the resonant frequency due to the interaction of the tip with the sample. The amplitude is either recorded directly or kept constant using a feedback loop [8, 30, 31, 33]. In this mode, further spectroscopic information may be obtained by measuring the phase shift between the excitation signal and the cantilever vibration [30].

Frequency Modulation

The alternative approach is frequency modulation AFM (FM-AFM), which uses a second, faster feedback loop to keep the measured amplitude constant by varying the excitation frequency and amplitude. This way, the shift in the resonance frequency can be measured directly; the excitation amplitude contains information about energy transfer to the sample [30, 31, 33, 34]. FM-AFM is preferable to AM-AFM under high vacuum conditions, because the absence of a damping medium implies a very high quality factor and the oscillation amplitude responds slowly to a change in the resonant frequency [34].

Dynamic Force Microscopy in a Liquid

DFM in a liquid is interesting because it is required to observe many biological samples and potentially allows atomic resolution. Modelling the dynamics is more involved, because the

movement of the liquid between the cantilever and the sample becomes important and the probe can no longer be approximated by a mass on a spring [30, 35]. Compared to the same cantilever in air or vacuum, there are more and broader resonances at lower frequencies [30] and the quality factor decreases by two orders of magnitude. The latter problem can be overcome by using a positive feedback loop to drive the oscillation [36].

2.4.6 True Atomic Resolution

Like the STM, the AFM can in principle achieve atomic resolution [12, 28, 33, 35]. In practice, this goal is considerably harder to achieve, since important contributions to the tip–sample interaction decay slowly. However, only short-range chemical forces can be used for imaging at the atomic scale. Sharp tips are necessary to reduce the contribution of the van der Waals force [28]. In air, the meniscus force dominates the interaction; measurements must hence be taken in a liquid [35,37] or in UHV [28,33]. While the repulsive forces used in contact AFM are short range and images with correct atomic spacings have been obtained this way, the forces on typical tips in contact mode are too large to be supported by a single microtip, preventing true atomic resolution [28, 33]. Additionally, the atomically clean surfaces of many tips and samples can stick together under UHV conditions by forming chemical bonds [28]. Atomic resolution images in UHV have only been obtained by large amplitude DFM, which prevents sticking and jump-to-contact. As with the STM, atomic resolution imaging of many materials may call for elaborate sample preparation, which can in itself require working in UHV.

2.4.7 Contact, Non-Contact, and Tapping Mode

In the literature the operating mode of a force microscope is often characterized as contact, non-contact (NCM), or tapping mode. In contact mode the repulsive part of the surface potential is probed. This is the usual situation for static force measurements, for which the main feedback loop is set up to increase the tip-sample distance when the detected force increases. Non-contact mode refers to operation in the attractive part of the potential. This regime is

usually associated with dynamic force microscopy, which allows for increased sensitivity in detecting small attractive forces and prevents jump-to-contact. Finally, tapping mode designates a DFM experiment in which the tip samples the attractive part of the potential for most of its oscillation cycle but penetrates into the repulsive part on closest approach to the sample. In practice distinguishing between non-contact and tapping modes in DFM experiments may be difficult and requires a detailed understanding of the tip-sample interaction. FM-AFM is often identified with NCM and AM-AFM with tapping mode [30, 31]; this usage frequently, but by no means necessarily agrees with the straightforward definitions given here.

2.4.8 Artefacts in Atomic Force Microscope Images

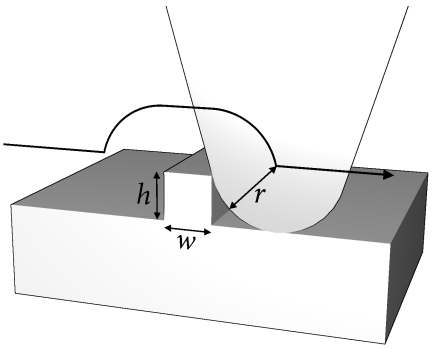
Finite Tip Size

As shown in Fig. 2.2(b), real AFM tips have a finite radius of curvature r and angle α. When scanning sample features with high aspect ratios, the point of contact is not always at the apex of the tip. Consequently, such structures cannot be traced accurately. This effect is illustrated in Fig. 2.5: The lateral size of small features is overestimated, while the tip cannot reach the bottom of deep holes or trenches. Vertical sidewalls are imaged with rounded or slanted profiles, depending on the shape of the tip at the scale of the feature. Artefacts due to finite tip size can usually be distinguished from sample features by the fact that they do not change their orientation when the sample is rotated relative to the tip. They often appear as the repetition of a pattern that corresponds to the shape of the tip apex.

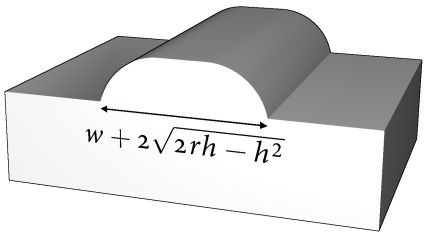
If the surface of the sample is given by $S(x, y)$ and the surface of the tip by $T(x, y)$, the image or apparent sample surface is

$$S'(x, y) = -\min_{\xi,\eta\in\mathbb{Z}} \left[T(\xi - x, \eta - y) - S(\xi, \eta) \right]. \tag{2.13}$$

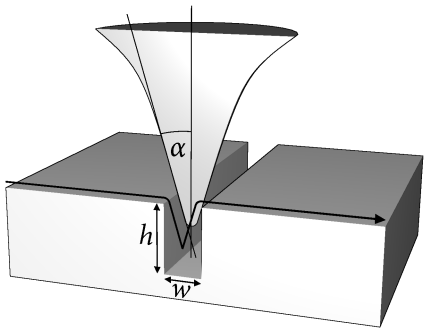
At each point (x, y), the minimal distance between the entire shifted tip $T(\xi - x, \eta - y)$ and



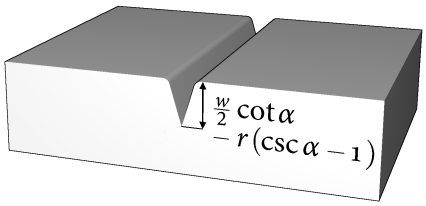
(a) Measurement using a tip with finite radius r



(b) Apparent topography measured in (a)



(c) Measurement using a tip with finite angle α



(d) Apparent topography measured in (c)

Figure 2.5: Finite tip size effects in an AFM measurement. Arrows indicate the trajectory of the centre of curvature of the tip.

the entire sample surface $S(\xi, \eta)$ determines the distance by which the tip can be lowered towards the sample before contact is established.

For a known tip shape and surface Eq. (2.13) allows a straightforward numerical simulation of the expected image. VILLARRUBIA [38] has developed a formulation in terms of mathematical morphology and used it to show that the operation can be reversed: Given the shape of the tip and the image, the sample surface can be recovered. The reconstruction is exact where the surface has been in contact with the tip and provides an upper bound elsewhere. Similarly, the tip can be reconstructed by the same method if the sample geometry is known.

In practice, the reconstruction of images is hampered by the fact that the shape of the tip apex is usually not well defined, as it can change during operation—even if care is taken to assess the specific probe by imaging a characterizer or by other means. If the tip is known to be sufficiently sharp that its area of interaction with the sample is much smaller than a given image, this image can be used to estimate the current shape of the tip via a self-consistent iterative 'blind reconstruction' method [38].

For the idealized finite size tip shown in Fig. 2.2(b) and surface features with simple geometries, closed expressions for the imaging error can be obtained in many cases by a simple geometrical construction; such expressions can be useful in the quick assessment of AFM images. A mesa with a rectangular cross-section of width w and height h, when imaged using a tip of finite radius of curvature $r \geq h$—Fig. 2.5(a) and (b)—produces an image with apparent width

$$w' = w + 2\sqrt{2rh - h^2}. \tag{2.14}$$

If a hole or groove is imaged with a tip that cannot reach its bottom as in Fig. 2.5(c) and (d), the apparent depth of the feature is given by

$$h' = \frac{w}{2} \cot \alpha \tag{2.15}$$

for a tip with a sharp point. If a finite radius of curvature $r \gtrsim h$ is taken into account, the length of the tip is reduced, and Eq. (2.15) becomes

$$h' = \frac{w}{2} \cot \alpha - r(\csc \alpha - 1). \tag{2.16}$$

Other Effects

AFMs are naturally susceptible to artefacts affecting all SPMs, such as scanner nonlinearity, smoothing and overshoot due to the finite response time of the feedback loop, and narrow-band noise mimicking a periodic surface structure. Specifically in an AFM, the tip can momentarily stick to the surface and then jump off, resulting in a spike in the image that does not correspond to a sample feature. In a DFM, for some operating parameters there can be

two stable oscillation states corresponding to a single value of the tip–sample force [30]. In this case, the apparent topography can jump between two levels with an approximately constant offset between them. The jumps are usually random, but can sometimes mimic a step in the sample.

2.5 Scanning-Probe-Based Lithography

2.5.1 Manipulation of Individual Atoms

A famous example of the ability of the SPM to shape matter at an atomic scale is the 'quantum corral' of iron atoms on a copper surface created by Crommie *et al.* [39] via the controlled movement of adatoms using an STM. The sensitivity of the STM to the local electronic density of states has been used to image the standing wave corresponding to an electron trapped in the artificial enclosure.

In order to manipulate individual atoms, the STM must operate under UHV, which is required for atomic resolution imaging. Performing the experiment at liquid helium temperatures improves the stability and may be necessary to prevent diffusion of the adatoms. The tip can be used to move individual atoms either in a parallel process, in which the connection to the surface is never broken, or in a perpendicular process, in which the atom detaches from the surface and is adsorbed at the tip. In the parallel process, the electric field at the apex of the tip can be increased to enhance diffusion in the desired direction. Alternatively, the tip can be lowered towards an adatom by changing the feedback parameter until the chemical forces allow for sliding the atom across the surface. The perpendicular process transfers atoms between the surface and the tip by approaching and retracting the probe or by applying a voltage pulse to overcome the potential barrier [40].

2.5.2 Nanoindentation

AFMs can also be used to modify a sample mechanically. The tip is pressed against the workpiece with a force exceeding the threshold for inelastic deformation and indents or chips the

surface. It must be attached to a sufficiently stiff cantilever to support the interaction force and commonly consists of a hard material, *e.g.*, diamond, to minimize wear. Apart from the sharpness and durability of the tip, the performance of this method depends critically on the mechanical properties of the sample itself; under optimal conditions, a resolution of 10 nm can be realized [41,42]. Since high resolution modifications can only be achieved with certain substrates, a separate surface layer consisting of a soft polymer, metal, or oxide is sometimes applied and the pattern transferred using an auxiliary etch technique. Nanoscratching with an oscillating force as in a DFM helps avoid sticking of the tip [41, 43, 44].

2.5.3 Electrical and Optical Surface Modification

A voltage pulse applied to a conducting STM or AFM tip can modify the substrate in various other ways. Field emission from an STM tip can be used to expose organic resists at length scales comparable with the best electron beam writers. Joule heating due to local high current densities and field-assisted evaporation have been proposed as mechanisms for the surface modification observed in some experiments. The electrical field in the vicinity of an STM or AFM tip can activate the chemical vapour deposition (CVD) of some materials, especially metals. Finally, the voltage difference can be used to locally enable electrochemical processes [42]. A special case of the latter technique is the local anodic oxidation of metal or semiconductor surfaces using adsorbed water as the electrolyte, which is discussed in more detail in Chapter 3. Similar to the situation with nanoindentation, these methods critically depend on the chemical and electrical properties of the involved materials; sometimes an additional pattern transfer process can circumvent this limitation. Optical techniques include the exposure of conventional photoresist and other photoactive surfaces by illumination mode NFSOM.

2.6 Summary

The SPM is a versatile tool that can measure and manipulate a surface at length scales from fractions of an Ångström up to several microns. The concept can be used with a large range

of different tip–sample interactions, and many different physical properties of the sample can be mapped. Advances in scanner and noise control technology mean that the fundamental limit to the resolution of an SPM method is determined by the size and shape of the probe and the range of the interaction it uses. Correct interpretation of images requires a detailed understanding of this interaction at the relevant scale.

Bibliography

[1] E. T. Yu. Nanoscale characterization of semiconductor materials and devices using scanning probe techniques. *Materials Science and Engineering*, R17:147–206, 1996.

[2] G. Schmalz. Über Glätte und Ebenheit as physikalisches und physiologisches Problem. *Zeitschrift des Verbandes Deutscher Ingenieure*, 73:144–161, October 1929.

[3] R. Young, J. Ward, and F. Scire. Observation of metal-vacuum-metal tunneling, field emission, and the transition region. *Physical Review Letters*, 27(14):922–924, October 1971.

[4] R. Young, J. Ward, and F. Scire. The topographiner: An instrument for measuring surface microtopography. *Review of Scientific Instruments*, 43(7):999–1011, July 1972.

[5] G. Binnig and H. Rohrer. Scanning tunneling microscopy—from birth to adolescence. *Reviews of Modern Physics*, 59(3):615–625, July 1987. Nobel Lecture.

[6] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel. Surface studies by scanning tunneling microscopy. *Physical Review Letters*, 49(1):57–61, July 1982.

[7] G. Binnig, C. F. Quate, and C. Gerber. Atomic force microscope. *Physical Review Letters*, 56(9):930–933, March 1986.

[8] Y. Martin, C. C. Williams, and H. K. Wickramasinghe. Atomic force microscope-force mapping and profiling on a sub-100 Å scale. *Journal of Applied Physics*, 61(10):4723–4729, May 1987.

[9] U. Dürig, D. W. Pohl, and F. Rohner. Near-field optical-scanning microscopy. *Journal of Applied Physics*, 59(10):3318–3327, May 1986.

[10] R. C. Reddick, R. J. Warmack, and T. L. Ferrell. New form of scanning optical microscopy. *Physical Review A*, 39(1):767–770, January 1989.

[11] D. Drakova. Theoretical modelling of scanning tunnelling microscopy, scanning tunnelling spectroscopy and atomic force microscopy. *Reports on Progress in Physics*, 64:205–290, 2001.

[12] W. A. Hofer, A. S. Foster, and A. L. Shluger. Theories of scanning probe microscopes at the atomic scale. *Reviews of Modern Physics*, 75:1287–1331, October 2003.

[13] S. Gasiorowicz. *Quantum Physics*, chapter 5. John Wiley & Sons, New York, second edition, 1996.

[14] J. Tersoff and D. R. Hamann. Theory and application for the scanning tunneling microscope. *Physical Review Letters*, 50(25):1998–2001, June 1993.

[15] J. G. Simmons. Generalized formula for the electric tunnel effect between similar electrodes separated by a thin insulating film. *Journal of Applied Physics*, 34(6):1793–1803, June 1963.

[16] C. Girard, C. Joachim, and S. Gauthier. The physics of the near-field. *Reports on Progress in Physics*, 63:893–938, 2000. read 19th July 2005; C.

[17] J. W. P. Hsu. Near-field scanning optical microscopy studies of electronic and photonic materials and devices. *Materials Science and Engineering*, 33:1–50, 2001.

[18] J. E. Griffith, G. L. Miller, C. A. Green, D. A. Grigg, and P. E. Russell. A scanning tunneling microscope with a capacitance-based position monitor. *Journal of Vaccum Science and Technology B*, 8(6):2023–2027, November/December 1990.

[19] K. Dirscherl, J. Garnæs, and L. Nielsen. Modeling the hysteresis of a scanning probe microscope. *Journal of Vacuum Science and Technology B*, 18(2), 621–625 2000.

[20] B. Cappella and G. Dietler. Force-distance curves by atomic force microscopy. *Surface Science Reports*, 34:1–104, 1999.

[21] L. Libioulle, A. Ronda, M. Taborelli, and J. M. Gilles. Deformations and nonlinearity in scanning tunneling microscope images. *Journal of Vacuum Science and Technology B*, 9(2):655–658, March/April 1991.

[22] H.-C. Yeh, W.-T. Ni, and S.-S. Pan. Digital closed-loop nanopositioning using rectilinear flexure stage and laser interferometry. *Control Engineering Practice*, 13:559–566, 2005.

[23] K. R. Koops, P. M. L. O. Scholte, and W. L. de Koning. Observation of zero creep in piezoelectric actuators. *Applied Physics A*, 68:691–697, 1999.

[24] L. E. Cross. Ferroelectric materials for electromechanical transducer applications. *Materials Chemistry and Physics*, 43:108–115, 1996.

[25] J. E. Sader. Susceptibility of atomic force microscope cantilevers to lateral forces. *Review of Scientific Instruments*, 74(4):2438–2443, April 2003.

[26] R. Erlandsson, G. M. McClelland, C. M. Mate, and S. Chiang. Atomic force microscopy using optical interferometry. *Journal of Vacuum Science and Technology A*, 6(2):266–270, March/April 1988.

[27] S. Alexander, L. Hellemans, O. Marti, J. Schneir, V. Elings, P. K. Hansma, M. Longmire, and J. Gurley. An atomic-resolution atomic-force microscope implemented using an optical lever. *Journal of Applied Physics*, 65(1):164–167, January 1989.

[28] F. J. Giessibl. Atomic resolution of the silicon (111)-(7×7) surface by atomic force microscopy. *Science*, 267(5194):68–71, January 1995.

[29] T. Akiyama, S. Gautsch, N. F. de Rooij, U. Staufer, P. Niedermann, L. Howald, D. Müller, A. Tonin, H.-R. Hidber, W. T. Pike, and M. H. Hecht. Atomic force microscope for planetary applications. *Sensors and Actuators A*, 91:321–325, 2001.

[30] R. García and R. Pérez. Dynamic atomic force microscopy methods. *Surface Science Reports*, 47:197–301, 2002.

[31] G. Couturier, R. Boisgard, L. Nony, and J. P. Aimé. Noncontact atomic force microscopy: Stability criterion and dynamical responses of the shift of frequency and damping signal. *Review of Scientific Instruments*, 74(5):2726–2734, May 2003.

[32] H. Hölscher, U. D. Schwarz, and R. Wiesendanger. Calculation of the frequency shift in dynamic force microscopy. *Applied Surface Science*, 140:344–351, 1999.

[33] U. D. Schwarz, H. Hölscher, and R. Wiesendanger. Atomic resolution in scanning force microscopy: Concepts, requirements, contrast mechanisms, and image interpretation. *Physical Review B*, 62(19):13089–13097, November 2000.

[34] R. Bennewitz, M. Bammerlin, M. Guggisberg, C. Loppacher, A. Barato, E. Meyer, and H.-J. Güntherodt. Aspects of dynamic force microscopy on NaCl/Cu(111): Resolution, tip–sample interactions and cantilever oscillation characteristics. *Surface and Interface Analysis*, 27:462–466, 1999.

[35] F. M. Ohnesorge. Towards atomic resolution non-contact dynamic force microscopy in a liquid. *Surface and Interface Analysis*, 27:179–385, 1999.

[36] J. Tamayo, A. D. L. Humphris, R. J. Owen, and M. J. Miles. High-Q dynamic force microscopy in liquid and its application to living cells. *Biophysical Journal*, 81:526–537, July 2001.

[37] F. M. Ohnesorge and G. Binnig. True atomic resolution by atomic force microscopy through repulsive and attractive forces. *Science*, 360(5113):1451–1456, June 1993.

[38] J. S. Villarrubia. Algorithms for scanned probe microscope image simulation, surface reconstruction, and tip estimation. *Journal of Research of the National Institute of Standards and Technology*, 102(4):425–454, July/August 1997.

[39] M. F. Crommie, C. P. Lutz, and D. M. Eigler. Confinement of electrons to quantum corrals on a metal surface. *Science*, 262(5131):218–220, October 1993.

[40] J. A. Stroscio and D. M. Eigler. Atomic and molecular manipulation with the scanning tunneling microscope. *Science*, 254(5036):1319–1326, November 1991.

[41] Y. Kim and C. M. Lieber. Machining oxide thin films with an atomic force microscope: Pattern and object formation on the nanometer scale. *Science*, 257:375–377, July 1992.

[42] R. M. Nyffenegger and R. M. Penner. Nanometer-scale surface modification using the scanning probe microscope: Progress since 1991. *Chemical Reviews*, 97:1195–1230, 1997.

[43] M. Wendel, B. Irmer, J. Cortes, R. Kaiser, H. Lorenz, J. P. Kotthaus, A. Lorke, and E. Williams. Nanolithography with an atomic force microscope. *Superlattices and Microstructures*, 20(3):349–356, 1996.

[44] U. Kunze. Nanoscale devices fabricated by dynamic ploughing with an atomic force microscope. *Superlattices and Microstructures*, 31(1):3–17, 2002.

3 Local Anodic Oxidation

Der Worte sind genug gewechselt,
Laßt mich auch endlich Taten sehn!
Indes ihr Komplimente drechselt,
Kann etwas Nützliches geschehn.

(Johann Wolfgang von Goethe, Faust I)

3.1 Introduction

3.1.1 Overview

A NEGATIVE BIAS applied to the tip of a scanning probe microscope (SPM) can locally oxidize nearby metal or semiconductor surfaces. This anodization process, which is illustrated in Fig. 3.1, is known as local anodic oxidation (LAO) or SPM lithography.[1] It requires that the instrument operate in air under ambient conditions, so that oxide-passivated hydrophilic surfaces will be covered by a thin layer of adsorbed water [1].

After the initial oxide layer has been created, further oxidation requires the migration of oxyanions, substrate cations, or both through the oxide film. The electric field set up in the oxide by the voltage difference between the probe and the substrate enhances their diffusion into the surface. If an SPM probe is brought close to the surface and a bias is applied to it, a water meniscus, which acts as a source of oxygen anions, forms between the tip and the sample [2–4]. As the formation of this meniscus is essential for successful anodization, the

[1] From Greek λίθος, 'stone', and γράφειν, 'to draw'. Originally referring to printing with a limestone, the word has come to mean any pattern transfer technique.

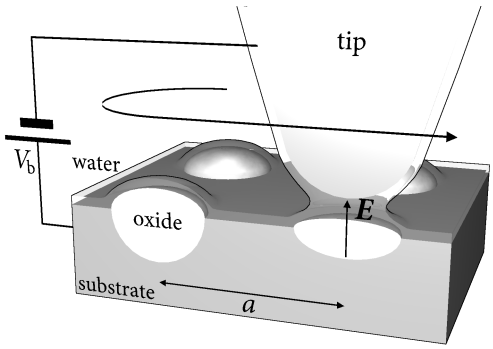


Figure 3.1: Local anodic oxidation

technique depends on the ambient humidity, which affects the thickness and continuity of the adsorbed water layer.

3.1.2 Motivation

LAO has successfully been demonstrated on a variety of materials, including III–V semiconductors. Thanks to improvements in SPM technology, such as feedback mechanisms that reduce drift originating from the creep and the hysteresis of the piezoelectric actuators, the method has evolved to the point where it allows the creation of complex nanoscale features with a resolution comparable to electron beam lithography (EBL).

In contrast to EBL, which will be considered in more detail in Sec. 4.2.2, LAO with an AFM does not suffer from proximity effects and can achieve optimal resolution even for dense patterns [5, 6]. The demands on the instrumentation are not as strict and optimal resolution can be obtained with relatively inexpensive equipment. Moreover, many materials can be manipulated directly, and a separate mask and etching procedure, which can incur damage to the bulk of the substrate, may be avoided. On the other hand, LAO is restricted to shallow etching of a limited number of materials, suffers from a poor reproducibility, and it now appears that the ultimate resolution obtainable by this technique[2] is less than that of EBL.[3]

[2] This is certainly not true for SPM-based lithographic techniques in general

[3] Another limitation, which is of little importance for scientific research and prototyping, but precludes the use of LAO in industry, is the extremely low yield of this technology even if compared to EBL.

The aim of the present work is to investigate lateral modulations of the quasi-two-dimensional carrier gases that form in GaSb–InAs–GaSb double heterojunctions (DHETS). I have set out to evaluate the use of LAO for creating such modulations by removing part of the upper GaSb layer, exploiting the sensitivity of these carrier gases to the proximity of surface states. Chapter 5 discusses the fundamental electronic properties of the DHETS, while chapter 7 considers lateral modulations in the form of antidot patterns and reports on the effect of dot patterns created by LAO on the magnetotransport in these structures.

3.2 Oxide Growth

3.2.1 The Mott-Cabrera Model

LAO is a practical technique for materials that form thin protective oxide layers in the presence of oxygen. In the temperature range where such layers are formed, the oxide growth on these materials starts out quickly but virtually stops after an oxide of finite thickness has been formed. For the materials approachable by LAO, the ultimate thickness is of the order of several nm to several tens of nm. The formation of this native oxide can be explained within the framework of the theory originally developed by NEVILL MOTT and NICOLÁS CABRERA [7, 8].

Discussing this model, I shall follow the usual conventions [8, 9] and assume a quasi-one-dimensional system consisting of:

1. the parent material, *i.e.*, bulk metal or semiconductor, for $x \in (-\infty, 0)$
2. an oxide layer of thickness X in the region $x \in [0, X]$
3. a source of oxygen, *e.g.*, air or water, for $x \in (X, \infty)$.

Once the initial oxide layer has formed, further oxidation requires mass transport through this film, which in general may proceed through diffusion either of ions or ionic defects or of neutral atoms. If no external circuitry is present, electrons and holes must pass through

the oxide as well. In the steady state approximation, the concentration n_s of each diffusing species s is assumed to be constant in time. Particle conservation, in the form of the familiar continuity equation $\partial n_s/\partial t = -\partial j_s/\partial x$, then dictates that $\partial j_s/\partial x = 0$ and the particle current j_s is constant throughout the oxide. One can accordingly write the oxide growth rate as

$$\frac{dX}{dt} = \sum_s \Omega_s j_s, \tag{3.1}$$

where Ω_s is the volume of oxide formed by the transversal of a single ion of species s [9].

According to Mott and Cabrera, the effect of the space charge created by the diffusing charged particles can be neglected for thin oxide layers: The contribution V_ρ to the potential due to the space charge $\rho(x) = \sum_s q_s n_s(x)$, where q_s is the charge of a single ion of species s, can be found from Poisson's equation and evaluates to $(1/2\varepsilon)X^2\rho$ if ρ is independent of x; if $X \ll \sqrt{\varepsilon k_B T/e\rho}$, eV_ρ is small compared with $k_B T$. This condition is typically fulfilled for native oxide films [7–9].

They further assume that electrons are transported through thin oxide films via tunnelling or thermal emission and the electronic motion is rapid compared to the ionic motion. If a layer of oxygen is present at X, some of the oxygen atoms will be converted to O^- ions until the chemical potential is equal on both sides of the oxide. This charge separation causes a built-in voltage V_M across the oxide, which is sometimes designated the Mott potential [10]. The magnitude of V_M may be estimated by $eV_M \approx EA_O + W - \varphi_m$, where EA_O is the electron affinity of oxygen, W the binding energy of the oxygen ion, and φ_m the work function of the parent material; V_M is typically of the order 1 V.

For thin films, the electric field $E = V_M/X$ becomes large and controls the ionic transport. If the particle current j_s of ions of species s obeys the linear relation $j_s = \mu_s E$, where μ_s is the mobility, the assumption of single carrier transport leads to a growth rate [7, 8]

$$\frac{dX}{dt} = \frac{\Omega\mu V_M}{X}, \tag{3.2}$$

which can be integrated to give the well-known parabolic law [7–9, 11, 12]

$$X(t) \propto \sqrt{t - t_0} \tag{3.3}$$

with constant of proportionality $\sqrt{2\Omega\mu V_M}$.

For very thin oxide layers, however, the electric field becomes so strong that j_s is no longer proportional to E. In this case, the relationship between the electric field and the particle current can be derived from the hopping model originally formulated by Verwey [7–9]: Let the distance between adjacent potential minima in the oxide be a and the energy barrier between them U. The classical Boltzmann probability that a particle will move from one site to the next is then $\nu \exp(-U/k_\mathrm{B}T)$, where the attempt frequency ν can be identified with the typical phonon frequency [7–9]. An electric field $E(x)$ lowers the activation energy barrier for movement in the positive x-direction by $\frac{1}{2}aq_sE(x)$. The hopping probability thus becomes $\nu \exp\left\{-\left[U - \frac{1}{2}aq_sE(x)\right]/k_\mathrm{B}T\right\}$ in the forward direction, while the probability for hopping in the reverse direction is reduced to $\nu \exp\left\{-\left[U + \frac{1}{2}aq_sE(x)\right]/k_\mathrm{B}T\right\}$ and can be neglected in the strong field limit $aq_sE(x) \gg k_\mathrm{B}T$. The particle current of species s is therefore given by

$$j_s \approx \frac{n_s(x)}{a}\nu \exp\left(\frac{-U}{k_\mathrm{B}T}\right) 2\sinh\left[\frac{\frac{1}{2}aq_sE(x)}{k_\mathrm{B}T}\right]. \tag{3.4}$$

In the strong field limit, $\frac{1}{2}aq_sE(x) \gg k_\mathrm{B}T$ and $2\sinh\left[\frac{1}{2}aq_sE(x)/k_\mathrm{B}T\right]$ can be approximated by $\exp\left[\frac{1}{2}aq_sE(x)/k_\mathrm{B}T\right]$.[4] Assuming that space charges are negligible, $E(x) = V_\mathrm{M}/X$, $n_s(x) = n_s(0)$ and Eq. (3.4) becomes [8]

$$j_s \approx \frac{n_s(0)}{a}\nu \exp\left(\frac{-U}{k_\mathrm{B}T}\right) \exp\left(\frac{\frac{1}{2}aq_s}{k_\mathrm{B}T}\frac{V_\mathrm{M}}{X}\right). \tag{3.5}$$

For single carrier transport, Eqs. (3.1) and (3.5) imply

$$\frac{dX}{dt} \approx \Omega\frac{n(0)}{a}\nu \exp\left(\frac{-U}{k_\mathrm{B}T}\right) \exp\left(\frac{\frac{1}{2}aq}{k_\mathrm{B}T}\frac{V_\mathrm{M}}{X}\right). \tag{3.6}$$

If transport through the oxide is rate determining, $n(0)$ is approximately independent of the oxide film thickness and determined by interfacial equilibrium conditions [9]. In contrast, if we follow Mott and Cabrera in assuming that the transport in the presence of

[4] On the other hand, if $\frac{1}{2}aq_sE(x) \ll k_\mathrm{B}T$, $2\sinh\left[\frac{1}{2}aq_sE(x)/k_\mathrm{B}T\right]$ is approximately $aq_sE(x)/k_\mathrm{B}T$. By considering the total current between x and $x + dx$, the linear diffusion equation $j_s = -D_s\, dn_s/\, dx + \mu_s E n_s$ is then recovered.

a strong field is very rapid and the limiting step is the dissolution of the ions at the interface, the rate of oxidation is given by

$$\frac{dX}{dt} \approx \Omega N' \nu \exp\left(-\frac{W_i + U}{k_B T}\right) \exp\left(\frac{\frac{1}{2} bq}{k_B T} \frac{V_M}{X}\right), \tag{3.7}$$

where W_i is the heat of solution of the ion, b the width of the energetic barrier it has to overcome, and N' is the sheet density of ions available for dissolution [7, 8]. The derivation is analogous to the argument given above.

In both cases, dX/dt has the same functional form

$$\frac{dX}{dt} = u_0 \exp\left(\frac{-W}{k_B T}\right) \exp\left(\frac{X_0}{X}\right). \tag{3.8}$$

By considering growth to have stopped once $dX/dt \gtrsim u_L$, a limiting oxide thickness X_L can be defined as

$$X_L = \frac{X_0}{\frac{W}{k_B T} - \ln\left(\frac{u_0}{u_L}\right)}. \tag{3.9}$$

Following this argument, the oxide growth will stop at a thickness $X_L \propto X_0 \propto V_M$ provided that $T < W/[k_B \ln(u_0/u_L)]$, in agreement with the observed behaviour of many metals [7,8].

The Mott–Cabrera-model is also applicable to the anodic formation of oxide films in an electrolyte [7, 8] and LAO was originally explained as a straightforward extension to it. In the vicinity of the probe, the applied bias would lead to a larger voltage V dropped across the oxide. The oxide growth would terminate at a larger final oxide thickness X_L in the affected areas [2, 4].

3.2.2 Beyond the Mott-Cabrera Model

Several observations have called this simple model into question.

1. According to Eq. (3.9), the total oxide thickness including any native oxide layer should be determined by the bias voltage. Experimentally, a strong sensitivity to the thickness of the native oxide has not been seen, and on Si, features of similar height have been grown on several nm of thermal oxide as on the thin native oxide using the same (pulsed) bias [4].

2. An increase in the limiting thickness is observed when a modulated bias is used. The magnitude of the increase depends on the material. This effect cannot be explained by the Mott–Cabrera-model of anodic oxidation [3, 4, 13].

3. The very high initial growth rates of tens or hundreds of nm/s observed for LAO indicate a high carrier concentration that is not compatible with MOTT's assumption of negligible space charge and E, n_s independent of x [4].

VERWEY's hopping model of ionic transport can be extended to include the effect of space charge and forgo the steady-state approximation which assumes that n_s is independent of time. Numerical calculations indicate that the effect of a non-negligible space charge is to reduce the ionic current through the oxide [9]. DAGATA [4] suggests that the rapid decline in the oxidation rate should not be attributed to the oxide thickness but to the effect of such built-up space charges. This is consistent with the observed effect of a modulated voltage: Depending on the migrating ion species present in the oxide, the modulation of the field can allow recombination to occur during each cycle and prevent the accumulation of space charge.

3.3 Background

3.3.1 Instrumentation

LAO was originally demonstrated on hydrogen-passivated Si surfaces with an STM operating in air [1]. Anodization of silicon and metals with an STM has been explored by a number of studies [14, 15]. However, in the STM the tunnelling current is the controlling parameter for imaging. Consequently, anodized regions, which have a higher resistance, appear as indentations since the feedback loop has to decrease the tip–sample distance to maintain a constant current. This does not reflect the topography of the sample, as the lower density of the oxide causes these regions to become higher than the surrounding substrate. The interplay between the measurement of electrical and topographical properties not only makes *in*

situ assessment of the grown oxide unreliable, to avoid tip–sample contact and damage to the probe operating conditions have to be chosen that result in ineffective anodization. A large tip–sample distance is needed, which results in limited field enhancement, so that the oxide features are thin and the operating speed is low [15].

The AFM, in contrast, separates exposure and imaging processes and is therefore now the preferred instrument for LAO under ambient conditions [5, 15]. Operation of the AFM as a DFM, *i.e.*, in non-contact or tapping mode, is possible and has empirically been found to yield better resolution and reproducibility than operation in contact mode [3, 15, 16]. This is explained with the smaller forces between probe and sample and consequently reduced wear on the tip. However, the electrostatic force between the probe and the sample (BELAIDI *et al.* [17] have considered this subject in more detail) has a much larger effect on the feedback loop of a DFM than in the case of an AFM operating in the steep repulsive part of the tip–sample potential. Both the average tip–sample distance and the oscillation amplitude decrease as a result of the applied bias [3]. Suitable feedback settings for exposure and imaging are therefore different, and optimal operation might require disconnecting the feedback loop altogether. In this case, it is necessary to explicitly compensate for the sample topography if long continuous oxide lines (as opposed to single oxide dots) are to be created [18].

Setting up a bias voltage between the apex of the tip and the sample surface requires a conducting probe; as the resolution that can be achieved with LAO depends to a large extent on the radius of curvature of the tip, sharp probes are required. Practical considerations favour n^+-Si tips, which can be obtained with an apex radius as small as 10 nm and typically have a resistivity of the order 0·1 Ω m. Tips coated with thin (5 to 100 nm) layers of diamond [10], titanium [2, 3, 10, 15], cobalt [19], gold [16, 20], or platinum [15] have also been used. Coatings reduce the sharpness of the tip and result in a less well-defined apex. Very thin metal layers may readily wear off, even if normal operating conditions keeps the tip–sample force small [10, 16]. On the other hand, metal films improve the conductivity of the probe and protect it from oxidation and, to a limited extent, wear. Consequently, some authors show a preference for coated tips [3, 15].

3.3.2 Materials

Most of the work on LAO originally concentrated on silicon surfaces [1–5,15,16,20,21]; it was noted that hydrogen passivation of Si in HF actually hinders the anodization process and optimal results are obtained with a thin layer of native oxide [2,21]. Minimal feature widths of approximately 40 nm [20] and maximal feature heights of 5 to 6 nm [21] are typical for the LAO of silicon surfaces. Silicon forms a stable oxide, which may be used as an etch mask for pattern transfer [5,15].

Anodization with an AFM has also been demonstrated for thin titanium [2,4,10,13,15,22,23] and aluminium [18, 24, 25] layers. LAO of titanium creates an insulating oxide, and a thin titanium layer can be oxidized completely to form nanoscopic electronic devices. Oxides as narrow as 30 nm and fully penetrating a 7 nm thick layer of Ti have been reported. Using *in situ* monitoring techniques to control the resistance of a metal strip remaining between two oxides, line widths of 10 nm have been realized [22]. A particularly large increase in the oxide height with modulated bias voltage, up to a factor of 4, is seen on Ti [4]. LAO of aluminium layers has been proposed as a means of creating robust etch masks on Si and SiO_2 [24]. Both Al and Al_2O_3 can be removed selectively, allowing for a positive and negative process. This idea will be revisited in Secs. 3.7 and 4.2.4. The full oxidation of Al layers 7 nm thick has been observed, producing features with a minimal width of approximately 10 nm and a height of approximately 3·5 nm.

By oxidizing the GaAs capping layer with an AFM, one can locally deplete the quasi two-dimensional electron gas contained in a shallow δ-doped GaAs–$Al_xGa_{1-x}As$ double heterojunction. The effect has been used to confine the electrons to one- or zero-dimensional areas and to form other nanoscale devices, including antidot lattices such as those discussed in Chapter 7 [6,23,26,27]. SASA *et al.* [28–30] have used LAO to modulate a InAs–GaSb heterostructure either by modifying the InAs layer directly or by completely anodizing a thin GaSb or $Al_xGa_{1-x}Sb$ capping layer and dissolving the oxide in water. Complete oxidation of GaSb caps with a thickness of up to 10 nm has proved possible, and line widths of 200 nm and 100 nm have been achieved on GaSb and InAs, respectively.

3.4 Experimental Setup

A ThermoMicroscopes (now Veeco) AutoProbe M5 SPM was employed for anodization experiments. This instrument uses a scanner operating in a closed loop configuration with a displacement detector to compensate for actuator drift and hysteresis. It is therefore suitable for writing large lithographic patterns, which can take several hours to complete. The SPM was installed on an air table equipped with active vibration cancellation. Air table and microscope were kept in an acoustic enclosure.

The assembly was located in a low-grade clean room; a commercial humidifier was used to control the relative humidity of the entire room. This arrangement was found to provide higher achievable humidity and better stability than an alternative setup based on a bubbler feeding a controlled mixture of dry and humidified air to the acoustic enclosure only.

I evaluated a number of different probes based on sharpened silicon tips (Veeco Ultralever A and B). According to the manufacturer, these tips had a typical radius of curvature of 10 nm and a tip angle of 12°. They were attached to cantilevers with a force constant of 2·1 N m^{-1} (3·2 N m^{-1}) and resonant frequency of 80 kHz (90 kHz). The tips and cantilevers were made from Boron doped n^+-Si with a resistivity of 0·1 Ω m. Apart from bare probes, tips coated with approximately 50 nm of cobalt and 20 nm to 40 nm of gold were tested. Cobalt coated tips, intended for magnetic force microscopy (MFM), were previously found to be useful for LAO [19] and could be purchased from the supplier. Gold coated tips were manufactured in the Clarendon Laboratory from uncoated Si tips in an evaporator.

Samples were fixed to the grounded x–y-stage with adhesive carbon pads. In some experiments, the top of the sample was additionally connected to the stage with the help of a metal clip or conductive silver paint. A bias was applied between ground and the probe. For optimal flexibility, the chip carrying the tip was contacted independently of the standard instrument setup and connected to a computer-controlled relay. The relay switched between ground and an external input, to which a constant voltage supply or function generator could be attached. In line with earlier work [4, 15, 16] a current limiting resistor was connected in

series with the AFM probe. During normal operation, the voltage dropped across a 1 MΩ resistor can be neglected; should the tip make direct contact with the substrate, the damage induced by local Joule heating is limited, as at most 30 µA can flow at a bias of −30 V. All anodization experiments were performed at or slightly above room temperature (28 ± 3 °C).

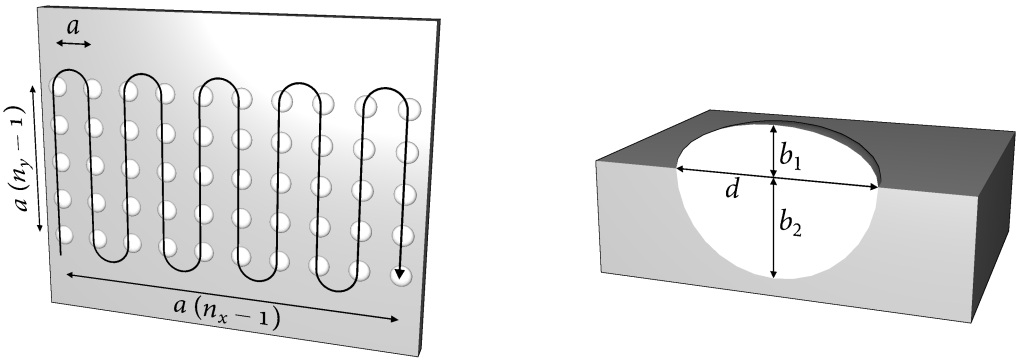
3.5 Antidot Patterns

3.5.1 Pattern Geometry

The objective of my AFM lithography work was to create a two-dimensional lateral modulation of the two-dimensional electron and hole gases present in a InAs–GaSb double heterostructure as described in Chapter 7. This required the creation of a lattice of sample modifications with well defined unit vectors **a** and **b** and individual diameters d, where $|\mathbf{a}|$ and $|\mathbf{b}|$ are of the order of tens to hundreds of nanometres. Although the creation of rectangular and hexagonal lattices using LAO was demonstrated in principle, the samples considered in this thesis are restricted to square lattices with a lattice constant a so that $\mathbf{a} = a\hat{\mathbf{x}}$ and $\mathbf{b} = a\hat{\mathbf{y}}$.

Although the problem of small antidot lattices with a total size smaller than the coherence length of the carriers in the sample is interesting in its own right [6], this was not the motivation of the present work. In order to measure the electrical properties of the patterned heterostructure, the objective was, instead, to modify the entire area of a Hall bar such as could be easily defined with good reproducibility using standard optical lithography (see Chapter 4). In practice, this means an area of approximately 5 µm × 15 µm or a total of 7,500 dots for a square lattice with $a = 100$ nm. Assuming a realistic oxidation time of 0·5 s per dot, a tip speed of 1 µm/s, and a meandering tip trajectory as shown in Fig. 3.2(a), the exposure of such an area takes approximately 1 h 15 min.

As discussed in Sec. 2.3, the piezoelectric scanner of an AFM is subject to a number of effects that compromise its ability to reliably visit a specific absolute position. These scanner limitations are particularly important when reproducibility over the length and time scales



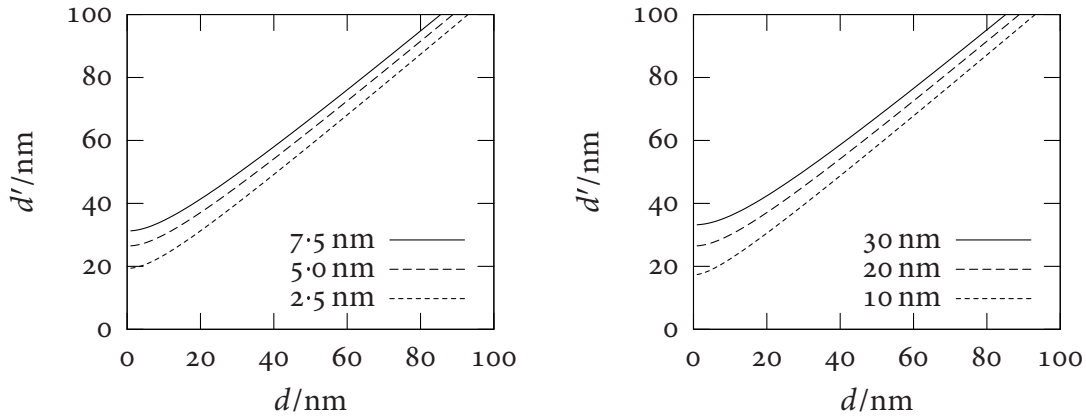
(a) Writing an antidot pattern (b) Simple geometric model of an oxide dot

Figure 3.2: Geometrical considerations for antidot patterns

relevant for the present antidot patterns is required. Operating the scanner in a closed loop setup with independent detectors avoids such problems to a large extent; this is explained in more detail in Sec. 2.3.2. Moreover, as illustrated in Fig. 3.2(a), the smaller extent of a rectangular pattern was chosen as the fast scanning direction for writing if practical, thereby reducing the shift between adjacent rows caused by slow drift. I successfully created patterns consisting of up to 7,500 dots with a local distortion of the unit cell from the nominal shape smaller than the uncertainty of the individual dot positions. No attempt was made to verify the correlation between dot positions in distant parts of the pattern. As the size of the entire pattern is considerably larger than the mean free path of the carriers in the sample, such long range distortions are not expected to significantly affect the electronic properties.

3.5.2 Dot Geometry

The precise shape of the oxide dots is not known. Since the size of the dots is inherently of the same order of magnitude as the radius of curvature of the AFM tip, convolution effects (*cf.* Sec. 2.4.8) are important. Consequently the dot diameter is generally overestimated and the dot profile cannot be determined accurately. Even so, it is reasonable to assume that dots are rotationally symmetric; considering the electric field between a spherical tip and a flat surface [17] furthermore suggests more oxide is produced at the centre and the height of the



(a) different oxide heights

(b) different tip radii

Figure 3.3: Apparent oxide dot diameter in AFM micrographs, calculated for spheroidal dots with diameter d and height b_1 and idealized tips with a uniform radius of curvature r. The tip radius in (a) is $r = 20\,\mathrm{nm}$ and the dot height in (b) is $b_1 = 5\,\mathrm{nm}$.

dot gradually decreases towards the perimeter. Both assumptions are supported well by AFM measurements within the limitations laid out above.

The dots are most easily described by two half spheroids as illustrated in Fig. 3.2(b). Based on this approximation, it is possible to calculate the size of the dots when imaged with an idealized AFM tip of finite size. The relationship between apparent dot diameter d' and assumed real diameter d is illustrated in Fig. 3.3 for realistic oxide dot and tip sizes.[5] The usefulness of these calculations is limited in so far as the tip size is not generally known exactly and the assumption of a uniform tip radius of curvature may not be accurate.

In the spheroid model of Fig. 3.2(b), the volume of the oxide dot is given by $V_\mathrm{o} = \frac{1}{6}\pi(b_1 + b_2)d^2$, whereas the volume of the oxidized material is $V_\mathrm{m} = \frac{1}{6}\pi b_2 d^2$. If v_o and v_m are the molar volumes of the oxide and the unoxidized material, oxidation of one mole of the substrate yields q moles of the oxide, and the thickness of the native oxide layer is neglected, it follows that

$$\frac{b_2}{b_1} = \left(q\frac{v_\mathrm{o}}{v_\mathrm{m}} - 1\right)^{-1}. \tag{3.10}$$

[5] The calculation is straightforward, but the result involves the solution of a quartic equation, the presentation of which in a closed form is not useful.

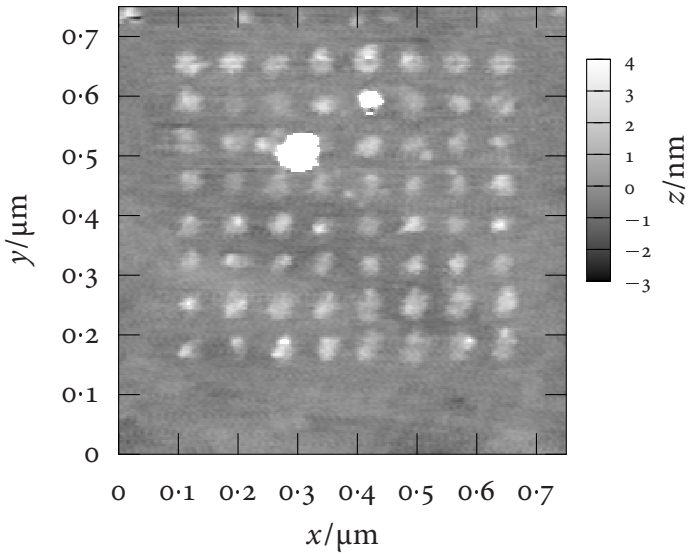


Figure 3.4: AFM micrograph of a square oxide dot array on the free GaSb surface of a GaSb–InAs–GaSb DHET with a nominal lattice constant of $a = 75$ nm. The dots have an average diameter of 30 ± 5 nm and a height of 4 ± 1 nm. They were created using a gold coated silicon tip by applying a bias of -15 V for 125 ms per dot.

Conversely,

$$q\frac{v_\mathrm{o}}{v_\mathrm{m}} = \frac{b_1}{b_2} + 1. \tag{3.11}$$

These relations hold in fact true for any dot shape as long as $V_\mathrm{upper} = c_1 b_1 d^2$ and $V_\mathrm{lower} = c_2 b_2 d^2$ with $c_1 = c_2$. In particular, if this condition is met the height increase is independent of d and the same as would be expected for a flat surface.

3.6 Local Anodic Oxidation of III-V Semiconductors

3.6.1 Observations

Performance

LAO of GaSb allowed reliable drawing of large arrays of oxide dots with diameters down to 30 ± 5 nm and heights of up to 8 ± 1 nm. Lattices with spacings as small as $a = 75$ nm were realized without any indication of individual sites interfering with each other (Fig. 3.4). The

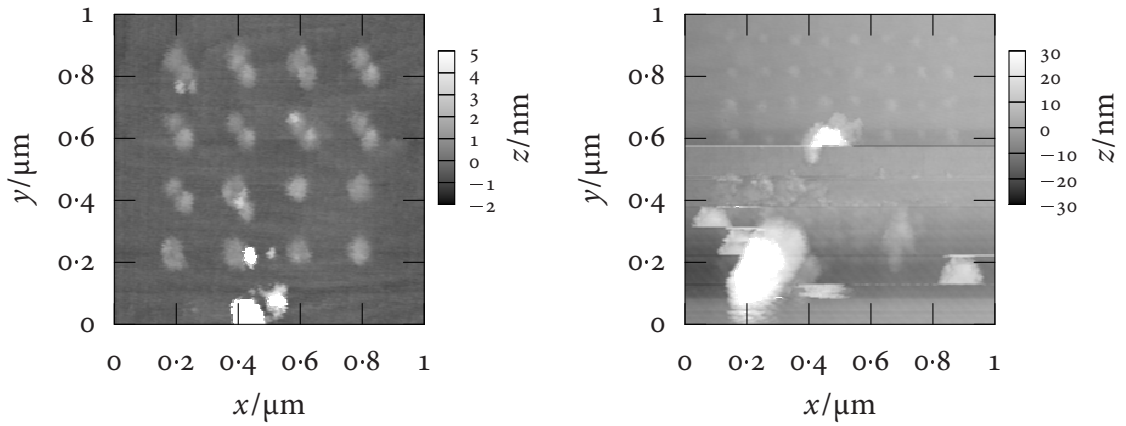
resolution depends on both the tip and the operating conditions; in practice, the tip determined the achievable resolution, while the operating conditions—within the range suitable for reasonably reliable operation—only had a small influence on the feature size. A limited number of experiments were performed on InAs and GaAs for comparison. If the differences between tips were taken into account, LAO performance was similar on these materials, and no significant difference in the lateral size of oxide dots was observed. Anodization of both InAs and GaAs was generally more reliable than that of GaSb, and oxide dots exceeding 10 nm in height could repeatably be obtained.

Anodic oxide of GaSb can be dissolved in water or dilute hydrochloric acid without attacking the bulk semiconductor [28]. Deoxidation revealed a depth b_2 of the oxide dots between $1{\cdot}0b_1$ and $2{\cdot}0b_1$, which indicated a volume expansion of 1·5 to 2·0 according to Eq. (3.11).

In most cases, lithography was performed on GaSb epilayers grown by metal-organic vapour phase epitaxy (MOVPE) in Oxford within a month of the growth process. Some experiments used MOVPE-grown samples or commercially obtained 'epi-ready' wafers that had been in storage for over a year. Both the height of the oxide features and that of the holes left after oxide removal were comparable for both sets of samples.

The Influence of the Tip

The condition of the probe had the most pronounced effect on the shape and size of the oxide features. In many cases, tips instantaneously changed in such a way that the diameter of the dots that they created increased in size from approximately 30 nm to several hundred nm. Individual dots would then consist of a number of oxide lumps and a 'rubber stamp' behaviour was observed in which this pattern repeated itself at each dot, as illustrated in Fig. 3.5(a). The repeated features in Fig. 3.5(a) evidently resulted from the shape of the tip during anodization, not during imaging: When the shape changed, it did so from one *dot* to the next, not from one (horizontal) *scan line* to the next, as would be expected if the effect were caused during data acquisition. While shadowing was sometimes observed, imaging properties were more often not noticeably affected. On occasion, tips also failed catastroph-



(a) Rubber stamp effect

(b) Tip contamination

Figure 3.5: AFM micrographs of oxide dot arrays on GaSb, showing the effect of tip contamination on LAO. (a) A square array with a lattice constant of 200 nm. The tip was damaged during use and then produced large dots with a complex shape, creating a 'rubber stamp' effect. At the bottom of the plot reversal of the tip change can be observed. (b) A square array with a lattice constant of 100 nm. The tip failed catastrophically halfway through the run, which proceeded from the top to the bottom of the image.

ically, as shown in Fig. 3.5(b). This was accompanied by the creation of large mounds and holes of sizes up to a few micron. Afterwards, these tips either continued to produce such features or created very wide dots indicative of blunt tip apices, which was consistent with the observed imaging artefacts. Finally, some tips started to create increasingly low dots and ultimately ceased to produce any oxide features at all. Such probes occasionally reverted to a usable condition either spontaneously or after forcefully making contact with the sample surface.

Uncoated silicon tips were less likely to lose their lithographic resolution; the behaviour of different new tips was consistent, with some variation in the lithographic resolution. However, they would often stop working after a few days of use. Cobalt or gold coated probes were more prone to instantaneous failure, but less susceptible to aging, which was however still observed for some probes. A gold coating of 10 to 20 nm did not significantly affect either the resolution or the speed at which the oxide grew. Commercially obtained cobalt

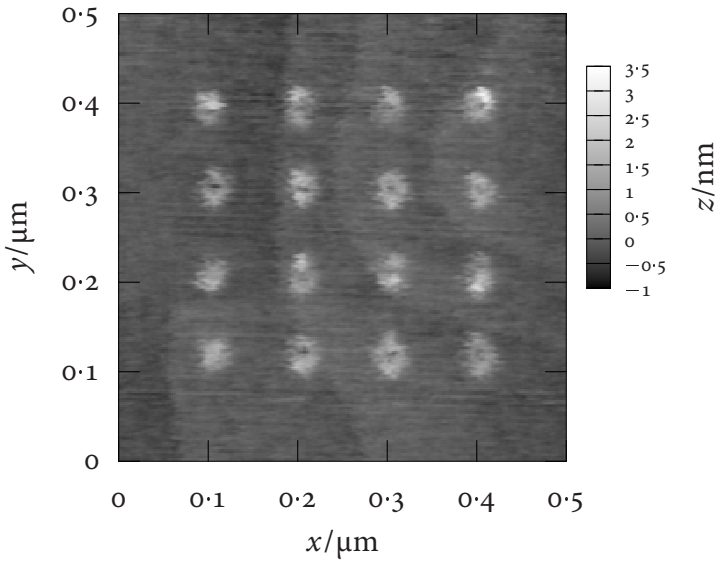


Figure 3.6: AFM micrograph of a square oxide dot array on GaSb with a nominal lattice constant of $a = 100\,\text{nm}$, showing the central indentations observed on some oxide dots. The curved lines correspond to atomic terraces.

coated tips, on the other hand, had on average a worse resolution and required up to 3 s per dot for the oxide height to saturate, as opposed to less than 1 s for the gold coated variant. All probe varieties were used successfully in both contact mode and non-contact mode (NCM). It sometimes occurred that a particular tip ceased to produce oxide features in NCM, but could still be used in contact mode.

Oxide dots frequently showed an indentation at their centre, as demonstrated in Fig. 3.6. Such an indentation could be observed for all substrate materials, but was more pronounced on GaSb. The effect was generally associated with tips that produced wide dots but had a good imaging resolution.

Current and Voltage

A straightforward check of the current between tip and sample was made using a source measure unit (SMU) as the pulse generator. Although the current was normally less than the

detection limit of approximately 10 nA,[6] currents in excess of 10 µA could be drawn if the tip failed or came in contact with the side wall of a large surface feature.

Depending on the tip and operating parameters, there was a threshold voltage in the range 6 to 10 V (negative tip) below which oxide formation did not occur. For voltages exceeding 20 V, the probability of tip failure increased rapidly, which put an effective limit on the oxide thickness that could be achieved with this method.

Motivated by the increased aspect ratio reported as a result of using a modulated bias for some substrate materials, in particular titanium [3, 4], I investigated the effect of modulated voltages on the LAO of GaSb. A square waveform centred around zero with a peak-to-peak voltage difference of 20 V and a duty ratio of 50 % was used at various frequencies up to 1 MHz on its own or in conjunction with a constant offset voltage. When anodizing for twice as long, modulation frequencies up to approximately 200 Hz combined with a negative DC bias on the tip of at least approximately 8 V created oxide dots very similar in size to those obtained with a constant bias. However, no enhancement of the dot height or the aspect ratio was seen. With a small or no DC bias and at high frequencies, the oxide height was significantly *reduced* compared to the unmodulated case. In all experiments with a modulated bias, the uniformity of the oxide dots was improved and the probability of tip failure resulting in a 'stamp effect' was reduced. After such a failure, the tip reverted to a usable state more readily than without the voltage modulation.

3.6.2 Quantitative Results—Dot Patterns on GaSb

Preliminary Remarks

As described in Sec. 3.6.1, the state of the tip is the one parameter that has the most drastic effect on the lithographic performance. In order to gain insight into the influence of the other operational parameters, special care has to be taken to keep the tip as well-defined as

[6] There are several conceptually straightforward ways to improve the sensitivity, analogous to the current measurement in an STM. As currents during LAO were previously found to be negligible [10, 16] and not a controlling parameter for oxidation [15], I did not explore these options.

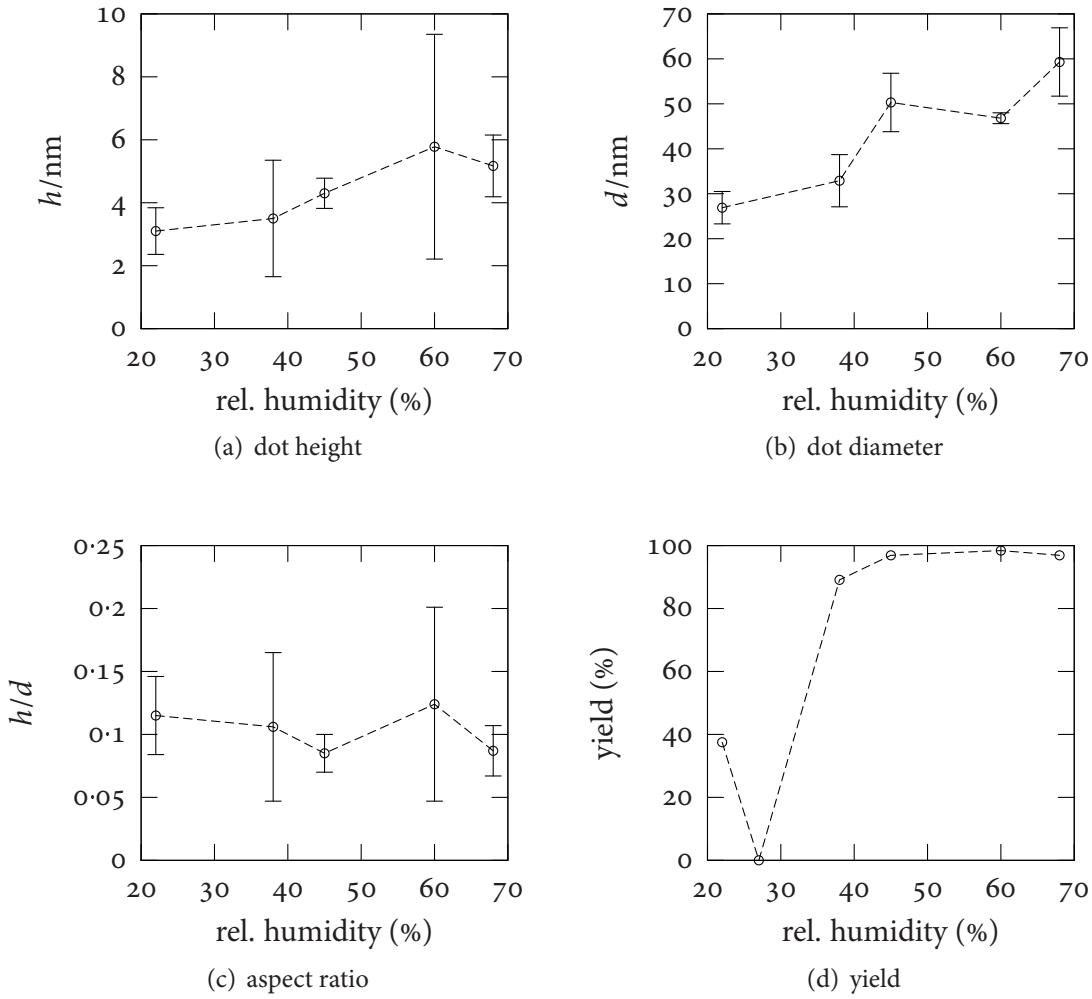
possible during the experiment. Even so, the possibility that a change in the tip configuration has happened between measurements cannot be excluded and needs to be considered in the interpretation of the results. Moreover, the numerical values for dot diameters, and, to a lesser degree, heights, should be expected to differ among different tips.

The results in this section were obtained with an unused silicon tip (Veeco Ultralever A) on a polished GaSb wafer intended for epitaxial growth. I have selected the series presented here for consistency and because the absence of obvious imaging artefacts indicates a relatively unchanged tip; it does not represent optimal lithographic performance.

To determine the height of each dot, the mean height of the surface—excluding the patterned area—was subtracted from the highest value measured within the dot. The height is not expected to be affected significantly by systematic errors inherent to the AFM, because it is much larger than the scale at which the distance dependence of the tip-sample interaction is relevant and much smaller than the scale at which scanner nonlinearity becomes important. In this range, the instrument can be calibrated reliably. The diameter of a dot at its base was estimated by counting the number of pixels lying higher than two standard deviations of the unpatterned surface and assuming a circular geometry—which was generally well supported by the data. Unlike the height, the measurement of the dot diameter *is* affected by the finite size of the tip—inherently of the same order of magnitude as the dot!—and can only be regarded as a upper limit of the actual size. The problem is discussed in more detail in Sec. 2.4.8; it should result in an approximately constant offset. Finally, the yield was defined as the number of oxidation sites actually producing an oxide dot.

Ambient Humidity

As seen in Fig. 3.7(a–c), both the dot height h and the dot diameter d increased with increasing humidity, while no significant change in the aspect ratio d/h could be demonstrated. The increase could be quantified as 0·47 nm per 10 % for the height and 5·3 nm per 10 % for the diameter, but the limited nature of the data did not allow me to accurately prove or disprove linearity. The most pronounced effect of the humidity was on the success rate shown in



(a) dot height (b) dot diameter

(c) aspect ratio (d) yield

Figure 3.7: Humidity dependence of oxide dot growth on GaSb. 64 oxide dots in a square lattice with lattice constant $a = 300\,\text{nm}$ were grown by applying a bias of $-15\,\text{V}$ to the tip for 2 s. The AFM was operated in 'non-contact' mode with a set point of $-0{\cdot}02$. The data points have been connected with straight lines to aid visualization.

Fig. 3.7(d): At low ambient humidity, only some of the anodization sites showed the development of an oxide dot, whereas the yield approached 100 % for relative humidities exceeding approximately 45 %. The behaviour of the series shown here is consistent with observations made using different tips and operating parameters, although the threshold behaviour was not always as clear.

Bias Voltage

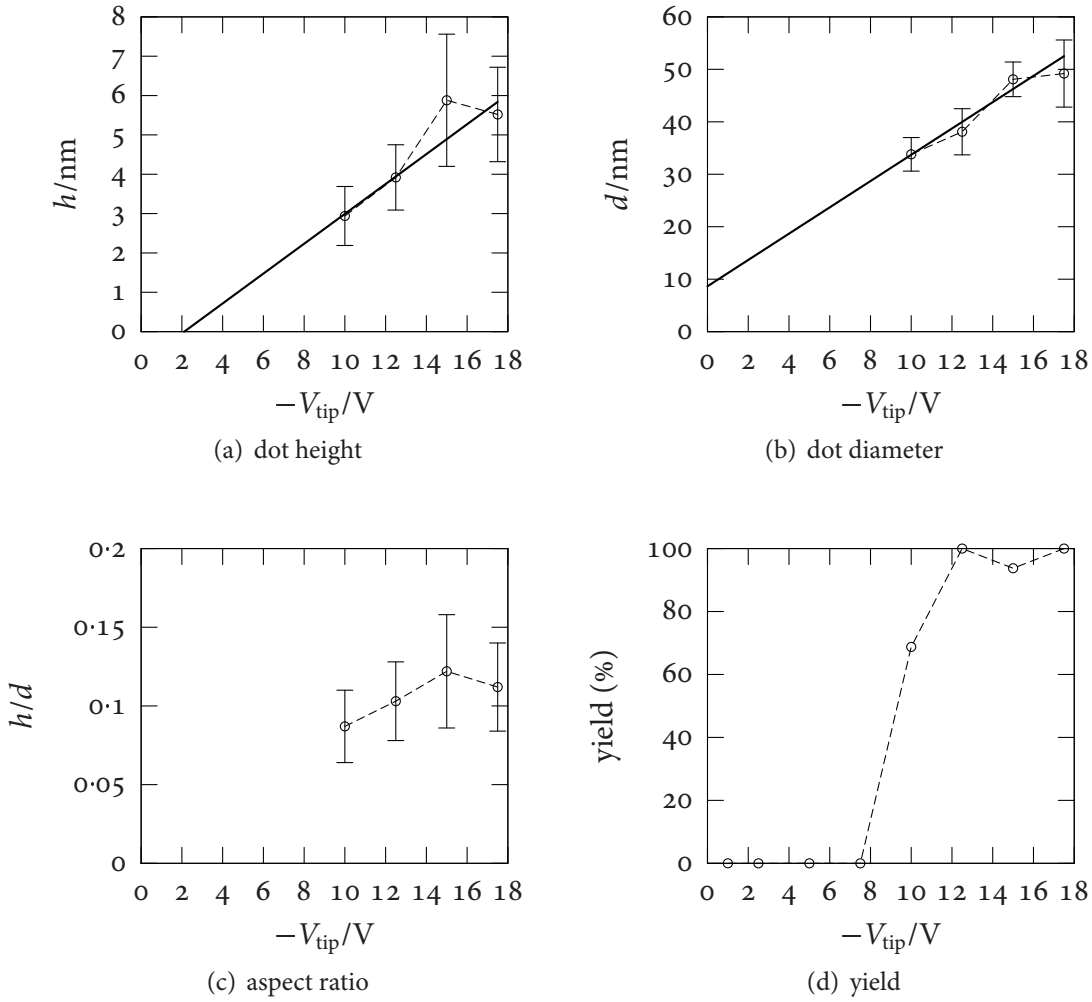
The dependence on the (negative) bias applied to the tip with respect to the sample is illustrated in Fig. 3.8. Both height and diameter of the oxide dots increased approximately linearly with voltage; the height increased by 0·38 nm per V as seen in Fig. 3.8(a), while the diameter, shown by Fig. 3.8(b), was best fit by a line with a slope of 2·51 nm/V. There was a small increase in the aspect ratio, Fig. 3.8(c), with increasing bias. As seen in Fig. 3.8(d), the potential difference had to exceed a threshold of 8 ± 2 V before anodization was observed.

Anodization Time

The dependence of the oxide formation on the anodization time (Fig. 3.9) was characterized by a very high initial growth rate, which decreased steeply with increasing time. The dot height h and the diameter d are shown in Figs. 3.9(a) and 3.9(b). Both quantities appeared to saturate for $t \gtrsim 1$ s. No significant effect on the aspect ratio, Fig. 3.9(c), was detected, and the success rate for oxide formation, Fig. 3.9(d), remained close to 100 % even for short anodization times.

Tip–Sample Distance

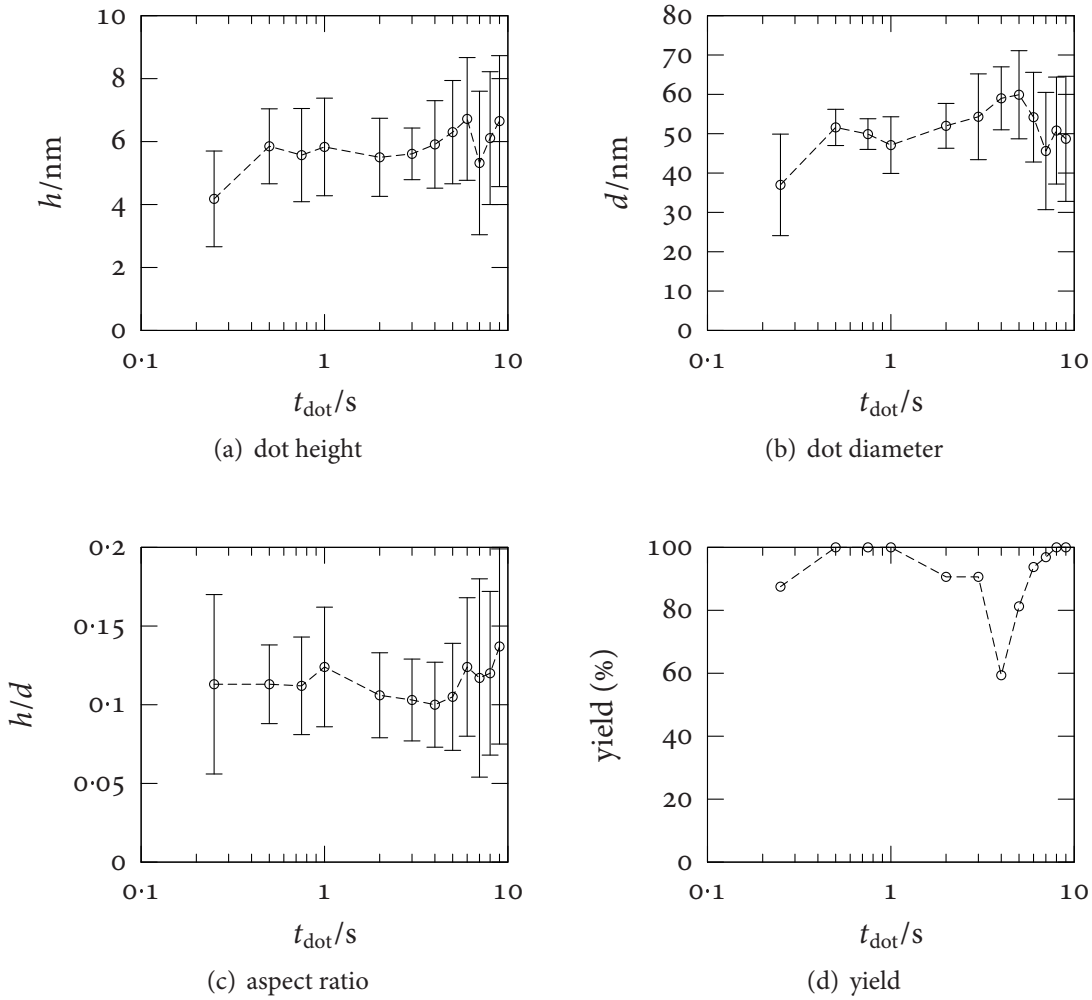
The parameter controlling the tip–sample distance in an AFM operating as an amplitude modulation DFM is the amplitude reduction factor (*cf.* Sec. 2.4.5). For the instrument used in the present study, a large negative value of the set point corresponded to a large tip–sample



(a) dot height (b) dot diameter

(c) aspect ratio (d) yield

Figure 3.8: Voltage dependence of oxide dot growth on GaSb. Oxide dots in a square lattice with lattice constant $a = 300\,\text{nm}$ were grown at 45 % relative humidity for 2 s. Each data point was averaged over 32 dots. The AFM was operated in 'non-contact' mode with a set point of −0·02. The data points have been connected with straight lines to aid visualization. Thick solid lines are linear fits to the data.



(a) dot height (b) dot diameter

(c) aspect ratio (d) yield

Figure 3.9: Anodization time dependence of oxide dot growth on GaSb. Oxide dots in a square lattice with lattice constant $a = 300\,\text{nm}$ were grown at 45 % relative humidity by applying a bias of $-15\,\text{V}$ to the tip for 2 s. Each data point was averaged over 32 dots. The AFM was operated in 'non-contact' mode with a set point of $-0{\cdot}02$. The data points have been connected with straight lines to aid visualization.

distance. The exact relationship between the set point and the distance depends on the tip-sample interaction and the elastic properties of the cantilever. While the *cantilever* position was fixed during anodization, both the oscillation amplitude and the mean tip–sample distance were affected by the electrostatic force resulting from the bias voltage.

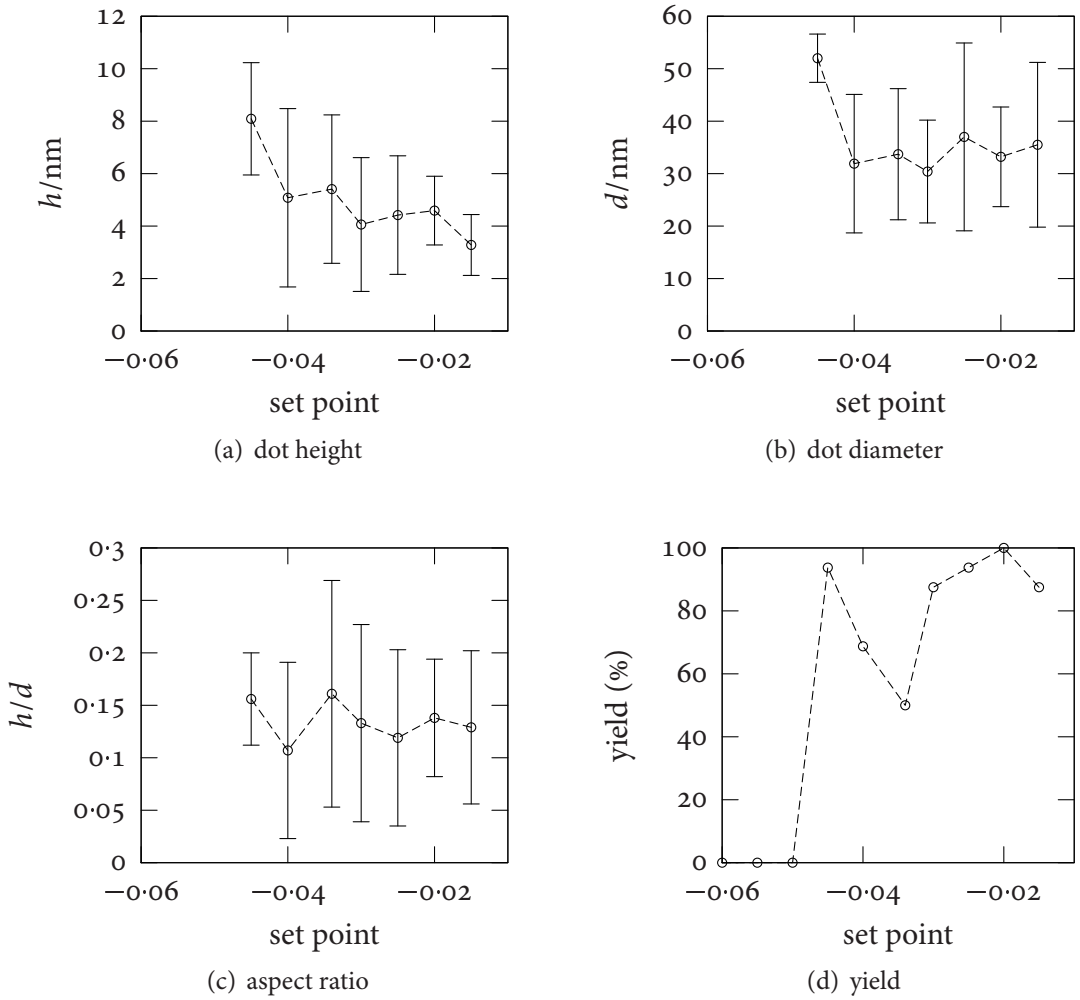
No systematic dependence of the shape and size of the oxide dots on the set point (and hence on the *initial* tip–sample distance) was detected (Fig. 3.10). As seen in Fig. 3.10(d), no oxide dots were formed for large tip-sample distances and a small (negative) set point was required for reliable operation. All set point values used in this study could be used to successfully obtain micrographs of previously grown oxide dots.

3.6.3 Discussion

The resolution of the oxide features is sufficient to create antidot patterns with a period as small as 75 nm. Height and resolution compare favourably with the values reported previously for the LAO of GaSb [28].

The composition of the anodic oxide is not known exactly. While gallium oxide generally occurs as Ga_2O_3, antimony can have different oxidation states and may form Sb_2O_3, Sb_2O_5, or a mixture thereof. While thermal oxidation of GaSb yields a mixture of Ga_2O_3 and Sb_2O_3, the native oxide contains predominantly Ga_2O_3 and Sb_2O_5 [31]. GaSb has a molar volume of approximately $34{\cdot}1\,\mathrm{cm}^{-3}$ [32] and Ga_2O_3 one of $31{\cdot}2\,\mathrm{cm}^{-3}$; the molar densities of Sb_2O_3 and Sb_2O_5 are $52{\cdot}2\,\mathrm{cm}^{-3}$ and $85{\cdot}6\,\mathrm{cm}^{-3}$ respectively [33]. Using these bulk values, one predicts a volume expansion factor between 1·2 for Ga_2O_3/Sb_2O_3 and 1·7 for Ga_2O_3/Sb_2O_5, which is consistent with the range of 1·5 to 2·0 observed experimentally. An oxide mixture containing predominantly Sb_2O_5 appears to be indicated by these calculations, but the volume expansion may be further affected by differences from the bulk crystal structure of the oxides and by incorporation of water.

The sudden changes in the dot shape are related to changes in the geometrical configuration of the actual tip, as is illustrated by the observed 'rubber stamp' effect. The re-



(a) dot height (b) dot diameter

(c) aspect ratio (d) yield

Figure 3.10: Set point dependence of oxide dot growth on GaSb. Oxide dots in a square lattice with lattice constant $a = 300\,\mathrm{nm}$ were grown at 45 % relative humidity by applying a bias of −15 V to the tip for 2 s. Each data point was averaged over 16 dots. The data points have been connected with straight lines to aid visualization.

versibility of this phenomenon, the frequent lack of an effect on the imaging resolution, and the size of the features call into question damage to the tips. Instead I propose that the probes pick up contaminants such as oxide particles from the surface. These contaminants could strongly influence the field distribution and the shape of the water meniscus, while imaging would be affected to a lesser degree if they do not touch the surface or their mechanical coupling to the tip is weak. The surface coating affects the ability of tips to pick up contaminants. Catastrophic tip failure is associated with a noticeable increase in current and can therefore be explained by electric breakdown and damage due to Joule heating. Loss of the ability to create any oxide can be attributed to the formation of a non-conducting layer on the tip by oxidation or contamination: Then most of the voltage is dropped across this layer and the tip–sample potential difference falls below the threshold required for anodization. Metal-coated tips can still be affected as the thin metal film can wear off. Both blunted and oxidized tips may rarely revert to a usable condition if collision with the surface chips off part of the tip and a fresh, sharp apex is exposed.

Central indentations have previously been observed in both dots and lines grown using LAO on different materials [4]. Their origin is not fully understood; an explanation based on tip shape can be excluded as a sufficiently blunt tip would be unable to later resolve the observed indentations, which is not in agreement with experiment. Instead, they are likely to be related to a change in the oxide composition at the centre. This is consistent with the observation of DAGATA *et al.*, who see a similar ring structure in the polarizability of oxide dots even in the absence of a corresponding topographical feature.

The lack of success in forming individual oxide dots that is seen for low humidities and large tip–sample distances does not significantly correlate with a decrease of the size of the oxide features. The total absence of oxide dots on some sites is attributed to a failure to form a water meniscus, which is either caused by a thin or non-uniform water layer on the substrate or a large tip–sample separation when the bias is switched on.

3.7 Local Anodic Oxidation of Thin Aluminium Layers

3.7.1 Overview

I have performed a limited number of experiments to assess the possibility of patterning thin aluminium films using LAO. This allows to use the AFM to create etch masks for reactive ion etching as described in Sec. 4.2.4, thus overcoming the inherent limitation to shallow features.

3.7.2 Performance

Using the procedure employed for the anodization of GaSb surfaces, features could successfully be drawn on thin (7 to 40 nm) aluminium layers created by evaporation on either GaSb or SiO_2. As aluminium rapidly forms a native oxide in contact with air, LAO was performed within a week of evaporation on these samples. For each dot, a pulsed bias of −20 V was applied for 0·5 to 2 s to a conducting silicon tip tracing the sample topography in non-contact mode after disconnecting the feedback loop. For high-quality oxide layers, a reference potential was provided by contacting the metallization separately.

In this way, large dot arrays could be created with good reproducibility as illustrated in Fig. 3.11. The smallest dot diameters and largest dot heights that could reliably be realized were 100±10 nm and 10±1 nm. Individual dots with diameters comparable to those achieved on GaSb could be created. The dot geometry depended on the specific tip used for anodization. The occurrence of repeating features indicated that both the imaging and the lithographic resolution was limited by finite tip size effects.

3.7.3 Oxide Removal

The aluminium oxide created by anodization with the AFM can be etched selectively in a mixture of 85 % phosphoric acid, deionized water, and chromiumtrioxide [24]. Fig. 3.12 shows a cross-section through the same six oxide dots before and after removal of the oxide.

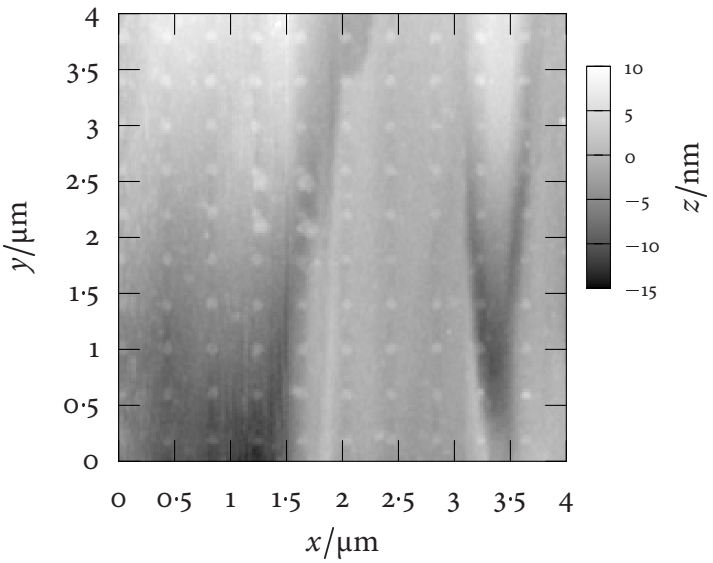
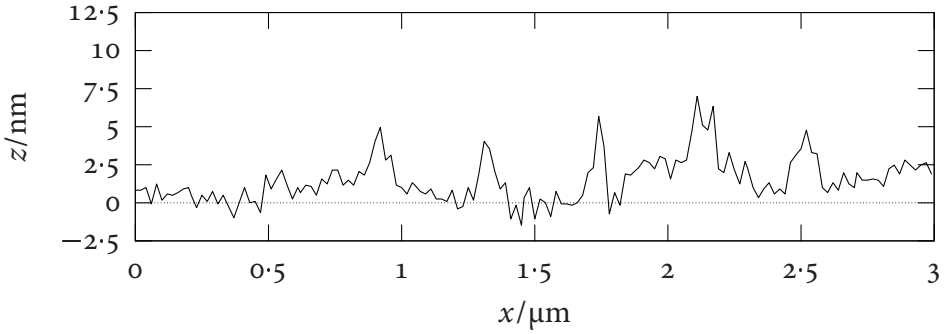


Figure 3.11: AFM micrograph of oxide dots created on aluminium by LAO. The array has a lattice constant of $a = 400\,\text{nm}$. The dots have an average diameter of $120 \pm 20\,\text{nm}$ and a height of $3 \pm 1\,\text{nm}$. They were created using a conducting silicon tip by applying a pulsed bias voltage of $-20\,\text{V}$. The corrugation is caused by the underlying DHET, on which the aluminium was evaporated. The four additional dots near the centre were drawn separately.

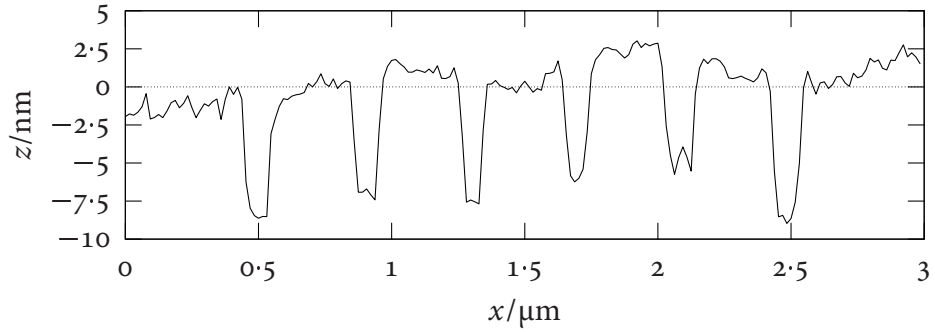
Comparatively large oxide dots with a period of 400 nm and a diameter of $140 \pm 10\,\text{nm}$ were chosen to minimize the impact of artifacts caused by a finite tip size, *e.g.*, as a result of incomplete penetration of narrow holes. The oxide dots had a height of $4 \pm 1\,\text{nm}$ and left holes with an estimated depth of $8 \pm 1\,\text{nm}$ and a diameter of $130 \pm 10\,\text{nm}$ after removal of the oxide. It is notable that the holes were much more uniform in their shape and vertical size than the oxide mounds observed before etching. The bottoms of the individual holes appeared flat.

3.7.4 Discussion

Aluminium layers have successfully been patterned in a way analogous to LAO of GaSb. The oxide thickness that has been achieved is similar to the values reported previously [18,24,25]. The lateral resolution is less than either obtained on GaSb in the context of the present work or achieved by a comparable method on Al before [18]. It appears to be limited by finite



(a) Dots created using LAO



(b) After selective etching of Al_2O_3

Figure 3.12: Removal of oxide dots created on Al using LAO. Both parts show a surface profile of the same region on the sample. The vertical scale is strongly exaggerated, while identical for both plots.

tip size, even though the same tips are capable of producing smaller features on GaSb. One possible explanation for this is the roughness of the surface on the scale of the radius of the tip apex, which may allow a larger meniscus to form or promote the contamination of the tip by small oxide particles.

According to Eq. (3.11), a hole depth $b_2 \approx 4 \pm 1\,\mathrm{nm}$ corresponding to a dot height $b_1 \approx 8 \pm 1\,\mathrm{nm}$ indicates a volume expansion factor qv_o/v_m of $1{\cdot}5 \pm 0{\cdot}2$, which agrees with the results reported for similar experiments in the literature [18, 24]. Aluminium generally occurs in a single oxidation state, and aluminium (III) oxide has a molar volume of $v_o \approx 25{\cdot}5\,\mathrm{cm}^{-3}$, while solid aluminium has a molar volume of $v_m \approx 10{\cdot}0\,\mathrm{cm}^{-3}$ [33]. Since the chemical formula for aluminium (III) oxide is Al_2O_3, $q = 0{\cdot}5$ and the expected volume expansion factor

for full oxidation is approximately 1·3. Given the uncertainty of the experimental value, this is in reasonable agreement with the observation. However, a different experimental volume expansion can also be attributed to a difference in the structure of the anodic oxide from bulk α-Al_2O_3; in particular, water may be incorporated during oxidation. The presence of a native oxide prior to anodization does not affect this argument, provided the selective oxide removal is performed soon after the LAO step. The native oxide layer is removed in the etch and the remaining hole reflects the volume of metal actually anodized. After the etch, the metal will oxidize again, but the effect on the hole shape is small if the measurement is performed immediately. As the maximal achievable oxide height is approximately 10 nm, complete oxidation of aluminium layers up to 20 nm should be possible.

3.8 Conclusions

An attempt has been made to elucidate the underlying principle of local anodic oxidation (LAO), which uses the electric field created by the negatively biased tip of an SPM to enhance the migration of ions through a native oxide layer and thus locally cause further oxidation. I have also reviewed previous attempts at LAO, which illustrate the fundamental capabilities of the method, especially its ability to modify semiconductor surfaces with a resolution comparable to electron beam lithography (EBL). Considerations for using the technique in the creation of regular dot patterns have been put forward.

I have evaluated LAO of several III–V semiconductors and metals, with a focus on Al and especially GaSb because of their intended use in sample fabrication. The importance of the shape and configuration of the tip apex for lithographic performance has been pointed out, and explanations for failure modes have been suggested. It has been noted that reliability can be improved by modulating the bias voltage. On GaSb and Al, the origin of the observed mounds from oxidation has been conclusively verified by selectively removing the oxide in a suitable etchant and showing that holes remained that possessed a depth compatible with the expected volume expansion.

I have shown that AFM lithography on GaSb surfaces is able to produce oxide dot arrays with lattice periods of less than $a = 75\,\text{nm}$ (Fig. 3.4). The feature size can be as small as 30 nm and oxide heights of up to 10 nm can be achieved. On thin aluminium layers created by evaporation on GaSb or SiO_2, a minimal feature size of 130 nm has been realized. Aluminium films with a thickness of up to 20 nm can in principle be oxidized completely, opening up holes in the layer after oxide removal.

In summary, LAO using an AFM is a lithographic technique that combines a high resolution with features that make it attractive for prototyping. The required instruments are highly accessible and allow for a variety of *in situ* monitoring techniques, most importantly an imaging mode that is independent of the exposure mechanism. Many interesting systems can be modified directly with little or no damage to the surrounding areas. However, LAO suffers from a relatively poor repeatability and reliability, being reliant on a well-defined AFM tip, a condition that can often not be fulfilled as variation between different tips and changes in the tip during operation are quite evident. Moreover, the technique is limited to very shallow modifications of materials that naturally form thin stable oxide films in the presence of oxygen. These restrictions may be circumvented by using a separate sensitive layer and a means of pattern transfer, but in doing so some advantages of the method, especially its simplicity and non-invasiveness may be negated.

Bibliography

[1] J. A. Dagata, J. Schneir, H. H. Harary, C. J. Evans, M. T. Postek, and J. Bennett. Modification of hydrogen-passivated silicon by a scanning tunneling microscope operating in air. *Applied Physics Letters*, 56(20):2001–2003, May 1990.

[2] A. E. Gordon, R. T. Fayfield, D. D. Litfin, and T. K. Higman. Mechanisms of surface anodization produced by scanning probe microscopes. *Journal of Vacuum Science and Technology*, 13(6):2805–2808, November/December 1995.

[3] M. Calleja, J. Anguita, R. García, K. F. Pérez-Murano, and J. A. Dagata. Nanometre-scale oxidation of silicon surfaces by dynamic force microscopy: reproducibility, kinetics and nanofabrication. *Nanotechnology*, 10:34–38, 1998.

[4] J. A. Dagata, T. Inoue, J. Itoh, K. Matsumoto, and H. Yokoyama. Role of space charge in scanned probe oxidation. *Journal of Applied Physics*, 84(12):6891–6900, December 1998.

[5] P. M. Campbell, E. S. Snow, and P. J. McMarr. AFM-based fabrication of Si nanostructures. *Physica B*, 227:315–317, 1996.

[6] A. Dorn, M. Sigrist, A. Fuhrer, T. Ihn, T. Heinzel, K. Ensslin, W. Wegscheider, and M. Bichler. Electric properties of antidot lattices fabricated by atomic force lithography. *Applied Physics Letters*, 80(2):252–254, January 2002.

[7] N. F. Mott. The theory of the formation of protective oxide films on metals.—III. *Transactions of the Faraday Society*, 43:429–434, 1947.

[8] N. Cabrera and N. F. Mott. Theory of the oxidation of metals. *Reports on Progress in Physics*, 12:163–184, 1948.

[9] A. T. Fromhold, Jr. Metal oxidation kinetics from the viewpoint of a physicist: The microscopic motion of charged defects through oxides. *Langmuir*, 3(6):886–896, 1987.

[10] R. Vullers. *Oxidation of Titanium Films with an Atomic Force Microscope*. PhD thesis, Katholieke Universiteit Leuven, 2000.

[11] N. F. Mott. A theory of the formation of protective oxide films on metals. *Transactions of the Faraday Society*, 35:1175–1177, 1939.

[12] N. F. Mott. The theory of the formation of protective oxide films on metals, II. *Transactions of the Faraday Society*, 39:472–483, 1940.

[13] Z. Shen, S. Hou, H. Sun, X. Zhao, and Z. Xue. Local oxidation of titanium thin films using an atomic force microscope under static and pulsed voltages. *Journal of Physics D: Applied Physics*, 37:1357–1361, May 2004.

[14] H. J. Song, M. J. Rack, K. Abugharbieh, S. Y. Lee, V. Khan, D. K. Ferry, and D. R. Allee. 25 nm chromium oxide lines by scanning tunneling lithography in air. *Journal of Vacuum Science and Technology B*, 12(6):3720–3724, November/December 1994.

[15] P. A. Fontaine, E. Dubois, and D. Stiévenard. Characterization of scanning tunneling microscopy and atomic force microscopy-based techniques for nanolithography on hydrogen-passivated silicon. *Journal of Applied Physics*, 84(4):1776–1781, August 1998.

[16] F. Pérez-Murano, G. Abadal, N. Barniol, X. Aymerich, J. Servat, P. Gorostiza, and F. Sanz. Nanometer-scale oxidation of Si(100) surfaces by tapping mode atomic force microscopy. *Journal of Applied Physics*, 78(11):6797–6801, December 1995.

[17] S. Belaidi, P. Girard, and G. Lévêque. Electrostatic forces acting on the tip in atomic force microscopy: Modelization and comparison with analytic expressions. *Journal of Applied Physics*, 81(3):1023–1030, February 1996.

[18] Z. J. Davis, G. Abadal, O. Hansen, X. Borisé, B. Barnibol, F. Pérez-Murano, and A. Boisen. AFM lithography of aluminum for fabrication of nanomechanical systems. *Ultramicroscopy*, 97:467–472, 2003.

[19] G. A. C. Jones. Private communication, 2002.

[20] T. Hattori, Y. Ejiri, K. Saito, and M. Yasutake. Fabrication of nanometer-scale structures using atomic force microscope with conducting probe. *Journal of Vacuum Science and Technology A*, 12(4):2586–2590, July/August 1994.

[21] T. Fayfield and T. K. Higman. Fabrication and transport measurements of atomic force microscope modified silicon metal–oxide–semiconductor field-effect transistors. *Journal of Vacuum Science and Technology B*, 13(3):1285–1289, May/June 1995.

[22] E. S. Snow and P. M. Campbell. AFM-fabrication of sub-10-nanometer metal-oxide devices with in situ control of electrical properties. *Science*, 270(5242):1639–1641, December 1995.

[23] R. Nemutudi, N. J. Curson, N. J. Appleyard, D. A. Ritchie, and G. A. C. Jones. Modification of a shallow 2DEG by AFM lithography. *Microelectronic Engineering*, 57–58:967–973, 2001.

[24] A. Boisen, K. Birkelund, O. Hansen, and F. Grey. Fabrication of submicron suspended structures by laser and atomic force microscopy lithography on aluminum combined with reactive ion etching. *Journal of Vacuum Science and Technology B*, 16(6):2977–2981, November/December 1998.

[25] G. Abadal, A. Boisen, Z. J. Davis, O. Hansen, and F. Grey. Combined laser and atomic force microscope lithography on aluminum: Mask fabrication for nanoelectromechanical systems. *Applied Physics Letters*, 74(21):3206–3208, 1999.

[26] N. J. Curson, R. Nemutudi, N. J. Appleyard, M. Pepper, D. A. Ritchie, and G. A. C. Jones. Ballistic transport in a GaAs/$Al_xGa_{1-x}As$ one-dimensional channel fabricated using an atomic force microscope. *Applied Physics Letters*, 78(22):3466–3468, May 2001.

[27] U. F. Keyser, H. W. Schumacher, U. Zeitler, R. J. Haug, and K. Eberl. Fabrication of quantum dots with scanning probe nanolithography. *physica status solidi (b)*, 224(3):681–684, 2001.

[28] S. Sasa, T. Ikeda, A. Kajiuchi, and M. Inoue. AFM fabrication and characterisation of InAs/AlGaSb nanostructures. *Solid-State Electronics*, 42(7–8):1069–1073, 1998.

[29] S. Sasa, A. Nakashima, S. Yodogawa, T. Kita, and M. Inoue. Magnetotransport properties of InAs nanostructure devices produced by AFM oxidation. *Physica B*, 314:95–98, 2002.

[30] S. Sasa, A. Nakashima, Y. Nakajima, and M. Inoue. Aharonov–Bohm oscillations observed in nanoscale open-dot structures fabricated in an InAs surface inversion layer. *Physica E*, 20:224–227, 2004.

[31] K. Möller, L. Töben, Z. Kollonitsch, C. Giesen, M. Heuken, F. Willig, and T. Hannappel. In-situ monitoring and analysis of GaSb(100) substrate deoxidation. *Applied Surface Science*, 242(392–398), 2005.

[32] Ioffe Physico-Technical Institute. New semiconductor materials. characteristics and properties. http://www.ioffe.rssi.ru/SVA/NSM/, November 2005.

[33] M. Winter. WebElements periodic table. http://www.webelements.com/, November 2005.

4 Sample Preparation and Measurement Strategies

> [. . .] une œuvre d'homme n'est rien d'autre que ce long cheminement pour retrouver par les détours de l'art les deux ou trois images simples et grandes sur lesquelles le cœur, une première fois, s'est ouvert.
>
> *(Albert Camus, L'envers et l'endroit)*

4.1 Introduction

Science lives by the reproducibility and falsifiability of its results. Even if such accounts rarely make for exciting reading, it is therefore of vital importance to provide as much data on the experimental procedure as is necessary to reproduce experiments independently, at least as far as this is possible within the context of publication.

The motivation for this chapter is twofold: First, to provide the background information that—it is to be hoped—puts the reader in the position to fully assess and reproduce the results presented here. To this end I shall report on the details of the experimental procedure where relevant and discuss several technologies with a focus on their use in the present work. Second, to collect all technical information in a convenient place to which one can refer back if necessary. In this way, I can omit these particulars from the discussion of physical phenomena, while still ensuring that they are readily available if needed.

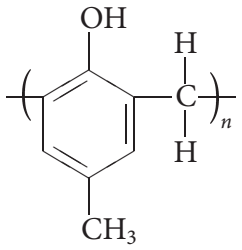
In Sec. 4.2 I shall describe the technologies used for producing samples from double heterostructures (DHETs) that are suitable for magnetotransport measurements and for creating modulations on them. In this context, Secs. 4.2.1, 4.2.2, 4.2.3, and 4.2.4 focus on lithographic

techniques, while 4.2.5 and 4.2.6 discuss methods for transferring patterns created in a surface layer into the semiconductor by etching. Sec. 4.3 explains the equipment and methodology employed in performing magnetotransport experiments on these samples.

4.2 Fabrication Techniques

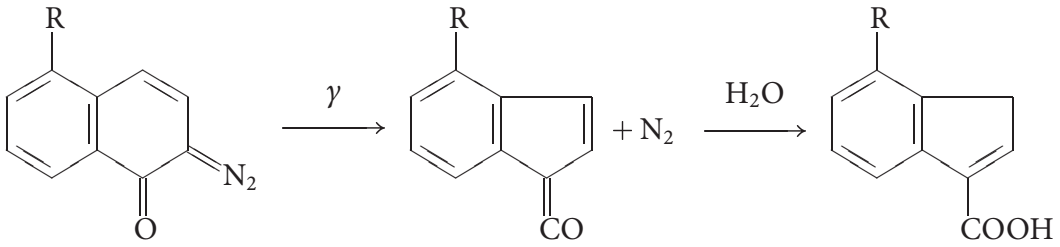
4.2.1 Optical Lithography

Optical lithography is a standard technique in semiconductor processing. One can distinguish positive processes, in which the exposed area is removed during development, and negative ones, in which exposed areas are retained while the rest of the photoactive film is removed [1]. Positive optical resist is generally based on the 'novolak' resin,



, (4.1)

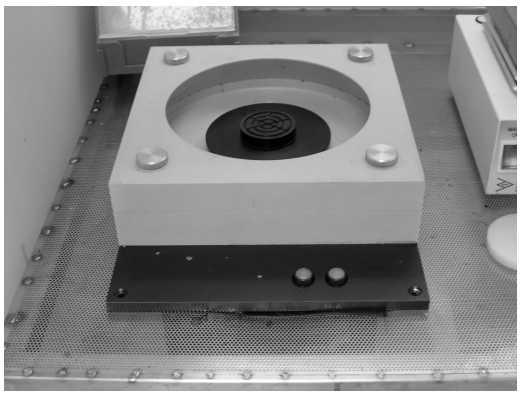
which is mixed with a strong dissolution inhibitor derived from diazonaphthoquinone (DNQ). Exposure to ultraviolet (UV) radiation changes DNQ into a ketene, which takes up water to form a carboxylic acid:



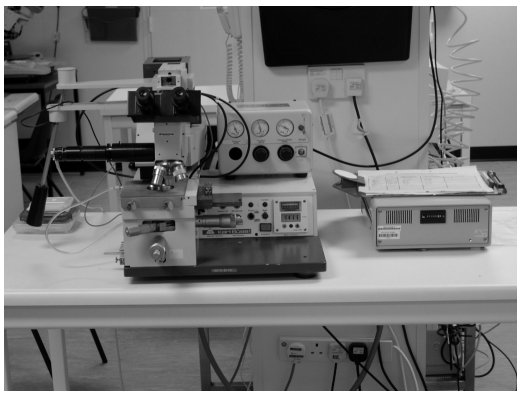
(4.2)

The carboxylic acid acts as a weak dissolution promoter [2].

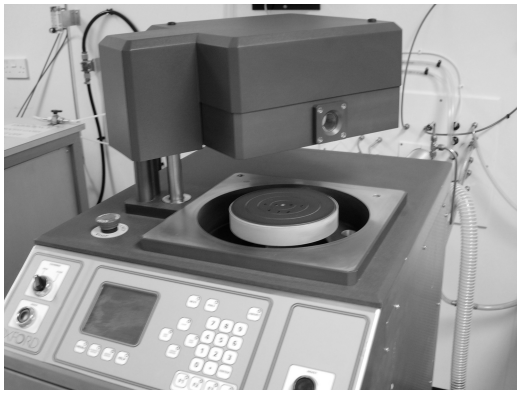
In this work, SHIPLEY MICROPOSIT S1813 positive resist was used together with SHIPLEY MF319 developer. The resist was spun on at 5,000 rpm for 50 s and then baked at 100 °C for



(a) Spinner



(b) Karl Süss MJB3 mask aligner

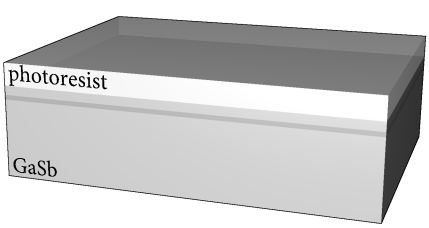


(c) Oxford Instruments RIE80+ reactive ion etcher

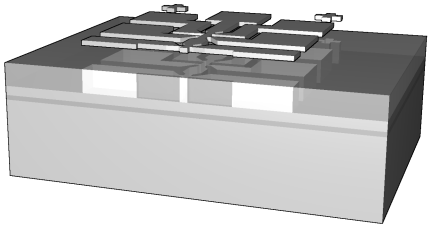
Figure 4.1: Clean room equipment

5 min by means of the conventional spinner seen in Fig. 4.1(a). Using a scanning electron microscope (SEM) to assess the transfer of strip patterns after development for 60 s, an optimal exposure time of 11 ± 1 s was determined for the Karl Süss MJB3 mask aligner of Fig. 4.1(b) with a new mercury vapour lamp.

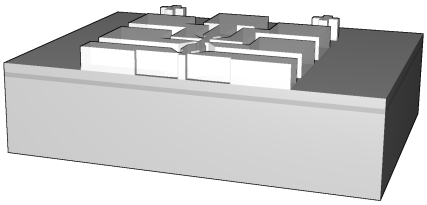
The complete optical lithography process is explained in Fig. 4.2 for the creation of a Hall bar-shaped mesa on a DHET. First, the photoresist was spun on the surface and baked to drive out the solvent and prevent sticking of the sample to the mask. The desired pattern was then exposed with a mask aligner and the exposed photoresist removed during development. In the next step, the surface was etched using one of the recipes described in Sec. 4.2.5 or 4.2.6. Usually a wet chemical etch was employed, as the inhomogeneous etch rate that can be



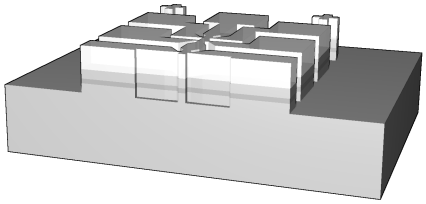
(a) Spinning on photoresist



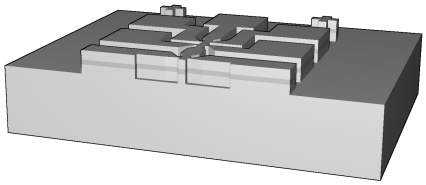
(b) Exposure of resist with contact mask



(c) Development of resist



(d) RIE or wet etch of semiconductor



(e) Removal of photoresist in acetone or stripper

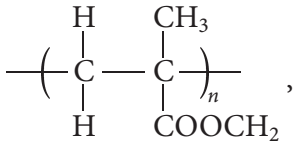
Figure 4.2: Optical lithography. In this example, a Hall bar pattern including contact pads is created using positive photoresist. The vertical scale is exaggerated.

achieved with a plasma etch was not vital for comparatively shallow etches at the resolution of the optical process. Finally, the remaining photoresist mask could be removed in acetone or resist stripper.

4.2.2 Electron Beam Lithography

Optical lithography is restricted by diffraction and has a finite depth of field. The practical resolution limit with deep UV light is approximately 0·15 µm, while 0·01 µm can be realized with hard X-rays [1]. Contact exposure with near UV light, which is useful in the laboratory for prototyping, achieves in practice a resolution of 0·8 to 1·0 µm. Electron beam lithography (EBL) is an alternative to optical lithography that uses electrons instead of photons and does not suffer from these effects. The wave length of the electrons is much smaller than the practical resolution limit given by the chemical properties of the sensitive layer and the quality of the beam focusing column, which is used instead of a mask to define the pattern.

As resist, EBL uses a polymer such as poly-(methyl-methacrylate) (PMMA),

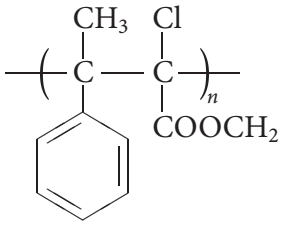


(4.3)

the chains of which can be broken by electron bombardment. The resulting oligomers become soluble and are removed during development, so that the polymer acts as a positive resist [2]. For very high doses, the opposite effect can be observed: exposed areas become insoluble even in solvents that readily attack the unmodified polymer. This effect is thought to be due to cross-linking and offers the possibility of creating negative patterns [3].

PMMA offers the highest available resolution—less than 10 nm is possible for isolated features [4]—but has a poor resistance to many etch plasmas, which erode the mask quickly by sputtering. This is particularly problematic for reactive ion etching (RIE) of III–V semiconductors, as the semiconductor etch rate is low. Several alternative chemistries have been

proposed to address this problem, one of the most successful ones being Nippon Zeon zep, which is based on poly-(methyl-chloroacrylate-co-α-methylstyrene):



(4.4)

Similar to pmma, exposure of zep causes chain scission and the chemical acts as a positive-tone resist.

ebl is affected by proximity effects caused by secondary electrons. Randomly scattered electrons induce a limited exposure of the resist film immediately around the nominally exposed areas. Although overexposure can be avoided by calculating the magnitude of the effect and correcting the dose accordingly, dense patterns suffer from a significant reduction in the contrast between exposed and unexposed areas and the resolution is worse than for isolated lines.

Most of the electron beam work was done in the Cavendish Laboratory in Cambridge with the help of Geb Jones, because the facilities available in Oxford would only allow for a limited resolution. We used both pmma and zep resist layers, each spun on the sample to give a film thickness of approximately 300 nm. pmma was baked at 180 °C for 1 h before exposure and developed for 10 s in a mixture of 583 parts methyl-ethyl-ketone, 1,617 parts ethanol, 30 parts 2-ethoxyethanol, 70 parts methanol, 1 part methyl-isobutyl-ketone (mibk), and 3 parts isopropanol (ipa). zep films were baked at 180 °C for 5 min. They were developed for 1 min in amyl-acetate followed by a 10 s rinse in a developer consisting of 9 parts mibk and 1 part ipa. Optimal exposure varied as a result of the proximity effect, with typical values around 90 µC/cm^2 for pmma and 300 µC/cm^2 for zep.

The use of zep to create more resistant etch masks for the rie of GaSb and InAs was only partly successful. Untreated zep films were etched with an etch rate of 20 to 30 nm/min in an argon-free hydrogen–methane plasma, comparable to etch rate of pmma. For some

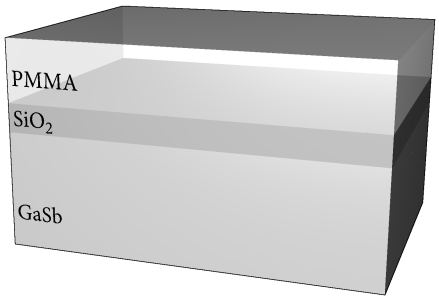
samples, this could be improved by thermal treatment at 145 °C for 2 h, yielding etch rates of 3 to 10 nm/min. The heat treatment appeared to be effective only if performed within a few days of development, and the results were inconsistent. The underlying chemical mechanism is not understood, and both the high etch rates observed initially and the thermal hardening may conceivably be connected to contaminants introduced during processing.

Another approach that can be employed to compensate for the poor etch resistance of PMMA is the use of an intermediate mask layer. SiO_2 has a negligible etch rate in hydrogen–methane plasmas, but can be etched readily using a freon-based recipe [5]. Fig. 4.3 illustrates the creation of antidot arrays by means of this technique: After a 40 nm thick silicon dioxide layer was evaporated on the sample and the resist was spun on, the dots were exposed in an electron beam writer. Development created a resist mask, which was used to transfer the pattern to the silicon dioxide layer using a freon-based etch plasma. The patterned oxide film then served as a mask for etching the InAs–GaSb DHET; during this process, the original PMMA mask was typically removed completely. Finally, the SiO_2 could be removed in hydrofluoric acid.

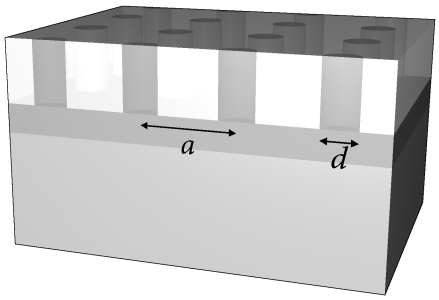
4.2.3 Direct Surface Modification by Local Anodic Oxidation

LAO can be used to oxidize a GaSb surface directly; the process has been described in detail in Chapter 3. The method makes it possible to fabricate oxide dot arrays covering a limited area such as the surface of a Hall bar. Oxide dots may then be dissolved in water or hydrochloric acid, leaving holes at most 20 nm deep. Fig. 4.4 illustrates the procedure.

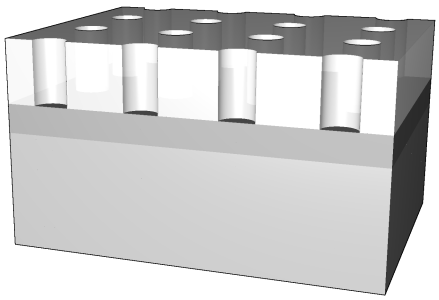
Since features created by LAO are generally not visible in an optical microscope [6], alignment marks or mesas must be defined on the surface before the atomic force microscope (AFM) patterning—alignment would be impossible the other way around. As a result of the standard lithographical process described in Sec. 4.2.1, organic residue will be present on the surface, which can stick to the AFM probe and interfere with anodization (see Fig. 4.5). Such residue can routinely be removed in a RIE setup with an oxygen plasma, a process commonly



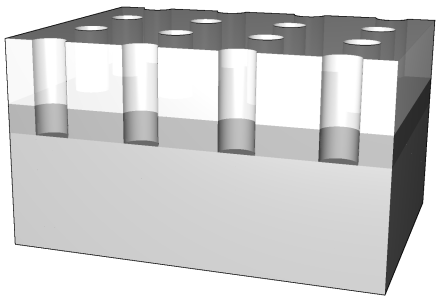
(a) Evaporation of silica and resist coating



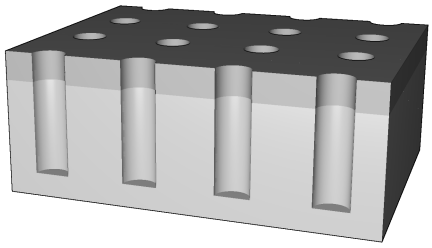
(b) Exposure of resist with electron beam



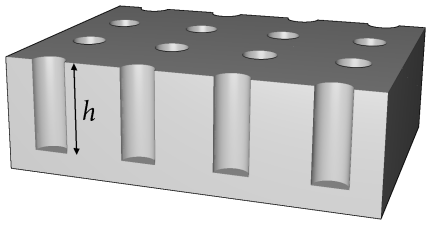
(c) Development of resist



(d) RIE of silica

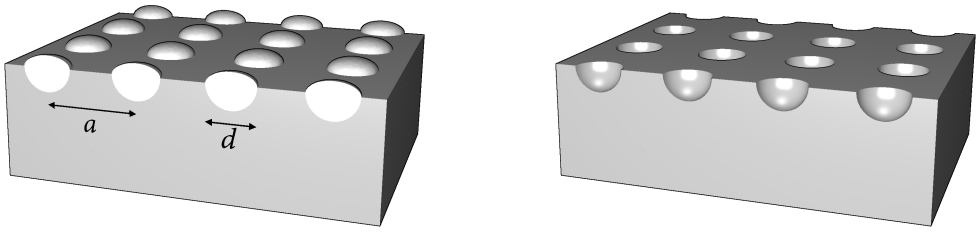


(e) RIE of GaSb



(f) Chemical removal of silica in HF

Figure 4.3: Electron beam lithography. In this example, holes with diameter d, lattice constant a, and depth h are created. An intermediate silica mask is used.



(a) Local anodization with the AFM

(b) Chemical removal of oxide dots

Figure 4.4: Direct surface modification with the AFM. In this example, holes with diameter d and lattice constant a are created by LAO.

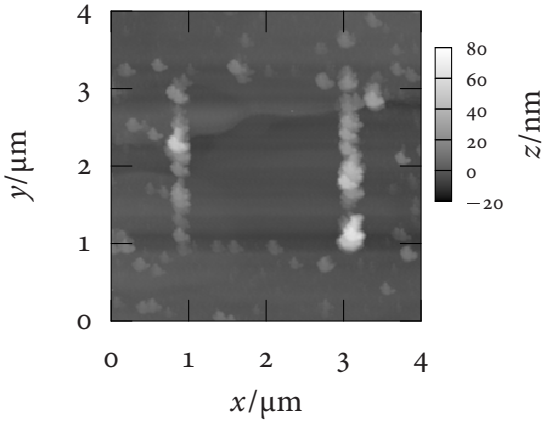


Figure 4.5: AFM micrograph showing organic residue after optical lithography. The central square has been swept clear by scanning in contact mode. Repeating features representative of tip geometry are clearly visible; they are believed to be caused by pick-up of contaminants.

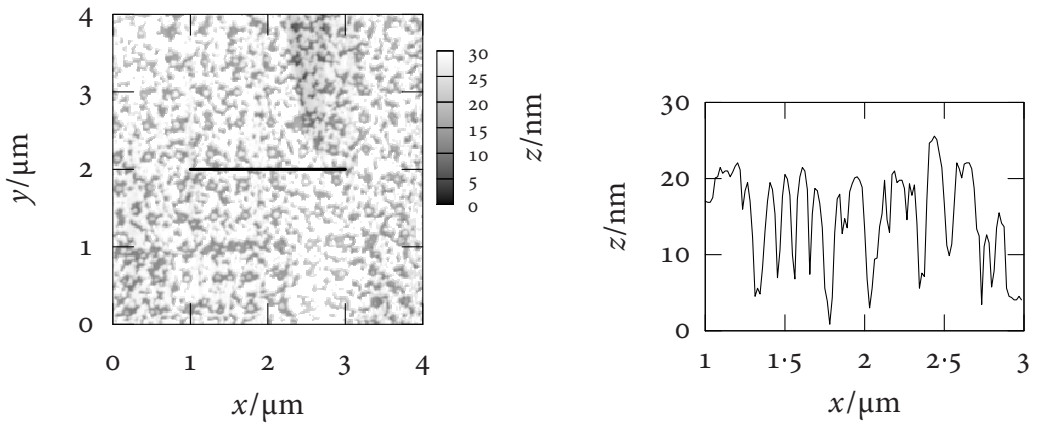
referred to as 'ashing' [7,8]. As an oxygen plasma was not available in the standard configuration of the RIE equipment available to me, an argon plasma was substituted in most cases, which is sufficient to remove thin organic layers by sputtering.

4.2.4 Creation of Aluminium Etch Masks by Local Anodic Oxidation

Direct modification of semiconductor surfaces by oxidation with an AFM is limited by the finite oxide thickness that can be achieved with this method. It is therefore desirable to have the option of transferring the lithographic patterns into the sample using an etch technique such as RIE that is capable of producing deeper features with a larger aspect ratio. This is possible if LAO can be used to create a suitable etch mask.

Silica withstands many useful RIE chemistries, including the hydrogen and methane based recipes usually employed to etch III–V semiconductors. The oxide formed by local anodization of silicon can therefore be used as an etch mask [9,10]. A similar approach using the oxidation of thin aluminium layers has been proposed by Anja Boisen *et al.* [11,12]. Both aluminium and aluminium oxide can be used as effective etch masks for RIE of both silicon and III–V semiconductors, and it is possible to selectively remove either the unoxidized aluminium or the oxide features in a wet etch solution without compromising the lithographic resolution. Such a metal layer may therefore be used as either a positive or negative resist. Al can be removed in 85 % phosphoric acid at room temperature, whereas selective removal of Al_2O_3 can be accomplished in a mixture of phosphoric acid, deionized water, and chromiumtrioxide at approximately 50 °C. If necessary, the aluminium or aluminium oxide may be removed in hydrofluoric acid.

The possibility of a positive process makes the method attractive for the creation of antidot patterns. Exposure of the entire area between the dots takes much longer than drawing individual dots. The risk of tip failure would be increased; moreover, protecting the contact leads would require an additional process step. I have therefore adapted the creation of aluminium masks to GaSb surfaces. Sec. 3.7 describes the LAO of thin aluminium layers and



(a) AFM micrograph of eroded Al layer

(b) Height profile at the line indicated in (a)

Figure 4.6: Erosion of aluminium layers on GaSb in oxide stripper. Remnants of an antidot pattern are discernible in (a).

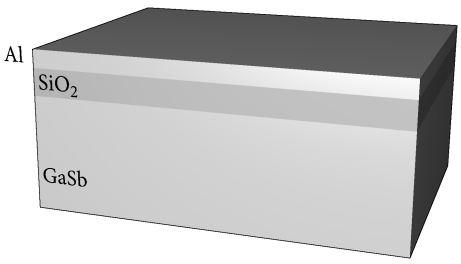
indicates that up to 20 nm thick films can be oxidized completely with a minimal feature size of approximately 130 nm. For mask applications, aluminium layers with a nominal thickness of 10 nm to 15 nm were chosen, and it was verified that such layers would withstand a typical RIE recipe for etching GaSb and InAs (recipe B from Sec. 4.2.6) for long enough to etch roughly 200 nm of GaSb.

For the selective removal of anodized aluminium, an etchant consisting of 3·5 ml H_3PO_4 (85 %), 100 ml DI-H_2O, and 2·0 g CrO_3 at 60 °C was used [13]. Initial attempts at selective removal of oxide dots in aluminium layers evaporated directly on GaSb epilayers were unsuccessful: the aluminium layer was quickly attacked in the oxide stripper before the oxide could be removed completely. Within 1 to 4 min, holes were opened in the aluminium layer down to the underlying GaSb, leading to a characteristic erosion pattern illustrated in Fig. 4.6. By comparing the behaviour of oxide layers on GaSb and Si processed under identical conditions, it was established that this effect is specific to aluminium films on GaSb and not caused by any other effect such as external contaminants.

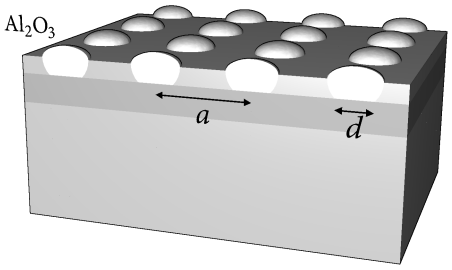
The chemical mechanism of this erosion caused by the presence of the GaSb surface is not fully understood. One possible explanation is the presence of Ga or Sb in the aluminium

layer because of interdiffusion. The problem could be avoided by introducing a high-quality SiO_2 spacer layer created by ion-beam assisted evaporation between the GaSb surface and the Al layer; a layer thickness of 40 nm was chosen. As explained in Sec. 4.2.6, silicon dioxide can be etched in a freon-based plasma and such an additional etch step can be added before the hydrogen–methane etch without having to remove the sample from the reaction chamber. As with direct manipulation of the surface, it is necessary to define either alignment marks or entire Hall bars by optical lithography prior to LAO, since it is not possible to align optical masks to features created by AFM patterning. This may be done either by defining and etching devices before the SiO_2 and Al layers are applied, or by patterning the Al layer and using it as a mask for the RIE of both the contact leads and the nanoscopic patterns. Both methods have drawbacks which caused the failure of a number of devices, leading to a reduced yield. On predefined mesas, the silica covering of sidewalls was imperfect, and the same erosion pattern as described above usually appeared in a strip of oxide several micron wide on both sides of the step. Since small Hall bars (typically 5 or 10 µm wide) were required to allow complete patterning in the AFM, such a defect could completely affect the active area of a device. If the Al layer was to be patterned by optical lithography, cracking of the underlying silica film sometimes occurred during development of the optical resist; such cracks often caused similar failures as edges and acted as starting points for peeling of the evaporated mask layers in the RIE.

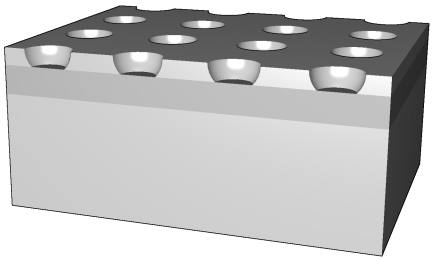
The full fabrication process is illustrated in Fig 4.7: First, SiO_2 and Al layers were created by evaporation, possibly patterning the Al layer using optical lithography. The Al film was then locally oxidized down to the underlying SiO_2 with an AFM and the oxide was removed in the oxide stripper, leaving holes in the aluminium layer. The aluminium layer served as a mask for the plasma etching of the silicon dioxide and the InAs–GaSb heterostructure. Finally, both the Al and the SiO_2 films were removed in hydrofluoric acid, exposing the patterned semiconductor. Fig. 4.8 shows a test pattern before and after the critical oxide removal step, corresponding to Fig. 4.7(b) and Fig. 4.7(c), respectively.



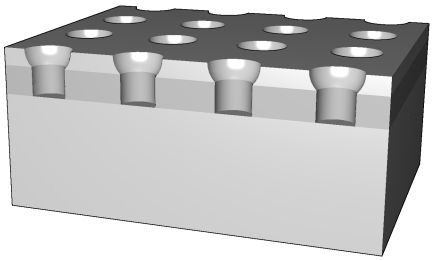
(a) Evaporation of silica and Al



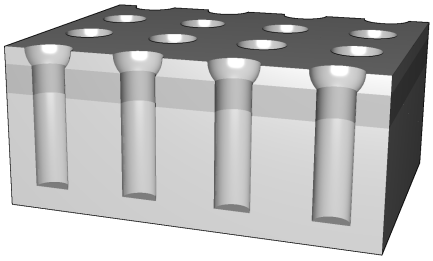
(b) Local anodization with the AFM



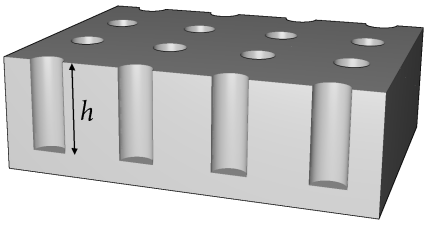
(c) Chemical removal of anodized Al



(d) RIE of silica

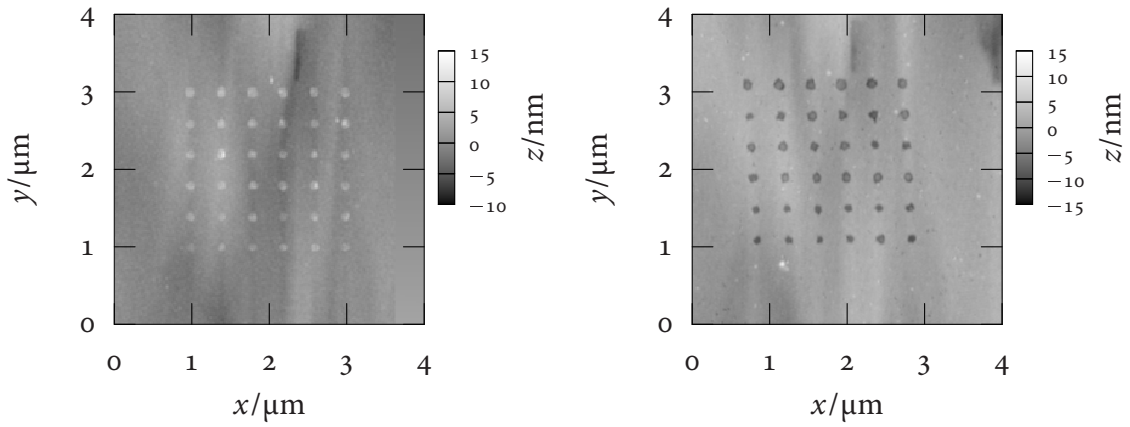


(e) RIE of GaSb



(f) Chemical removal of the mask in HF

Figure 4.7: Pattern transfer using an AFM-created RIE mask. In this example, holes with diameter d, lattice constant a, and depth h are created in GaSb, which is protected from the selective alumina etch by a silica layer.



(a) Before removal of oxide

(b) After removal of oxide

Figure 4.8: AFM micrographs of an aluminium etch mask created by LAO

4.2.5 Wet Chemical Etching

I used two different wet etch solutions for shaping InAs–GaSb DHETs; wet etching served predominantly to create mesa structures, such as Hall bars, that pattern the InAs quantum well and enable transport measurements in a well-defined geometry.

BUGLASS *et al.* [14] have proposed an etchant that is suitable for III–V semiconductors containing antimony, as it contains $(C_4H_4O_6)^{2-}$ tartrate ions that react with Sb to form a soluble complex, thereby preventing the formation of a protective layer of antimony oxide. The actual etchant is an aqueous solution containing 0·1 mol $(NH_4)_2$ $(C_4H_4O_6)$, 1·0 mol HCl, and 0·5 mol H_2O_2 per litre. During the etching of (100)-GaSb, gas bubbles develop and the surface is roughened noticeably. The etchant ages quickly and the etch rate varies accordingly. In the interest of reproducibility, the solution was allowed to stabilize for 20 min, whereafter an isotropic etch rate of approximately 100 nm/min was determined for GaSb by measuring etch steps using an AFM. This agrees with the values found by CHRISTOPHER BUMBY [15], who reports an initial rate of 200 nm/min, decreasing significantly within 30 min. The etch rate on InAs is much less, and a surface layer is eventually formed that prevents further erosion [15]; for the thin (30 nm) InAs quantum wells present in the DHETs of interest here, this effect did not pose a problem.

An alternative system, originally proposed by Adachi *et al.* [16] in a slightly different form, uses the dichromate ion, $(Cr_2O_7)^{2-}$, as an oxidizing agent. It etches most III–V semiconductors isotropically and with little roughening. The etchant was prepared as an aqueous solution containing 0·0367 mol $K_2Cr_2O_7$ (potassium dichromate), 3·97 mol HBr (hydrobromic acid), and 5·83 mol CH_3COOH (acetic acid) per litre; the comparatively low dichromate concentration originated from the use of a commercially available prepared solution. With this recipe, etch rates of approximately 2,000 nm/min for (100)-GaSb and 2,800 nm/min for (111)-InSb were observed for shallow ($\gtrsim 1\,\mu m$) etches. This does not contradict the the lower values around 200 nm/min reported by Bumby [15] for bulk GaSb, as he has investigated deep (several micron) etching and notes that agitation is required to maintain fast etch rates: the observation indicates that the concentration of the active species drops quickly close to the semiconductor surface, leading to lower rates. On semiconductors containing antimony, a brown discolouration of the surface was frequently seen after etching; the film could be removed by dipping in concentrated HCl. Overall, the etch gave higher etch rates, better repeatability, and smoother surfaces than the tartrate etch and was therefore preferred for most mesa structures.

4.2.6 Reactive Ion Etching

Reactive ion etching (RIE), also referred to as dry etching, allows highly anisotropic etch rates and is therefore the preferred method for creating features with high aspect ratios, especially at the submicron scale. In a low-pressure, continuous-flow reaction chamber such as the one shown in Fig. 4.1(c), the process gases are exposed to a radio-frequency (RF) oscillating electric field above a planar electrode. The RF excitation ionizes the gases and forms a plasma containing radicals that can react with the sample placed on the electrode [1]; the chemical process itself may be isotropic or exhibit a degree of crystal-plane dependence. Anisotropy is a consequence of the large (typically several hundred volts) DC voltage that builds up because the DC-isolated electrode is charged by absorption of electrons from the plasma: transport of

Recipe	Freon 14 (sccm)	Argon (sccm)	Hydrogen (sccm)	Methane (sccm)	Power (W/cm^2)	Pressure (mTorr)
A			20·0	4·0	1·25	40·0
B		14·0	40·0	10·0	0·83	40·0
C			40·0	10·0	0·83	40·0
D	50·0				0·42	50·0

Table 4.1: RIE etch recipes

Recipe	GaSb (nm/min)	SiO_2 (nm/min)	Al (nm/min)	PMMA (nm/min)	ZEP (nm/min)	baked ZEP (nm/min)
A	$1·6 \pm 0·2$	$< 0·05$	$< 0·05$	> 30	> 30	
B	$1·7 \pm 0·2$	$< 0·05$	$< 0·05$	> 30		
C	$1·0 \pm 0·2$	$< 0·05$	$< 0·05$	20 to 30	20 to 30	4 to 10
D		16 ± 1		> 30	> 30	

Table 4.2: RIE etch rates

reactive ions depends on the electric field[1] corresponding to the potential difference and is highly directional. Dry etching of III–V semiconductors typically exhibits slow etch rates in comparison with silicon. While other chemistries have been explored, a hydrogen–methane plasma is often used as it gives good anisotropy and reproducibility for all III–V materials [5, 7,17]. SiO_2, which is used as an intermediate dielectric mask, can be etched readily in a freon-based plasma.

For these experiments, I used the OXFORD INSTRUMENTS RIE80+ plasma etcher depicted in Fig. 4.1(c). Table 4.1 lists the etch recipes employed in the present work. Recipe A followed the work of WERKING *et al.* [17], while recipe B had previously been used in Oxford by REHMAN [18]; recipe C was based on recipe B, but omitted the argon to achieve lower sputter rates on organic masks. Recipe D was a standard prescription for dry etching of silica.

Etch rates for the individual recipes are listed in Table 4.2. The etch rates for GaSb were quite low compared with published values [17]. In particular the rate for recipe B was approximately a third of that reported by REHMAN [18] on the same equipment. The reason

[1] Of the order 1 kV/m in conventional reaction chambers.

for this discrepancy is not fully understood. The etch rates listed here were determined by measuring step sizes after different process times with the help of a calibrated AFM and were corroborated for recipe C by the exposure of the InAs layer, which can be identified by its colour, after a suitable etch time. RIE in a continuous-flow reactor is quite sensitive to the process conditions. However, no significant increase in etch rate was obtained by varying power, pressure, and methane flow. The etch rates for organic resists could only be determined approximately, as their behaviour is non-linear: with the plasma formulae discussed here, initial etch rates were very high, but the resist was hardened with time. Presumably this was caused by a mechanism similar to the crosslinking that happens as a result of electron beam overexposure—exposed EBL resist that is not removed during development also was more resistant to the etch plasma. A further complication was the poor reproducibility of the thermal hardening of ZEP described in Sec. 4.2.2. Aluminium and silica did not etch appreciably in the hydrogen–methane plasmas used for structuring GaSb and InAs. However, peeling of the mask layers off the semiconductor surface was sometimes observed during long plasma exposures. Such peeling could be reduced, but not avoided completely, by careful cleaning of the surfaces before evaporation of the mask layer.

4.3 Transport Measurements

4.3.1 Magnets and Cryogenic Setup

Two superconducting magnets were used for most[2] magnetotransport measurements: an Oxford Instruments magnet with a maximal field of 18 T, which is shown in Fig. 4.9(a), and a state-of-the-art magnet, also manufactured by Oxford Instruments, with a maximal field of 21 T and a novel cryostat—see Fig. 4.9(b)—designed to reduce the helium evaporation rate. In both cases, the quoted field could only be obtained when the magnet was cooled to a temperature of 2·19 K. This temperature corresponds to the λ-point of ^{4}He and could be

[2] Some work was done on two other superconducting magnet systems very similar to the 18 T-magnet described here.

(a) Oxford Instruments 18 T superconducting magnet

(b) Oxford Instruments 21 T superconducting magnet

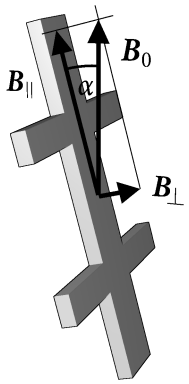
Figure 4.9: Magnets and cryogenic equipment

achieved in the volume below the lambda-plate without cooling down the entire ^{4}He bath. If used at the boiling point of ^{4}He, 4·23 K, which was often convenient for simplicity, the permissible fields were reduced to approximately 16 T and 19 T, respectively.

The sample under investigation could be immersed in the main bath either directly or within an insert filled with low pressure helium gas to facilitate heat exchange. For temperature dependent measurements, a variable temperature insert (VTI) was fitted to the 21 T-magnet, allowing the temperature to be controlled between 1·5 K and approximately room temperature; for the 18 T magnet, a ^{3}He-insert was available, allowing temperatures of 450 mK to be reached.

For both the VTI and the ^{3}He-refrigerator, sample holders including a rotation mechanism were available so that the sample could be turned about an axis perpendicular to the applied field from outside of the cryostat. The insert used with the VTI is shown in Fig. 4.10(a). Apart from making it possible to switch between perpendicular and parallel field configurations without having to take the sample out of the cryostat, such an arrangement facilitated studying the effect of an in-plane field on the magnetotransport. As illustrated in Fig. 4.10(b), in an external field $\mathbf{B}_0$ a sample rotated by an angle α from the parallel field configuration

(a) Rotating sample holder



(b) Rotated Hall bar

Figure 4.10: Sample rotation

experiences a field $\mathbf{B}_\perp$ with magnitude $B_\perp = B_0 \sin\alpha$ in the normal direction and a field $\mathbf{B}_\parallel$ with magnitude $B_\parallel = B_0 \cos\alpha$ parallel to the surface. As long as α is small, the perpendicular field in the presence of a strong parallel field can be varied straightforwardly by rotating the sample, as $B_\perp \approx \alpha B_0$ and $B_\parallel \approx B_0$. For larger rotation angles these approximations do not hold: although rotation then does not provide a simple plot versus the smaller field component, the full parameter space in $B_\perp$ and $B_\parallel$ can still be mapped. An alternative approach, which was not attempted, would be to adjust B_0 to keep one of the components constant as the sample is rotated, *e.g.*, $B_0(\alpha) = B_\parallel / \cos\alpha$.

4.3.2 Electronic Measurements

The full two-dimensional conductivity tensor $\boldsymbol{\sigma}$ of a quasi-two-dimensional carrier system can be determined most straightforwardly in the Hall bar geometry shown in Fig. 4.11 [19]. Owing to the use of four-contact measurements, there is no contribution due to contact and lead resistances. The longitudinal resistivity is given by

$$\rho_{xx} = \frac{w}{\ell}\frac{V_{xx}}{I}, \tag{4.5}$$

while the Hall resistivity is calculated as

$$\rho_{xy} = \frac{V_{xy}}{I}. \tag{4.6}$$

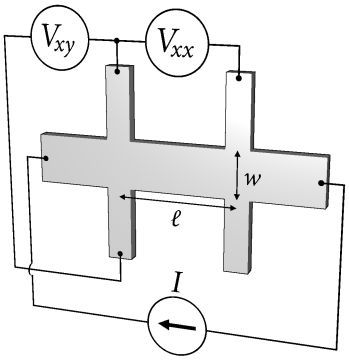


Figure 4.11: Measuring resistance in the Hall bar geometry

Assuming that $\rho_{yy} = \rho_{xx}$ and $\sigma_{yy} = \sigma_{xx}$,[3] the components of the conductivity tensor $\boldsymbol{\sigma} = \boldsymbol{\rho}^{-1}$ can be recovered from the well-known matrix relations

$$\sigma_{xx} = \frac{\rho_{xx}}{\rho_{xx}^2 + \rho_{xy}^2}; \qquad \sigma_{xy} = \frac{-\rho_{xy}}{\rho_{xx}^2 + \rho_{xy}^2}. \tag{4.7}$$

In practice, V_{xx} and V_{xy} were measured with a standard AC lock-in technique. The oscillator output of the lock-in amplifier was connected in series with a large adjustable resistor $R_{\text{series}} \gg R_{\text{sample}}$, where R_{sample} is the typical resistance between two contacts of the sample under investigation. If connected in an appropriate manner (*cf.* Fig. 4.11), the oscillator could thus act as a current source providing an excitation current $I \approx V_{\text{osc}}/R_{\text{series}}$. Because significant capacitive coupling between leads cannot be avoided in magnetotransport experiments, a low excitation frequency was desirable. Frequencies in the range 10 to 30 Hz were used, and care was taken to detune the reference frequency with respect to external sources of noise and circuit resonances. If a total of four lock-in amplifiers were used, V_{xx} and V_{xy} could be measured for two samples under identical experimental conditions as shown in Fig. 4.12. Alternatively, the voltages accross both pairs of contacts that a Hall bar provides for measuring each of ρ_{xx} and ρ_{xy} could be monitored simultaneously and compared for consistency. Measurements of this kind were routinely done over a limited field range after cooling down

[3] This assumption is reasonable for all samples discussed here, but will not hold true for anisotropic modulations.

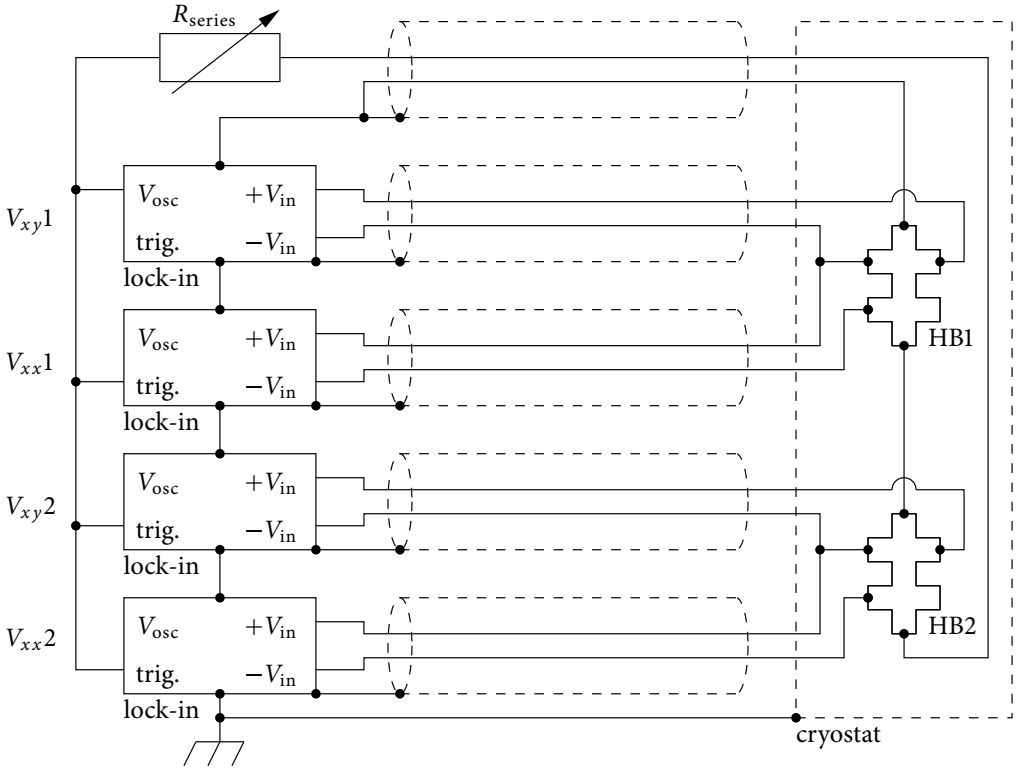
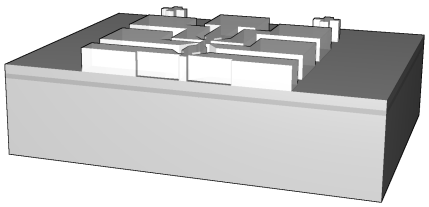


Figure 4.12: Set-up for measuring two samples with lock-in amplifiers

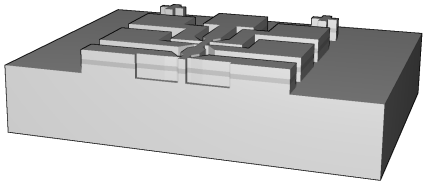
a sample and to investigate issues with individual samples (*cf.* Sec. 7.4.1), but most data were taken with a single pair of contacts for each Hall bar and independent resistivity component.

4.3.3 Fabrication of Hall bars

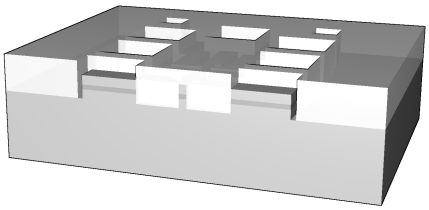
Hall bars for such measurements were produced by defining a suitably shaped mask on top of a DHET by optical lithography. The process is illustrated in Fig. 4.13: After lithography, wet chemical etching was used to create a Hall-bar-shaped mesa between 0·5 and 1·0 µm high. As the InAs layers lay between 50 and 150 nm below the surface, the quasi-two-dimensional electron and hole gases were confined to the shape of the mesa, which contained areas forming broad contact leads and contact pads as shown in Fig. 4.14. A second photoresist mask was then created on top of the mesa, which left holes at the contact pads. In an evaporator the sample was covered with a metallization consisting of 8 nm nickel, 140 nm eutectic



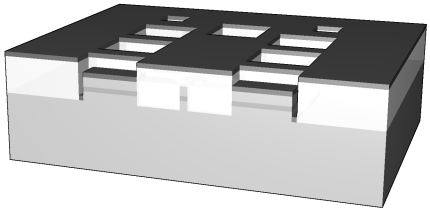
(a) Definition of Hall bar mask



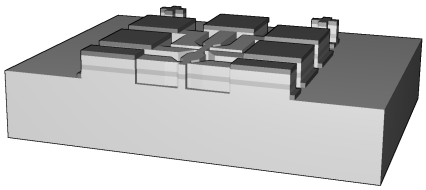
(b) Pattern transfer



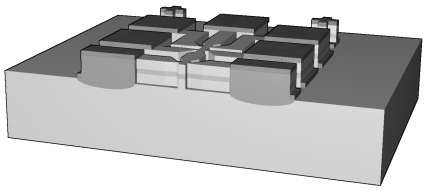
(c) Definition of contact mask



(d) Evaporation of contact metallization



(e) Lift-off



(f) Annealing

Figure 4.13: Fabrication of Hall bars for transport measurements. The vertical scale is exaggerated.

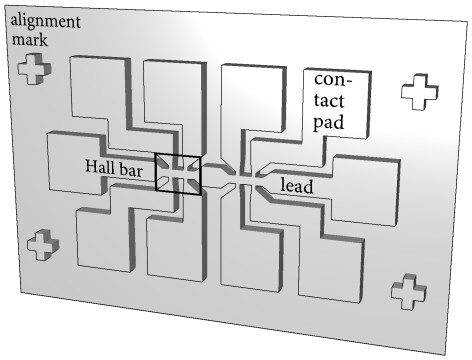


Figure 4.14: Hall bar mesa

gold–germanium alloy, a further 60 nm nickel, and finally 150 to 300 nm of gold. The nickel layers served to improve adhesion. In the next step, the metallization was lifted off the mesa except for the contact pads by dissolving the photoresist in acetone or resist stripper. Thin gold layers were preferred as they made lift-off easier. Finally, the contact metallization was annealed for 30 s at 390 °C. During this process, germanium diffused into the GaSb cap layer, forming an ohmic contact between the InAs quantum well and the remaining contact metal. The sample could then be attached to a suitable package, and the contact pads could be connected with a wire bonder. I used a Kulicke & Soffa 4124 manual gold ball bonder for this purpose.

4.4 Summary

As will be explained in Chapter 5, quasi-two-dimensional carrier systems appear at InAs–GaSb heterointerfaces. Magnetotransport measurements of a small portion of the carrier sheets that form in thin epitaxial heterostructures can be performed by patterning the surface layer of a suitable sample in the form of a conventional Hall bar. This is readily done using a combination of standard optical lithography and wet chemical etching techniques; details of their implementation in relation to the present work have been given. The resolution of these methods is sufficient to create Hall bars with sizes down to 5 µm × 15 µm. I have outlined how

such devices can be used to measure the resistivity (and hence conductivity) of the carriers at low temperatures as a function of external magnetic fields parallel or perpendicular to the plane of the surface.

Imposing a lateral modulation on the Hall bars that has a period much smaller than the mean free path in the carrier systems requires technologies capable of much higher resolution. One such technology is standard electron beam lithography (EBL). Access to expensive state-of-the-art equipment is, however, required to achieve sufficient resolution; this is particularly true as dense antidot patterns are especially susceptible to proximity effects caused by secondary electrons. Another problem with EBL is the incompatibility of the commonly used resist materials with deep reactive ion etching (RIE) of III–V semiconductors. Yet RIE, as a highly anisotropic etch technique, is required for the creation of pits with high aspect ratios. The difficulty can be circumvented by introducing an intermediate dielectric mask that stands up well to the etch plasma.

I have developed alternative lithographic techniques based on atomic force microscopy (AFM). Local anodic oxidation (LAO), described in Chapter 3, can be used to remove designated areas of the semiconductor surface with high resolution. However, the etch depth attainable with this technique is fundamentally limited and has ultimately proven insufficient for imposing a lateral modulation on the structures investigated in the present work. Special consideration has therefore been given to the possibility of using LAO to fabricate RIE etch masks. The chemical incompatibilities involved in the creation of a thin aluminium mask on GaSb can be avoided by introducing a silica spacer layer to isolate the mask from the antimonide.

Bibliography

[1] M. J. Kelly. *Low-Dimensional Semiconductors*, chapter 3. Series on semiconductor science and technology. Oxford University Press, Oxford, England, 1995.

[2] D. R. Medeiros, A. Aviram, C. R. Guarnieri, W.-S. Huang, R. Kwong, C. K. Magg, A. P. Mahorowala, W. M. Moreau, K. E. Petrillo, and M. Angelopoulos. Recent progress in

electron-beam resists for advanced mask-making. *IBM Journal of Research and Development*, 45(5):639–650, 2001.

[3] A. C. F. Hoole, M. E. Welland, and A. N. Broers. Negative PMMA as a high-resolution resist—the limits and possibilities. *Semiconductor Science and Technology*, 12:1166–1170, 1997.

[4] W. Chen and H. Ahmed. Fabrication of sub-10 nm structures by lift-off and by etching after electron beam exposure of poly(methylmethacrylate) resist on solid substrates. *Journal of Vacuum Science and Technology B*, 11(6):2519–2523, November/December 1993.

[5] J. R. Lothian, F. Ren, and S. J. Pearton. Mask erosion during dry etching of deep features in III–V semiconductor structures. *Semiconductor Science and Technology*, 7(9):1199–1209, 1992.

[6] R. M. Nyffenegger and R. M. Penner. Nanometer-scale surface modification using the scanning probe microscope: Progress since 1991. *Chemical Reviews*, 97:1195–1230, 1997.

[7] B. Schmidt, S. Illek, B. Borchert, J. Rieger, and M.-C. Amann. Tunable twin guide laser diodes with high output efficiency fabricated by an improved reactive ion etching technique. *Semiconductor Science and Technology*, 13:821–825, 1998.

[8] K. R. Williams, K. Gupta, and M. Wasilik. Etch rates for micromachining processing—part II. *Journal of Microeletromechanical Systems*, 12(6):761–778, December 2003.

[9] P. M. Campbell, E. S. Snow, and P. J. McMarr. AFM-based fabrication of Si nanostructures. *Physica B*, 227:315–317, 1996.

[10] P. A. Fontaine, E. Dubois, and D. Stiévenard. Characterization of scanning tunneling microscopy and atomic force microscopy-based techniques for nanolithography on hydrogen-passivated silicon. *Journal of Applied Physics*, 84(4):1776–1781, August 1998.

[11] A. Boisen, K. Birkelund, O. Hansen, and F. Grey. Fabrication of submicron suspended structures by laser and atomic force microscopy lithography on aluminum combined with reactive ion etching. *Journal of Vacuum Science and Technology B*, 16(6):2977–2981, November/December 1998.

[12] G. Abadal, A. Boisen, Z. J. Davis, O. Hansen, and F. Grey. Combined laser and atomic force microscope lithography on aluminum: Mask fabrication for nanoelectromechanical systems. *Applied Physics Letters*, 74(21):3206–3208, 1999.

[13] A. Boisen. Private communication, 2004.

[14] J. G. Buglass, T. D. McLean, and D. G. Parker. A controllable etchant for fabrication of GaSb devices. *Journal of The Electrochemical Society*, 133(12):2565–2567, December 1986.

[15] C. W. Bumby. *Thermophotovoltaic Devices based upon Heteroepitaxial GaSb grown by MOVPE*. DPhil thesis, Hertford College, University of Oxford, 2005.

[16] S. Adachi, H. Kawaguchi, and G. Iwane. A new etchant system, $K_2Cr_2O_7$–H_2SO_4–HCl, for GaAs and InP. *Journal of Materials Science*, 16(9):2449–2456, 1981.

[17] J. Werking, J. Schramm, C. Nguyen, E. Hu, and H. Kroemer. Methane/hydrogen-based reactive ion etching of InAs, InP, GaAs, and GaSb. *Applied Physics Letters*, 58(18):2003–2005, May 1991.

[18] Y. J. Rehman. *Processing and Magneto-transport Studies of InAs/GaSb Low Dimensional Structures*. DPhil thesis, The Queen's College, University of Oxford, 1999.

[19] C. Weisbuch and B. Vintner. *Quantum Semiconductor Structures*, chapter IV. Academic Press, San Diego, California, 1991.

5 GaSb–InAs Heterostructures

> When I hear of Schrödinger's cat, I reach for my gun.
>
> *(Steven William Hawking)*

5.1 Introduction

When in contact, GaSb and InAs have non-overlapping band gaps, which leads to a number of interesting phenomena in heterostructures created from these materials. Our group in Oxford has conducted extensive research on structures formed by alternating epitaxial GaSb and InAs layers grown by metal-organic vapour phase epitaxy (MOVPE). The purpose of Chapter 5 is to give an overview of these specimens both from a theoretical and a practical point of view.

These epitaxial strata are two-dimensional in the sense that they exhibit translational symmetry in the plane of the surface and the structure only changes along the perpendicular growth axis. In this thesis, the coordinate system is chosen such that the x and y axes lie parallel to the surface and the z axis points in the perpendicular direction. 'Parallel' refers to a direction in the (x, y)-plane, whereas 'perpendicular' or 'normal' refer to the z direction, as this direction is normal to the two-dimensional carrier systems which appear at the heterostructures and are the main focus of the present work.

The chapter starts out in Sec. 5.2 with a short reminder of subband energies in heterostructure quantum wells. In the following Sec. 5.3.1, a review of the 6·1 Å material system consisting of the nearly lattice-matched semiconductors InAs, GaSb, and AlSb and their het-

erostructures sets the stage for the discussion of the GaSb–InAs–GaSb double heterostructure (DHETS) studied in this thesis. After briefly touching the subject of InAs–AlSb quantum wells in Sec. 5.4, I shall concentrate on the peculiar broken-gap band line-up of GaSb–InAs heterostructures in Sec. 5.5 and explain how it leads to charge transfer, and—via quantization in the resulting electrostatic potential—to the formation of quasi-two-dimensional carrier gases. Electron–hole hybridization and the formation of the minigap are likewise discussed, hopefully clarifying the ramifications of this effect on observable properties, which will play a role in the remaining chapters.

Some remarks on Landau quantization and the in-plane charge transport in the presence of a perpendicular field follow in Sec. 5.6; while the focus is on sample characterization, they provide a useful reminder of some basic physical phenomena. Sec. 5.7 is dedicated to the growth of DHETS by MOVPE. A short introduction to the technology, which has not yet attained a high level of reproducibility, will allow the reader to better appreciate the limitations and uncertainties inherent in the creation of the samples discussed here. Finally, Sec. 5.8 will give an overview of the individual MOVPE-grown epitaxial structures discussed in this thesis for the purpose of reference. It will be pointed out how the observed properties of these samples can be understood in terms of the theoretical framework laid out earlier.

5.2 Subband Energies in a Heterostructure Quantum Well

The electrons and holes in a heterostructure quantum well experience a potential $V(z)$, where z is the position along the growth direction, but are free to move in the plane of the sample. Ignoring the influence of the crystal lattice, their wave functions are consequently of the form

$$\varphi(\mathbf{r}) = e^{i(k_x x + k_y y)} \chi(z), \tag{5.1}$$

where the envelope function $\chi(z)$ depends on z only. A form similar to Eq. (5.1) can still be found if the crystal lattice is taken into account, provided the interface potentials can be

neglected. This is know as the envelope function approximation [1]:

$$\varphi(\mathbf{r}) \approx \sum_{\alpha \in \{\mathrm{GaSb,InAs}\}} e^{i(k_x x + k_y y)} u_{\alpha \mathbf{k}}(\mathbf{r}) \chi(z), \tag{5.2}$$

where $u_{\alpha \mathbf{k}}$ is the Bloch function in material α.

As long as the effective masses in the materials forming the heterostructure are sufficiently similar, the Hamiltonian can be taken as

$$\hat{H} = \frac{\hat{\mathbf{p}}^2}{2\overline{m}^*} + V(z), \tag{5.3}$$

where $\hat{\mathbf{p}} = -i\hbar \nabla_{\mathbf{r}}$ is the momentum operator and $\overline{m}^*$ is the average effective mass. The resulting time-independent Schrödinger equation,

$$\hat{H}\varphi(\mathbf{r}) = E_{\mathrm{tot}} \varphi(\mathbf{r}), \tag{5.4}$$

is separable if $\varphi(\mathbf{r})$ is given by Eq. (5.1) or (5.2), leading to a one-dimensional Schrödinger-like equation in $\chi(z)$:

$$\hat{H}_0 \chi(z) = E_{\mathrm{tot}} \chi(z), \tag{5.5}$$

where the one-dimensional Hamiltonian reads

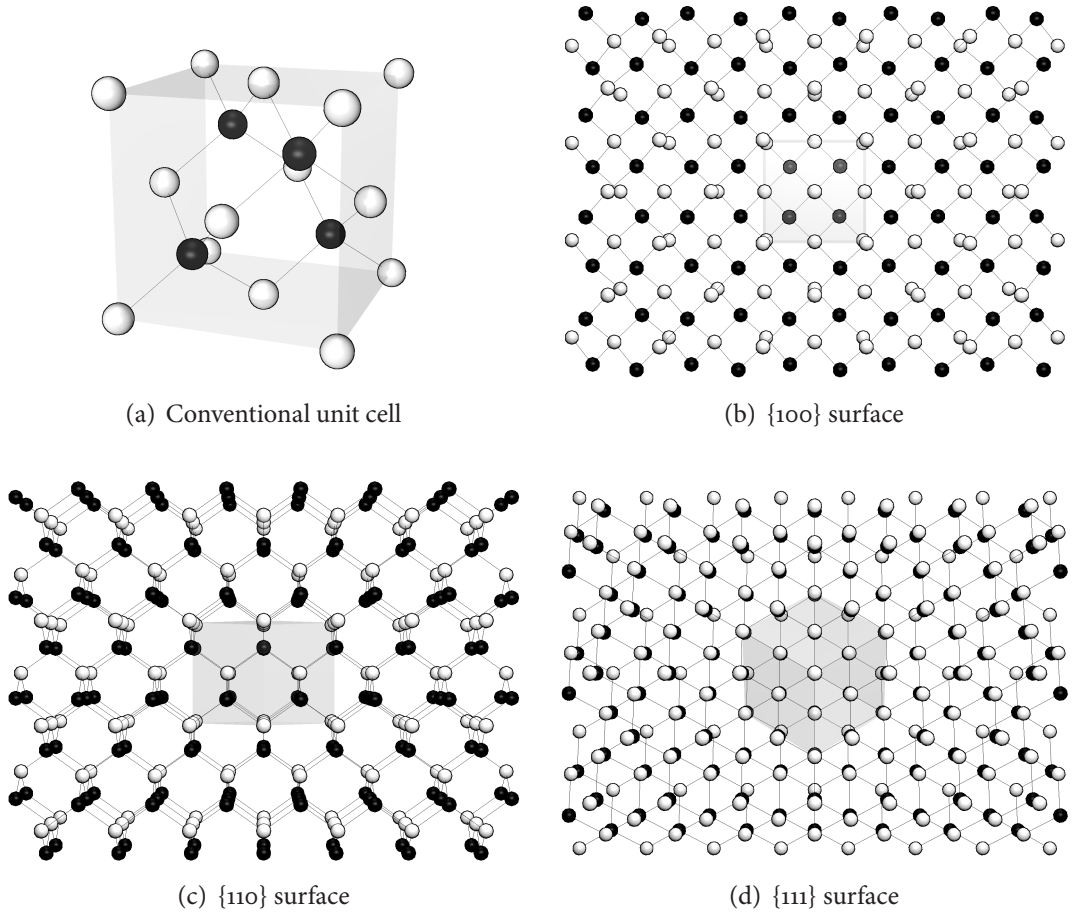
$$\hat{H}_0 = -\frac{\hbar}{2\overline{m}^*} \frac{\partial^2}{\partial z^2} + V(z) + E_{xy} \tag{5.6}$$

and E_{xy} is the in-plane kinetic energy resulting from the corresponding equation in x and y—in the free electron case, this would be $\hbar(k_x^2 + k_y^2)/2\overline{m}^*$. Eq. (5.5) leads to localized envelope functions $\chi_n(z)$ with corresponding subband energies $E_n = E_{\mathrm{tot}} - E_{xy}$, depending on the shape of the potential $V(z)$.

5.3 The 6·1 Å Family of III–V Semiconductors: InAs, GaSb, AlSb

5.3.1 Bulk Material Properties

InAs, GaSb, and AlSb crystallize in the familiar zincblende structure shown in Fig. 5.1, which consists of two interleaved face-centred cubic sublattices [2]. The lattice constant for all three



(a) Conventional unit cell

(b) {100} surface

(c) {110} surface

(d) {111} surface

Figure 5.1: The zincblende structure. Black and white spheres represent atoms of different species. Both sublattices are equivalent. Bonds are shown as thin black lines.

materials is around 6·1 Å, so that lattice-matched heterostructures can be grown [3]; the exact lattice spacings are 6·0584 Å for InAs, 6·0959 Å for GaSb, and 6·1355 Å for AlSb at 300 K [4]. Like GaAs, the materials preferentially cleave along the {110} family of planes.

The reciprocal lattice of the face-centred cubic lattice, and hence the diamond and zincblende lattices, is a body-centred cubic lattice. The Wigner-Seitz unit cell of the reciprocal lattice, which corresponds to the first Brillouin zone, forms a truncated octahedron and is shown in Fig. 5.2(a) with its conventionally labelled symmetry points [2, 4, 5].

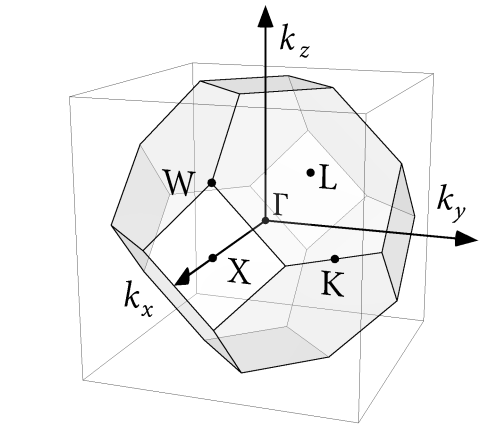
There are eight outer electrons per primitive unit cell—3 from the group III element (In, Ga, or Al) and 5 from the group V element (As or Sb)—which occupy four bonding sp^3

hybrid orbitals. The main features of the full band structure are shown in Fig. 5.2. Bulk InAs and GaSb are direct-gap semiconductors, since both the maximum of the valence band and the minimum of the conduction band lie at the Γ-point. The size of the energy gap is $E_g^{(\mathrm{InAs})} \approx 415\,\mathrm{m}e\mathrm{V} - 0{\cdot}276\,\mathrm{m}e\mathrm{V\,K^{-1}} \times T^2/(T - 83\,\mathrm{K})$ for InAs [9] and $E_g^{(\mathrm{GaSb})} \approx 813\,\mathrm{m}e\mathrm{V} - 0{\cdot}108\,\mathrm{m}e\mathrm{V\,K^{-1}} \times T^2/(T - 10{\cdot}3\,\mathrm{K})$ for GaSb [10]. Bulk AlSb is an indirect-gap semiconductor with the bottom of the conduction band at the X-point and $E_g^{(\mathrm{AlSb})} \approx 1696\,\mathrm{m}e\mathrm{V} - 0{\cdot}390\,\mathrm{m}e\mathrm{V\,K^{-1}} \times T^2/(T + 140\,\mathrm{K})$ [11]. At 300 K, the energy gaps are 350 m*e*V for InAs, 727 m*e*V for GaSb, and 1616 m*e*V for AlSb, while at 4·2 K the values are close to their zero-temperature limits of 813 m*e*V, 415 m*e*V, and 1696 m*e*V, respectively.

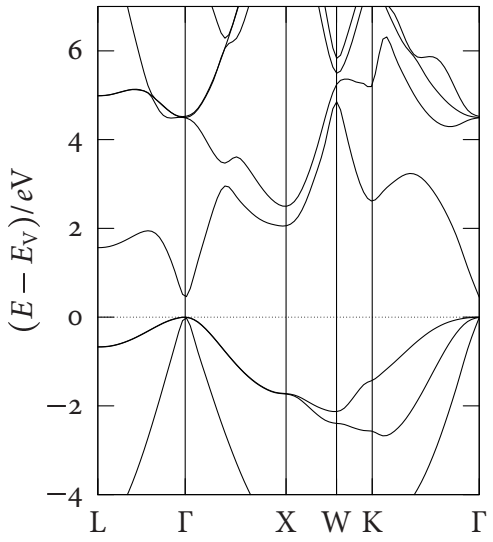
As there are eight electrons in bonding states, four spin-degenerate valence bands are filled: a single band corresponding to the atomic *s* orbital and three bands corresponding to the three *p* orbitals. The latter take part in conduction and are split by spin–orbit interaction into a 'split-off' hole band with total angular momentum $J = \frac{1}{2}$ and two 'heavy' and 'light' hole bands with $J = \frac{3}{2}$, which are degenerate at the Γ-point. The lowest conduction band corresponds to the next highest *s* orbital and is therefore singly degenerate (disregarding spin degeneracy).

5.3.2 Heterostructure Band Alignment

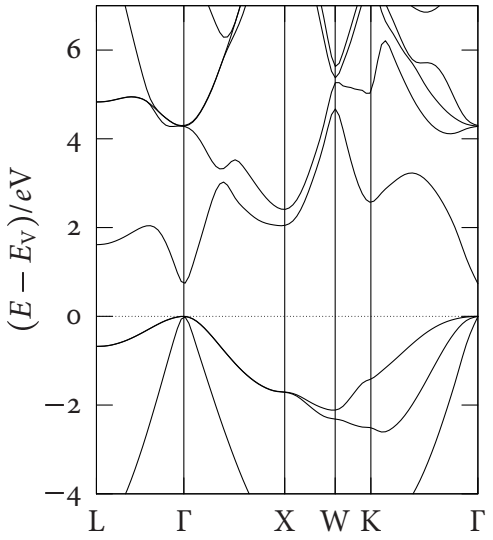
The band line-ups in heterostructures combining these materials, shown schematically in Fig. 5.3, differ drastically from those available in the widely studied GaAs–$Al_xGa_{1-x}As$ system. While the GaSb–AlSb heterojunction is characterized by a conventional type I band alignment in which the narrow gap material has both a higher valence and lower conduction band than the wide gap material, the AlSb–InAs and GaSb–InAs systems exhibit a staggered, or type II alignment [3]. The InAs–AlSb conduction band offset at *c.* 1·35 *e*V is exceptionally large compared to other practical semiconductor heterojunctions, facilitating very deep quantum wells and very high tunnelling barriers. The most remarkable band alignment,



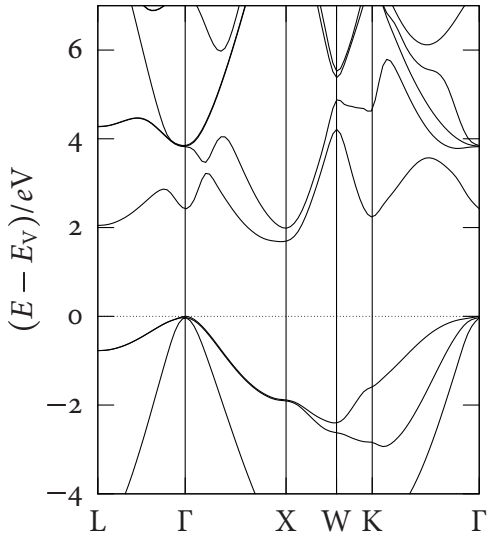
(a) The first Brillouin zone of the zincblende lattice



(b) InAs



(c) GaSb



(d) AlSb

Figure 5.2: bulk InAs, GaSb, and AlSb. The calculation uses the empirical pseudopotential method following Cohen *et al.* [6]. The parameters for AlSb are taken from Topol *et al.* [7]. Spin-orbit interactions and non-local correction terms are not considered; more realistic calculations are available in the literature [8].

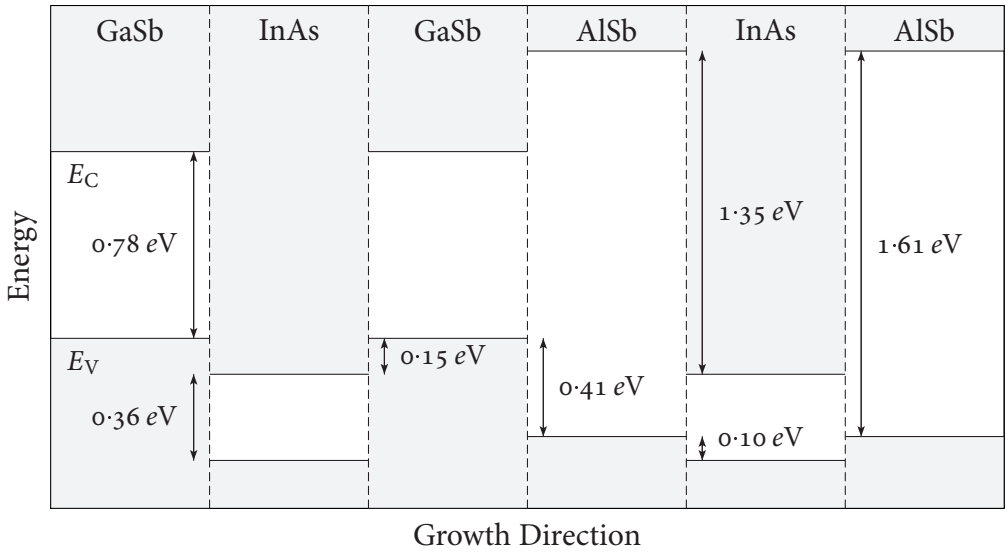


Figure 5.3: InAs, GaSb, and AlSb at room temperature. Adapted from Kroemer [3].

however, is that of GaSb–InAs heterointerfaces: The bottom of the InAs conduction band lies approximately 150 m*e*V below the top of the GaSb valence band, leading to a *broken gap* line-up [12], a configuration that is sometimes classified as 'type III'.

5.3.3 Sample Growth

Starting from initial work at IBM [12], the growth of the 6·1 Å materials by molecular beam epitaxy (MBE) has been developed systematically by a number of researchers, including Herbert Kroemer's group at the University of California [13–15], and can now be regarded as a fairly mature process.

Another viable option is MOVPE, and our research group at Oxford has demonstrated the growth of high quality GaSb–InAs heterostructures [16, 17] using the methods documented in Sec. 5.7. In particular, GaSb–InAs heterostructures which are nearly intrinsic, *i.e.*, in which carriers are created mostly by charge transfer between the GaSb and InAs layers as explained in Sec. 5.5.1, are grown more readily using this technique [18, 19]. However, the deposition of AlSb by MOVPE remains problematic, as most of the potential metal-organic precursors result in high levels of carbon and oxygen contamination. With alternative sources it has been dif-

ficult to control gas phase reactions, making them unsuitable for atmospheric pressure reactors and restricting the combinations of precursors that can be used simultaneously [20–22].

5.3.4 Interface Configuration

At a heterojunction between two semiconductors that do not share a common species, two distinct bond configurations are possible. For example, an atomically flat GaSb-InAs interface may consist of either In–Sb bonds or Ga–As bonds. Which kind of bond is formed depends on the details of the growth process. Methods have been developed for both MBE [14] and MOVPE [23] to control the interface composition to a certain degree. However, even for MBE-grown samples the interface flatness and the preference for one bond will not be perfect.

The two interfaces are not equivalent: Interface states differ, and the different equilibrium bond lengths cause different strain distributions. There is also theoretical and experimental evidence that the band offset at the GaSb–InAs heterojunction depends on the bond configuration [24,25]. The situation is further complicated by the influence of the bonding type–or the methods used to impose it–on the growth processes. The abruptness of the interface and the chemical composition of nearby layers may be affected, and the effect can depend on the ordering of the materials, with significantly different behaviour at the upper and lower interfaces of a quantum well [14, 26].

5.4 InAs–AlSb Heterostructures

The large conduction band offset between InAs and AlSb in combination with the high electron mobility in InAs suggests that it is possible to grow very deep quantum wells containing high-density, high-mobility two-dimensional electron gases (2DEGS) with potential applications to high-speed field-effect transistors (FETS). Considerable effort has been devoted to the optimization of such structures, and mobilities as high as 220,000 $\mathrm{cm}^2/(\mathrm{V\,s})$ at sheet densities of $5{\cdot}5 \times 10^{12}\,\mathrm{cm}^{-2}$ have been obtained [3].

For narrow InAs wells, the mobility is limited by interface roughness scattering and depends sensitively on the interface composition [13, 27]. Experimentally, it was found that predominantly InSb-like interfaces yield higher mobilities and that the heterojunction below the quantum well has a greater influence [13, 14, 26, 28].

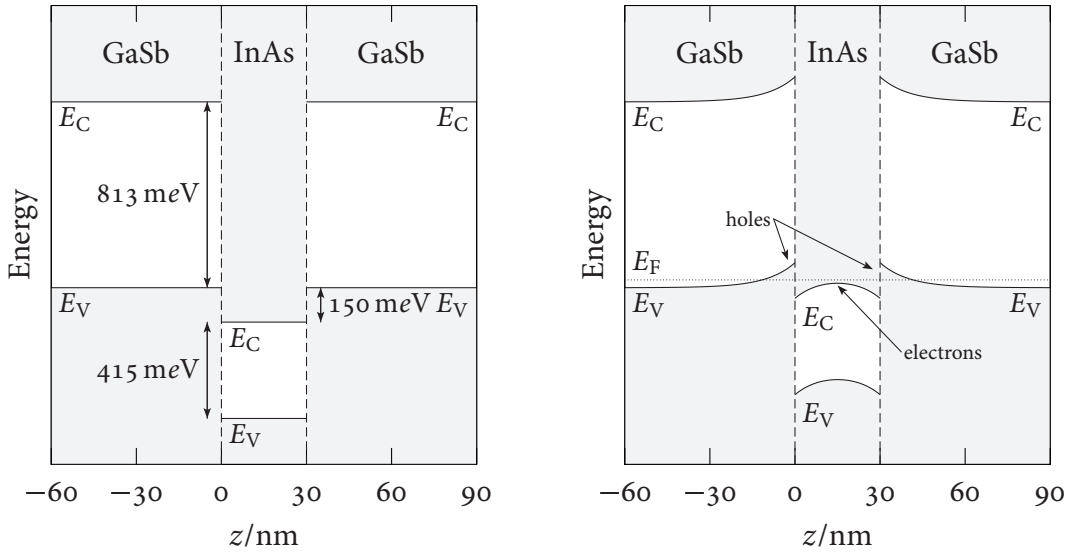
Nominally undoped InAs–AlSb quantum wells can have high electron sheet concentrations of the order 10^{12} cm^{-2}, much in excess of what can be explained by background doping. Because of the depth of the quantum well, which reaches almost down to the AlSb valence band, electrons drain readily into the InAs layer. It has been shown by Nguyen *et al.* [15] that a large fraction of the carrier concentration results from donor states at the surface of the GaSb cap that protects the top AlSb barrier from oxidation. Other proposed sources of excess electrons include Al_{Sb} antisite defects [13] and interface Tamm states [3]. Intentional modulation doping is achieved by embedding a heavily doped layer of AlSb:Te [29] or thin InAs:Si [30] in one or both of the barriers.

Heterostructure FETs (HFETs) with cut-off frequencies of up to 93 GHz have been demonstrated by Bolognesi *et al.* [30–32]. The long mean free paths of the order 10 µm also make high-mobility InAs–AlSb quantum wells attractive model systems for studying ballistic transport in quantum constrictions at low temperatures [33, 34].

5.5 Broken-Gap Heterostructures

5.5.1 Effective Band Gap and Charge Transfer

The broken-gap band line-up of InAs and GaSb makes it possible to create both semiconducting and semimetallic heterostructures with tunable virtual band gaps. As long as the confinement energy of the carriers in the effective electrostatic potential of the heterostructure is small, electrons can lower their energy by dropping from the GaSb valence band to the InAs conduction band until the Fermi level E_F is equal in both layers and the system is in equilibrium. This charge transfer is illustrated in Fig. 5.4 for a symmetric GaSb–InAs–



(a) Juxtaposition of individual band diagrams

(b) Symmetric DHET in equilibrium

Figure 5.4: Band alignment of a GaSb–InAs–GaSb DHET with a 30 nm InAs layer

GaSb DHET similar to the samples studied in this thesis. If the GaSb layer is replaced by a GaSb–AlSb superlattice [35] or an AlSb spacer layer is introduced between the GaSb and InAs layers [36], the charge transfer process can still take place, leaving behind mobile holes in the superlattice or the GaSb layer, respectively.

For narrow well widths, the lowest electron subband energy lies above the highest hole level and the heterostructure is a semiconductor. For wells wider than roughly 85 Å [16, 37], however, charge transfer is energetically favourable and the structures exhibit semimetallic behaviour. The GaSb–InAs–GaSb DHETs studied in this thesis have InAs layers 240 to 300 Å wide and fall well within the latter regime. It is also possible to tune the effective band gap by replacing the GaSb with $Al_xGa_{1-x}Sb$ and varying x.

The initial work on broken-gap heterostructures focused on the semimetal–semiconductor transition and the band gap engineering possibilities in superlattice structures [12, 38]. However, the practical usefulness of the tunable band gap is restricted by the fact that the electrons and holes are spatially separated and only the proportion of the device volume in which their wave functions overlap is optically active [3]. As the effective band gap can be

made arbitrarily small, and in particular smaller than the binding energy E_X of an exciton, it was also suggested that broken-gap heterostructures with an excitionic ground state can be created and used to study exciton gases [39]. However, evidence for an excitonic ground state remains elusive. Cyclotron resonance features originally attributed to excitons [40] are better explained by electron–hole hybridization (*cf.* Sec. 5.5.3), as demonstrated in experiments using AlSb barrier layers between the GaSb and InAs layers to control the wave function overlap and hence the hybridization gap [36].

Because of the quantization of the envelope wave function in the growth direction, the carrier sheets in semimetallic heterostructures form quasi-two-dimensional electron gases (2DEGs) and quasi-two-dimensional hole gases (2DHGs) [38, 41, 42]. These strongly interacting 2D carrier gases have attracted much interest and are also the essential feature of broken-gap systems from the point of view of the present work.

5.5.2 Band Bending and Electron-Hole Ratio

In an intrinsic heterostructure, all carriers are created by charge transfer from the GaSb to the InAs layers and the sheet densities of mobile electrons and holes are the same. In actual samples, localized states at interfaces, defects, or the sample surface can act as donors or acceptors, potentially resulting in unbalanced concentrations of mobile carriers.

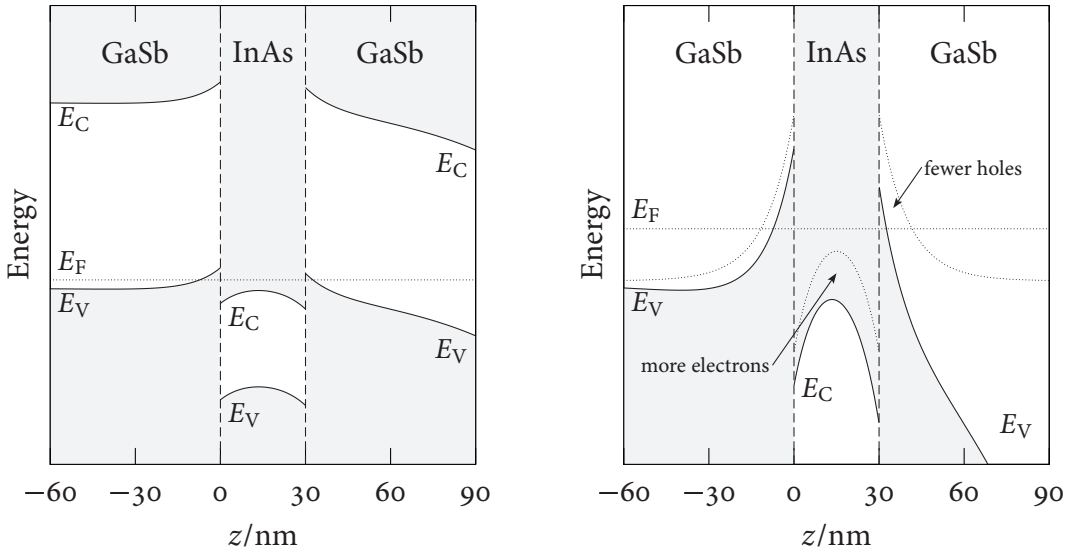
For a large number of experiments it is useful to be able to control the electron–hole ratio. While a limited control of the carrier concentrations is possible by optical pair generation [13, 27] or application of hydrostatic pressure [43, 44], the most flexible and conceptually most simple approach is the use of a gate electrode [45]. The energy bands bend to an extent determined by the voltage at the gate and the screening within the structure. As a result, the position of the Fermi level at the depth of the carrier gases can be changed over a considerable range. However, gating of heterostructures with GaSb top layers represents a certain technological challenge as high quality insulating layers cannot be grown by thermal oxidation of GaSb and even Schottky contacts are difficult to form reliably on this material.

Successful fabrication of insulated gates has been reported using anodic oxide [17, 46] as well as deposited polymers [35, 47] or oxides [48, 49]. Depleting the lower carrier sheet may also necessitate the use of a back gate to avoid the screening caused by the upper sheet [47]. Using top gates, Drndic *et al.* [35] have demonstrated the change from electron-dominated to hole-dominated conduction in a single structure. Yang *et al.* [48] have explored the non-monotonic transfer characteristic associated with this transition in a FET-like device.

Magnetotransport studies of GaSb–InAs–GaSb DHETs grow by MOVPE in Oxford [25, 50, 51] indicate that the electron density is always higher than the hole density. These DHETs lie relatively close (40 to 120 nm) to the surface of the semiconductor, and there is compelling evidence that the increased electron–hole-ratio is caused by surface states at the GaSb–air interface, which pin the Fermi level at a position different from its bulk value.

While the precise position of the surface pinning is generally not known and depends on the chemical composition of the interface, there is some evidence [52] that for surfaces covered with native oxide the Fermi level lies in the lower part of the band gap but above the Fermi level in the bulk material. As a result of the Fermi level pinning, the bands bend towards lower energy near the surface, and if the GaSb cap layer is sufficiently thin, the DHET lies in a region of strong band bending as illustrated in Fig. 5.5(a). The pulling down of the band configuration with respect to the Fermi level E_F means that more conduction band states in the InAs lie below E_F and fewer valence band states in the GaSb lie above E_F, leading to the observed electron–hole imbalance as illustrated in Fig. 5.5(b). The mechanism is hence expected to lead to a dependence of the electron–hole ratio on the GaSb cap thickness, which has been observed experimentally [51]. This dependence affords a limited control over the electron–hole ratio without the use of an insulated gate and should in principle make it possible to impose a potential and carrier density modulation on the carrier sheets by laterally varying the cap layer thickness.

The band bending induced by the surface states means that the DHET is no longer symmetrical. In fact, the GaSb valence band edge at the upper interface lies close to the Fermi energy, and it is questionable whether there are any mobile holes in the cap layer at all [51]. Another



(a) Effect of donor states at the surface

(b) Comparison with the symmetric DHET (dotted bands) at a magnified energy scale

Figure 5.5: Band bending due to surface states

possible source of band asymmetry is the dependence of the band overlap on the GaSb–InAs interface composition: As both the group III and the group V elements change at the interface, either a monolayer of InSb or GaAs may be formed, and it has been shown that interfaces that are predominantly of the GaSb type have a band overlap approximately 30 m*e*V smaller than those of the InSb type [44, 53]. If the interface types on both sides of the InAs layer are different, the resulting DHET will have an asymmetric band line-up even in the absence of surface states [25]. The samples discussed in this thesis did not have interfaces biased either towards InSb or GaSb, but the presence of a residual imbalance could not be excluded.

5.5.3 Hybridization and Minigap

The electron and hole dispersion relations $E_i(\mathbf{k})$, where $\mathbf{k}$ is the in-plane wave vector and i identifies the band, overlap as a result of the band alignment described in Sec. 5.3.2. Any finite coupling $\Delta_{ij}/2$ between electron and hole states will therefore lead to the formation of an energy gap Δ_{ij} at the crossing points. A simple two-band model can be used to capture

the essential behaviour of the hybridized system [42,54]: In this model, the electron band is assumed to be isotropic and is given by

$$E_{\mathrm{e}}(\mathbf{k}) = \frac{\hbar^2\mathbf{k}^2}{2m^*}, \tag{5.7}$$

where $m^* = 0{\cdot}03m_{\mathrm{e}}$ is the electron effective mass, while the warped heavy hole band is

$$E_{\mathrm{h}}(\mathbf{k}) = E_{\mathrm{g}} - \frac{\hbar^2}{m_{\mathrm{e}}}\left(|A|\mathbf{k}^2 - \sqrt{B^2\mathbf{k}^4 + C^2k_x^2k_y^2}\right), \tag{5.8}$$

where E_{g} is the band overlap and $A = -14{\cdot}3$, $B = -10{\cdot}4$, and $C = 13{\cdot}5$ are anisotropy parameters [55]. The Schrödinger equation for the coupled system leads to the eigenvalue problem

$$\begin{pmatrix} E_{\mathrm{e}}(\mathbf{k}) & \Delta_{\mathrm{eh}}/2 \\ \Delta_{\mathrm{eh}}/2 & E_{\mathrm{h}}(\mathbf{k}) \end{pmatrix}\begin{pmatrix} c_{\mathrm{e}} \\ c_{\mathrm{h}} \end{pmatrix} = E(\mathbf{k})\begin{pmatrix} c_{\mathrm{e}} \\ c_{\mathrm{h}} \end{pmatrix}, \tag{5.9}$$

where c_{e} and c_{h} are the coefficients of the electron and hole wave functions, respectively. The mixed energy bands obtained by solving Eq. (5.9) are

$$E_{\pm}(\mathbf{k}) = \frac{1}{2}\left[E_{\mathrm{e}}(\mathbf{k}) + E_{\mathrm{h}}(\mathbf{k})\right] \pm \frac{1}{2}\sqrt{\left[E_{\mathrm{e}}(\mathbf{k}) + E_{\mathrm{h}}(\mathbf{k})\right]^2 + \Delta_{\mathrm{eh}}^2}. \tag{5.10}$$

Because of this mixing, which opens an energy *minigap* at the Fermi energy E_{F}, it was originally expected that intrinsic heterostructures should be semiconducting [38]. However, the anisotropy of the hole band leads to the formation of disconnected Fermi contour islands for most values of E_{F} (*cf.* Fig. 5.7(a) below), reducing the effective minigap and explaining the metallic behaviour seen in most DHET samples [42].

The strength of the electron–hole interaction $\Delta_{\mathrm{eh}}/2$ can be modified either by reducing the overlap between the electron and hole wave functions or by applying a magnetic field in the plane of the heterojunction. An effective means of decreasing the overlap is to introduce an AlSb barrier between the GaSb and the InAs layers. The additional layer increases the spatial separation of the carrier gases; at the same time, the large AlSb band gap ensures that the penetration of the electron and hole states into the barrier is small. If samples with different barrier thicknesses are prepared, Δ_{eh} can be varied systematically [36, 47]. The in-plane magnetic field makes it possible to switch the interaction on and off in a single structure, but does not provide for continuous tuning of Δ_{eh}. The process by which it affects the coupling requires some explanation.

In the presence of a magnetic field, and neglecting spin, the Schrödinger equation (5.4) becomes

$$\frac{\left[\hat{\mathbf{p}} + e\mathbf{A}(\mathbf{r})\right]^2}{2\overline{m}^*}\varphi(\mathbf{r}) + V(z)\varphi(\mathbf{r}) = E_{\text{tot}}\varphi(\mathbf{r}), \tag{5.11}$$

where $\mathbf{A}(\mathbf{r})$ is the magnetic vector potential. If the magnetic field is homogeneous and lies in the plane of the heterostructure along the y-axis, *i.e.*, $\mathbf{B} = (0, B_y, 0)$, the magnetic vector potential can be chosen as $\mathbf{A}(\mathbf{r}) = (B_y z, 0, 0)$ in the Landau gauge.[1] Eqs. (5.1) and (5.11) then lead to the following Schrödinger-like equation in $\chi(z)$ [56]:

$$\frac{\hbar^2}{2m^*}\left(k_x^2 + \frac{2eB_y k_x z}{\hbar} + \frac{e^2 B_y^2 z^2}{\hbar^2} + k_y^2 - \frac{\partial^2}{\partial z^2}\right)\chi(z) + V(z)\chi(z) = E_{\text{tot}}\chi(z). \tag{5.12}$$

Eq. (5.12) is of the form

$$\hat{H}_0\chi(z) + \hat{H}_1\chi(z) = E_{\text{tot}}\chi(z), \tag{5.13}$$

where the Hamiltonian $\hat{H}_0$ in the absence of a magnetic field is given by Eq. (5.6) and

$$\hat{H}_1 = \frac{e^2 B_y^2 z^2}{2m^*} + \frac{\hbar e B_y k_x z}{m^*} \tag{5.14}$$

is the contribution due to B_y [57].

As long as $\hbar e B_y / 2m^* \ll E_0$—which will typically be fulfilled for InAs well thicknesses smaller than 400 Å in the samples discussed here [42, 58]—$\hat{H}_1$ can be treated as a perturbation, leading to a first order energy correction [56, 57]

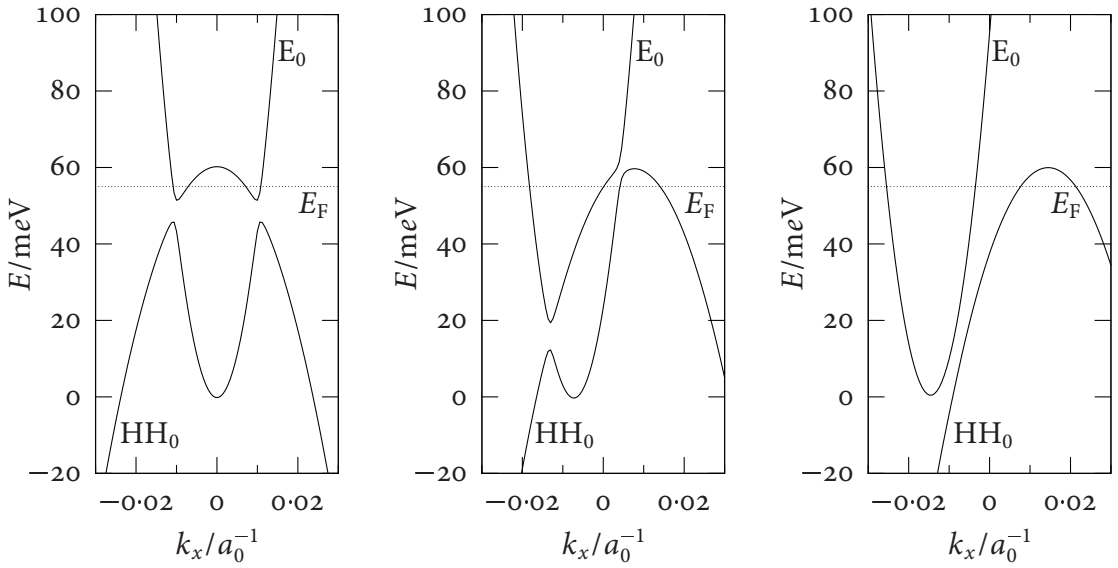
$$\langle\chi_n|\hat{H}_1|\chi_n\rangle = \frac{e^2 B_y^2}{2m^*}\langle z^2\rangle_n + \frac{\hbar e B_y k_x}{m^*}\langle z\rangle_n, \tag{5.15}$$

where $\langle z\rangle_n = \langle\chi_n|z|\chi_n\rangle$ and $\langle z^2\rangle_n = \langle\chi_n|z^2|\chi_n\rangle$ denote the expectation values of z and z^2. Rearrangement of $E_{\text{tot}} = E_n + E_{xy} + \langle\chi_n|\hat{H}_1|\chi_n\rangle$ gives [56, 59]

$$E_{\text{tot}} = E_n + \frac{\hbar^2}{2m^*}\left[\left(k_x + \frac{eB_y}{\hbar}\langle z\rangle_n\right)^2 + k_y^2\right] + \frac{e^2 B_y^2}{2m^*}\left(\langle z^2\rangle_n - \langle z\rangle_n^2\right). \tag{5.16}$$

From Eq. (5.16) it can be seen that the overall effect of the in-plane field B_y is a shift δk_x of the dispersion relation along the x-direction, which is proportional to the position of

[1] The argument can be extended straightforwardly to $\mathbf{B} = (B_x, B_y, 0)$. $\mathbf{A}(\mathbf{r})$ is then $(B_y z, -B_x z, 0)$.



(a) $B_y = 0\,\mathrm{T}$ (b) $B_y = 10\,\mathrm{T}$ (c) $B_y = 20\,\mathrm{T}$

Figure 5.6: Electron and hole dispersion in an in-plane magnetic field, calculated for a DHET with $\Delta = 7\,\mathrm{m}e\mathrm{V}$, $E_\mathrm{g} = 60\,\mathrm{m}e\mathrm{V}$, and $\delta z = 180\,\text{Å}$. Wave numbers are given in terms of the Bohr radius a_0.

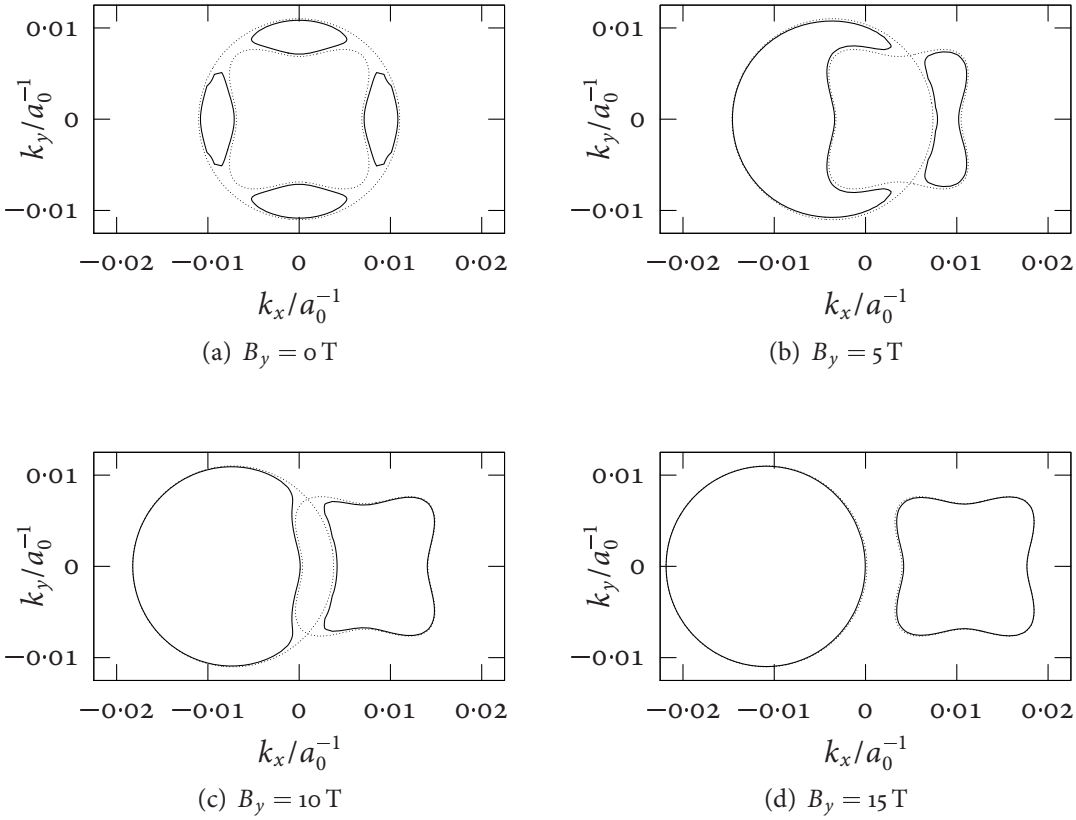
the electron sheet with respect to the coordinate origin, and a diamagnetic shift δE_n of the subband energy, which depends on the spread of $\chi_n(z)$ [25, 56, 58]:

$$\delta k_x = \frac{eB_y}{\hbar}\langle z\rangle_n; \qquad \delta E_n = \frac{e^2B_y^2}{2m^*}\left(\langle z^2\rangle_n - \langle z\rangle_n^2\right) \tag{5.17}$$

In the electron-hole system, $\langle z\rangle_n$ can be identified with the distance δz between the electron and hole gases along z, while δE can be neglected [25, 42, 51]; this is in agreement with experiments that have shown little dependence of the subband energies on an in-plane magnetic field [58].

Fig. 5.6 illustrates the effect of the magnetic field on the minigap: Plots of Eq. (5.10) with $E_\pm(k_x, k_y)$ replaced by $E_\pm(k_x \pm \delta k_x/2, k_y)$ are shown for different field strengths and typical values for E_g, E_F, and δz. With increasing field, the crossing points move away from the Fermi level and for sufficiently strong fields, the dispersion curves become decoupled and the minigap disappears.

The effect on the Fermi contours is shown in Fig. 5.7. The relative shift of the disper-



(a) $B_y = 0\,\mathrm{T}$ (b) $B_y = 5\,\mathrm{T}$

(c) $B_y = 10\,\mathrm{T}$ (d) $B_y = 15\,\mathrm{T}$

Figure 5.7: Fermi contours in an in-plane magnetic field, calculated for a DHET with $\Delta = 7\,\mathrm{m}e\mathrm{V}$, $E_\mathrm{g} = 60\,\mathrm{m}e\mathrm{V}$, $\delta z = 180\,\text{Å}$, and $E_\mathrm{F} - E_\mathrm{e}(0) = 55\,\mathrm{m}e\mathrm{V}$. Wave numbers are given in terms of the Bohr radius a_0.

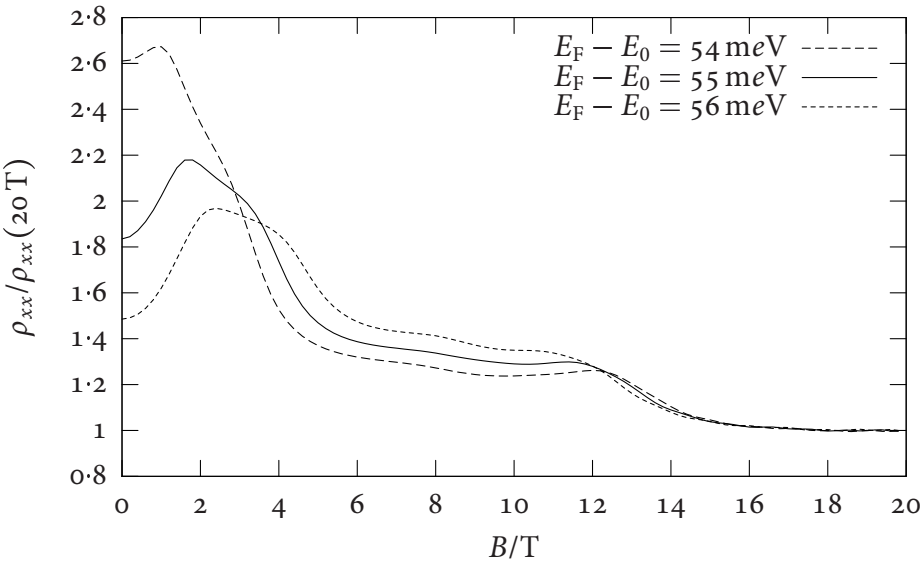


Figure 5.8: Negative magnetoresistance in an in-plane magnetic field, calculated for a DHET with $\Delta = 7\,meV$, $E_g = 60\,meV$, and $\delta z = 180\,Å$. The resistivity is given in terms of the resistivity at high field.

sion relations changes the shape of the contours drastically, varying from the small 'puddles' present at zero field to nearly unperturbed independent electron and hole Fermi contours at high field. Owing to the anisotropy of the hole band, the shape in the intermediate region will depend on the orientation of the in-plane field.

The in-plane conductivity, and hence resistivity, can be calculated from the Boltzmann equation (see Appendix B), which leads to an expression of the form

$$\sigma_{xx} = -\frac{2e^2}{(2\pi)^2} \sum_i \int \left[v_x^{(i)}(\mathbf{k})\right]^2 \tau f'\left[E_i(\mathbf{k})\right] d\mathbf{k}, \tag{5.18}$$

where $v_x^{(i)}(\mathbf{k})$ is the x-component of the group velocity, $f'(E)$ the derivative of the Fermi-Dirac distribution, and the transport relaxation time τ is taken to be independent of $\mathbf{k}$ [42]. Fig. 5.8 shows the resistivity $\rho_{xx} = 1/\sigma_{xx}$ (since $\rho_{xy} = 0$) calculated by numerical integration from Eq. (5.18) with the simple two-band model of Sec. 5.5.3 and using the same typical values for the sample parameters as in Figs. 5.6 and 5.7. Just as the Fermi contours, the magnetoresistance is highly dependent on the band overlap (not shown) and the Fermi level. Peaks correspond to the field strengths where anticrossing occurs at the Fermi energy, *i.e.*, $\delta k_x(B_y) \approx$

$|k_x^{(e)}(E_F)| - |k_x^{(h)}(E_F)|$ (around 2 T in Fig. 5.8) and $\delta k_x(B_y) \approx |k_x^{(e)}(E_F)| + |k_x^{(h)}(E_F)|$ (around 12 T in Fig. 5.8).

Several studies of GaSb–InAs superlattices and DHETs have shown a large decrease of up to 70 % of the two-dimensional resistivity in the presence of a strong in-plane magnetic field. [42, 50, 54, 58]. The measured magnetoresistance and its temperature dependence are described well by calculations such as those of Fig. 5.8, providing strong evidence for the existence of the minigap. A systematic study of a gated GaSb–InAs heterostructure by COOPER *et al.* [47] has revealed as strong increase in the longitudinal resistance for equal electron and hole concentrations. The temperature dependence of the peak in this experiment indicated an energy gap of *c.* 2 m*e*V, and the resistance increase was not observed in a second sample that was nominally identical save for an AlSb barrier suppressing the electron–hole coupling. POULTER *et al.* [58] have reported optical absorption peaks corresponding to excitation across the minigap, supporting the magnetotransport results. Evidence for a small positive band gap, which disappears in an in-plane electric field, has also been seen in capacitance measurements by YANG *et al.* [60].

5.5.4 Magnetic Breakdown

In a perpendicular magnetic field, carriers follow orbits in the plane of the DHET that in momentum space correspond to orbits along the Fermi contours shown in Fig. 5.7. As a result of the mixing between electron and hole states, the Fermi contour consists of several unconnected islands and the possible closed orbits include both electron- and hole-like states.

The hole states are characterized by a larger effective mass compared to the electron states, so that the carriers take a longer time to complete the hole-like parts of the orbits and are more likely to leave them because of scattering. On the other hand, once the cyclotron energy $\hbar\omega_c$ becomes comparable to the minigap Δ, carriers can tunnel readily between electron-like parts of the Fermi contour, forming an orbit with low effective mass similar to the electron orbit in the absence of electron–hole mixing [61].

The Shubnikov–de Haas-oscillations described in Sec. 5.6.3 and the commensurability features in the presence of a lateral modulation discussed in Chapter 7 are dominated by such *magnetic breakdown orbits* [50], and the evidence for a direct effect of the disjoint Fermi contour islands remains tenuous [51].

5.6 Magnetotransport in a Perpendicular Magnetic Field

5.6.1 Drude Model

It is tempting to describe the parallel magnetotransport in the presence of a perpendicular magnetic field $\mathbf{B} = (0, 0, B_z)$ in terms of a simple semiclassical two-carrier model based on the Drude formalism [25, 50, 51]. In this model, the relaxation time approximation is used and the electrons and holes are assumed to be independent and have individually well-defined conductivities $\boldsymbol{\sigma}_\mathrm{e}$ and $\boldsymbol{\sigma}_\mathrm{h}$. The total conductivity of the electron–hole system would then simply be

$$\boldsymbol{\sigma} = \boldsymbol{\sigma}_\mathrm{e} + \boldsymbol{\sigma}_\mathrm{h}. \tag{5.19}$$

In the Drude model [2], the equation of motion for a particle of species $i \in \{\mathrm{e}, \mathrm{h}\}$ is

$$m_i^* \dot{\mathbf{v}} = q_i \left(\mathbf{E} + \mathbf{v} \times \mathbf{B}\right) - m_i^* \frac{\mathbf{v}}{\tau_i}, \tag{5.20}$$

where $\mathbf{E}$ is the electric field, $\mathbf{v}$ the velocity, m_i^* the effective mass, q_i the charge, and scattering is accounted for by the phenomenological friction term $(m_i^*/\tau_i)\mathbf{v}$ with scattering relaxation time τ_i. If $\mathbf{E} = (E_x, E_y, 0)$, $\mathbf{v} = (v_x, v_y, 0)$, and $\dot{\mathbf{v}} = \mathbf{0}$ Eq. (5.20) simplifies to

$$\frac{v_x}{\operatorname{sgn}(q_i)\mu_i} = E_x + v_y B_z; \qquad \frac{v_y}{\operatorname{sgn}(q_i)\mu_i} = E_y - v_x B_z, \tag{5.21}$$

where $\mu_i \stackrel{\mathrm{def}}{=} |q_i|\tau_i/m_i^*$ is the mobility. Eq. (5.21) has the solution

$$v_x = \frac{\operatorname{sgn}(q_i)\mu_i E_x + \mu_i^2 B_z^2 E_y}{1 + \mu_i^2 B_z^2}; \qquad v_y = \frac{-\mu_i^2 B_z^2 E_x + \operatorname{sgn}(q_i)\mu_i E_y}{1 + \mu_i^2 B_z^2}. \tag{5.22}$$

As the conductivity $\boldsymbol{\sigma}$ is defined by $\mathbf{j} = \boldsymbol{\sigma}\mathbf{E}$ and $\mathbf{j}_i = q_i n_i \mathbf{v}$, where n_i is the carrier concentration, the (x, y)-components of the conductivity tensor $\boldsymbol{\sigma}_i$ are

$$\sigma_{xx}^{(i)} = \sigma_{yy}^{(i)} = \frac{|q_i| n_i \mu_i}{1 + \mu_i^2 B_z^2}; \qquad \sigma_{xy}^{(i)} = -\sigma_{yx}^{(i)} = \frac{q_i n_i \mu_i^2 B_z}{1 + \mu_i^2 B_z^2}. \tag{5.23}$$

For a two-carrier system consisting of electrons and holes Eq. (5.19) together with $\rho_{xx} = \sigma_{xx}/(\sigma_{xx}^2 + \sigma_{xy}^2)$ and $\rho_{xy} = -\sigma_{xy}/(\sigma_{xx}^2 + \sigma_{xy}^2)$ then leads to the following expressions for the in-plane resistivity:

$$\rho_{xx} = \rho_{yy} = \frac{1}{e} \frac{n_\mathrm{e}\mu_\mathrm{e}(1 + \mu_\mathrm{h}^2 B_z^2) + n_\mathrm{h}\mu_\mathrm{h}(1 + \mu_\mathrm{e}^2 B_z^2)}{(n_\mathrm{e}\mu_\mathrm{e} + n_\mathrm{h}\mu_\mathrm{h})^2 + \mu_\mathrm{e}^2\mu_\mathrm{h}^2 B_z^2 (n_\mathrm{e} - n_\mathrm{h})^2}; \tag{5.24}$$

$$\rho_{xy} = -\rho_{yx} = \frac{1}{e} \frac{n_\mathrm{e}\mu_\mathrm{e}^2 B_z(1 + \mu_\mathrm{h}^2 B_z^2) - n_\mathrm{h}\mu_\mathrm{h}^2 B_z(1 + \mu_\mathrm{e}^2 B_z^2)}{(n_\mathrm{e}\mu_\mathrm{e} + n_\mathrm{h}\mu_\mathrm{h})^2 + \mu_\mathrm{e}^2\mu_\mathrm{h}^2 B_z^2 (n_\mathrm{e} - n_\mathrm{h})^2}. \tag{5.25}$$

Eqs. (5.24) and (5.25) exhibit the positive magnetoresistance $\rho_{xx}(B_z)$ and non-linear Hall resistivity $\rho_{xy}(B_z)$ which are characteristic of bipolar systems and have been demonstrated in GaSb–InAs DHETs [50, 51]. Information about n_e, n_h, μ_e, and μ_h could in principle be obtained by fitting the functional dependence on B_z described by these equations to the measured magnetotransport traces.[2]

However, the assumption of independent electrons and holes is ultimately false in the light of the phenomena described in Sec. 5.5.3 and severely limits the applicability of the model to the system investigated here. Depending on the electron–hole ratio and the Fermi level at the DHET, the minigap causes a substantially higher resistivity than scattering alone, and the mobilities are significantly underestimated if such a fit is attempted. Only in the presence of an in-plane magnetic field that is strong enough to completely decouple the electrons and holes the Drude model is a realistic approximation.

[2] Eqs. (5.24) and (5.25) are of the form $y(x) = a(1 + bx^2)/(1 + cx^2)$ and $\eta(x) = \alpha x(1 + \beta x^2)/(1 + \gamma x^2)$, respectively. Each has only three independent parameters, and hence a fit to either $\rho_{xx}(B_z)$ or $\rho_{xy}(B_z)$ cannot be used to determine all four parameters n_e, μ_e, n_h, and μ_h unless one of them is known separately, *e.g.*, n_e from the Shubnikov-de Haas oscillations. If both equations are used, $\gamma = c$, $\alpha = a\sqrt{c}(c - b)/(c - \beta)$, and β remains variable.

5.6.2 Landau Quantization

For a homogeneous magnetic field perpendicular to the (x, y)-plane, *i.e.*, $\mathbf{B} = (0, 0, B_z)$, the magnetic vector potential can be chosen as $\mathbf{A}(\mathbf{r}) = (0, B_z x, 0)$ in the Landau gauge $\nabla_{\mathbf{r}} \cdot \mathbf{A}(\mathbf{r}) = 0$. With the *Ansatz*

$$\varphi(\mathbf{r}) = e^{ik_y y}\chi(z)\psi(x), \tag{5.26}$$

Eq. (5.11) again becomes separable. The equation for $\chi(z)$ is given by Eqs. (5.5) and (5.6), while $\psi(x)$ has to obey

$$\frac{\hbar^2}{2\overline{m}^*}\left(-\frac{\partial^2}{\partial x^2} + k_y^2 + \frac{2eB_z k_y x}{\hbar} + \frac{e^2 B_z^2 x^2}{\hbar^2}\right)\psi(x) + E_n\psi(x) = E_{\text{tot}}\psi(x), \tag{5.27}$$

which simplifies to

$$\left(-\frac{\hbar^2}{2\overline{m}^*}\frac{\partial^2}{\partial \xi^2} + \frac{e^2 B_z^2 \xi^2}{2\overline{m}^*}\right)\psi(\xi) = E_{xy}\psi(\xi) \tag{5.28}$$

with $\xi = x + \hbar k_y / eB_z$ and $E_{xy} = E_{\text{tot}} - E_n$. Eq. (5.28) is the Schrödinger equation of a one-dimensional quantum harmonic oscillator with eigenvalues

$$E_{xy}^{(N)} = \left(N + \frac{1}{2}\right)\hbar\omega_c, \tag{5.29}$$

where the quantum number N is known as the Landau level and $\omega_c = eB_z/\overline{m}^*$ is the cyclotron frequency.

As a result of the quantization, all electron states must be accommodated in the discrete Landau levels. The degeneracy of the N^{th} Landau level can hence be calculated by considering the number of states that would lie between $E_{xy}^{(N-1)}$ and $E_{xy}^{(N)}$ in the absence of a magnetic field. For the quadratic dispersion relation $E = \hbar(k_x^2 + k_y^2)/2\overline{m}^*$, which follows from Eq. (5.28) if $B_z = 0$, it is straightforward to show [1, 62] that the density of states per unit area of a two-dimensional system is $D(E) = \overline{m}^*/2\pi\hbar^2$ per spin state. In this case, the degeneracy of each Landau level is

$$g_{\text{LL}}(B_z) = \int_{E_{xy}^{(N-1)}}^{E_{xy}^{(N)}} D(E)\, dE = \frac{\overline{m}^*}{2\pi\hbar^2}\hbar\omega_c = \frac{eB_z}{h} = \frac{B_z}{\Phi_0^{(\text{D})}} \tag{5.30}$$

per unit area, where $\Phi_0^{(\mathrm{D})} = h/e$ is the Dirac flux quantum.[3] If n_e is the electron density per unit area, the number of filled spin-split Landau levels at a given field B_z is consequently

$$\nu(B_z) = \frac{n_\mathrm{e}}{g_\mathrm{LL}(B_z)} = \frac{n_\mathrm{e}}{B_z/\Phi_0^{(\mathrm{D})}} = \frac{n_\mathrm{e}}{eB_z/h}. \tag{5.31}$$

The quantity $\nu(B_z)$ is known as the (Landau level) filling factor and corresponds to the number of electrons divided by the number of flux quanta.

5.6.3 Shubnikov–de Haas-Oscillations

According to Eq. (5.29), in a magnetic field B_z perpendicular to the plane of the DHET each subband splits into a number of Landau levels that are equally spaced in Energy by $\hbar\omega_\mathrm{c} = \hbar eB_z/\overline{m}^*$. Each Landau level is further split because of the interaction $g^*\mu_\mathrm{B}\mathbf{B}\cdot\boldsymbol{\sigma}$ of the electron spin with the external magnetic field $\mathbf{B}$, which causes an energy shift $\delta E_{xy}^{(N)} = \pm\frac{1}{2}g^*\mu_\mathrm{B}B_z$; here μ_B is the Bohr magneton, g^* is the *effective* gyromagnetic factor of the subband, and $\boldsymbol{\sigma}$ a vector with the Pauli spin matrices as its elements. The corresponding density of states has a peak for each (spin-split) Landau level, which is broadened thermally and by scattering; the contribution due to scattering is $\hbar/\tau$, where τ is the relaxation time. As long as $\omega_\mathrm{c}\tau \gg 1$ and $\hbar\omega_\mathrm{c} \gg k_\mathrm{B}T$, the result is a density of states $D(E)$ that oscillates as a function of energy [1]. If the magnetic field B_z is varied, the period $\hbar\omega_\mathrm{c} = \hbar eB_z/\overline{m}^*$ of this oscillation varies and the Fermi energy E_F successively encounters different Landau levels and a varying density of states. The density of states reaches a minimum if E_F lies between two Landau levels. If spin splitting is not resolved, this occurs at the fields for which the filling factor $\nu(B_z) = \Phi_0^{(\mathrm{D})}n_\mathrm{e}/B_z$ is an even integer, leading to changes in a number of electronic properties that are periodic in $1/B_z$ with a period

$$\Delta_{1/B_z} = \frac{2}{\Phi_0^{(\mathrm{D})}n_\mathrm{e}} = \frac{2e}{n_\mathrm{e}h}. \tag{5.32}$$

[3] Not to be confused with the superconducting flux quantum $\Phi_0^{(\mathrm{s})} = h/2e$.

The periodic change of the magnetic susceptibility is called the de Haas–van Alphen-effect, while the corresponding magnetoresistance fluctuations are known as Shubnikov–de Haas-oscillations.

The oscillations are a useful tool for the characterization of GaSb–InAs DHETs. In these structures, a single $1/B_z$ oscillation frequency is typically observed in the magnetoresistance [50, 51], which can be attributed to electron-like magnetic breakdown orbits as explained in Sec. 5.5.4 and hence corresponds to the InAs conduction band. In this situation, Eq. (5.32) provides an accurate estimate of the two-dimensional electron density.

5.7 Sample Growth using Metal-Organic Vapour Phase Epitaxy

5.7.1 Motivation

Semiconductor heterostructures can be grown epitaxially[4] on the surface of suitable substrate wafers by several different techniques. The most widely used technology is molecular beam epitaxy (MBE), which uses ballistic beams of atoms or molecules created by controlled evaporation of precursor materials in ultra-high vacuum (UHV). The samples discussed in this thesis were grown in Oxford by a variety of vapour phase epitaxy (VPE) that is known as metal-organic vapour phase epitaxy (MOVPE) as it uses metal-organic precursor materials. A third option is liquid phase epitaxy (LPE).

MBE produces high quality materials and allows excellent reproducibility and control of the epitaxial growth down to monolayer accuracy.[5] Its drawback is the reliance on UHV, which not only imposes high demands on the quality of the vacuum equipment but also reduces its flexibility: If the vacuum in the growth chamber is compromised for any reason, reestablishing it usually takes several weeks. MOVPE can grow semiconductors with similar and in some cases better quality while operating at pressures that do not present a technological challenge. However, monolayer accuracy cannot be achieved and reproducibility can

[4] From Greek ἐπί, 'upon', and τάττειν, 'to arrange'. In Greek, ἐπίταξις actually means 'command'.
[5] Abrupt interfaces, however, may not be thermodynamically stable.

be an issue as the growth depends on a large number of parameters. Also, the metal-organic precursors are highly toxic[6] and costly hazard management is necessary.

5.7.2 Principles of Metal-Organic Vapour Phase Epitaxy

MOVPE was first described in 1969 by MANASEVIT and SIMPSON [63]. In this technique, the substrate wafer is placed in a reaction chamber on a heated sample holder. A continuous flow of hydrogen carrier gas, in which a controlled amount of the precursor materials has been dissolved, is maintained at the substrate. The metal-organic precursors are cracked by pyrolysis[7] at the elevated temperatures occurring close to the sample surface. Metal atoms then diffuse across the low-velocity boundary layer to the substrate, where they are adsorbed at the surface and lose any remaining organic groups. These groups react to form volatile organic compounds, while the metals diffuse along the surface until they are incorporated in the crystal lattice.

The controlled solutions of the precursors in hydrogen are obtained by driving dry hydrogen gas through bubblers each containing a liquid precursor. The bubblers are kept at a well-defined temperature and the H_2 gas passing through them takes up precursor molecules and forms a saturated solution, which is then mixed with pure hydrogen at the desired dilution. The dilution ratio and the flows of the process gases are regulated with the help of mass flow controllers. Individual precursor solutions can be quickly switched into and out of the mixing manifold that feeds the process chamber.

For epitaxial growth to be possible, the substrate must be free from contaminants and oxide. This can be achieved by creating an oxide layer with desirable properties by etching and controlled oxidation.[8] The oxide layer is then removed by thermal desorption in the MOVPE reactor in the presence of either pure hydrogen or the process gas providing the substrate species with the highest vapour pressure to prevent the preferential loss this element.

[6] In the proper sense of being quite lethal should they escape during normal operation.
[7] from Greek πῦρ, 'fire' and λύειν 'to dissolve'
[8] Nowadays, so-called 'epi-ready' wafers with suitable surfaces are available commercially, so that this process need not be performed in the laboratory.

The main controlling parameters of MOVPE growth are the relative flow rates of the process gases and the substrate temperature, while the fluid dynamics of the mixing of the individual gases and the flow in the process chamber also play an important part. Different reactor geometries in which the sample holder is oriented either parallel or perpendicular to the gas flow can be used.

5.7.3 Growth of Epitaxial GaSb–InAs Heterostructures in Oxford

The samples discussed in this thesis were grown in the CLARENDON LABORATORY in Oxford by NIGEL MASON and PHILIP SHIELDS. The reactor, which has now been decommissioned, operated near atmospheric pressure and used a reaction chamber made from glass in which the sample surface was oriented horizontally and aligned parallel to the main gas flow [37]. The substrate wafers rested on a molybdenum susceptor which was heated from underneath by a resistive heater cut from a sheet of manganese metal. Despite the fact that the temperature was monitored with a thermocouple, there is a large uncertainty of at least 30 °C in the measurement of the absolute growth temperature; the reproducibility and relative accuracy, however, was much better unless the replacement of the heater became necessary.

GaSb–InAs heterostructures were grown on lattice-mismatched (by approximately 7·3 %) semi-insulating undoped GaAs wafers. Such wafers are readily procurable commercially, while good quality GaSb wafers are still expensive. More importantly, GaSb wafers are inevitably *p*-type, which is undesirable for electrical measurements on epitaxially grown devices [16, 64, 65]. Following a thin GaAs layer and a low temperature heteroepitaxial GaSb layer, a relaxed GaSb buffer layer of roughly 2 μm thickness was grown on the GaAs substrate, creating a *virtual substrate* for the growth of the heterostructures. Under optimal conditions, much of the strain caused by the large lattice mismatch at the GaAs–GaSb interface is accommodated by 90° misfit dislocations in the plane of the interface [65]. Together with the mutual annihilation of threading dislocations during the growth of the buffer layer [66], this results in a virtual substrate relatively free of dislocations. The surface morphology of

such GaSb layers was generally characterized by flat elongated pyramidal shapes 10 to 30 nm high and several micrometres wide. The DHETs grown on these virtual substrates closely follow the shape of the surface; on the scale of these features, the surface roughness was *not* correlated with electrical indicators of the epilayer quality. While the pyramids may affect transport in the DHETs by causing a local tilt with respect to the plane of the sample and by changing the states at the resulting kink lines, the overall effect is assumed to be small.

Trimethylgallium (TMGa), trimethylantimony (TMSb), trimethylindium (TMIn), and tertiarybutylarsine (TBAs) were used as precursors. For growth of GaSb, the ratio of gallium to antimony at the surface needs to be close to 1 as Sb has a very low vapour pressure compared to Ga and would otherwise be precipitated at the growth front. A further complication is caused by the incomplete pyrolysis of the precursors at typical GaSb growth temperatures, leading to a growth rate that is dependent on temperature, *i.e.*, kinetically controlled [16, 22]. Fortunately the pyrolysis rate temperature dependence of TMGa and TMSb are very similar, so that the precursor ratio required at the gas inlet depends only weakly on temperature [16]. The 'epi-ready' GaAs wafers were deoxidized at 650 to 700 °C in the presence of TBAs; GaSb and InAs were typically grown at 525 to 600 °C and 500 to 525 °C, respectively. However, the values were affected by a relatively large absolute uncertainty as explained above, and the optimal growth conditions had to be determined anew after any significant modification to the reactor.

The crystal growth could be monitored *in situ* with the help of a surface photoabsorption (SPA) setup. In this method, a laser is incident on the semiconductor surface at an angle close to Brewster's angle, and the intensity of the reflected light with its (electric field) polarization parallel to the surface is recorded. In our reactor, a 633 nm laser was used at Brewster's angle for GaAs (75·3° [67]) and the reflected signal was detected by a silicon photodiode using a lock-in technique [68, 69]. Information on the surface roughness, surface composition, and heterostructure layer thickness could be obtained. SPA is sensitive to the surface material via the difference in Brewster's angle, and to epilayer thickness via interference between light

reflected from the surface and the heterointerface, which results in Fabry-Pérot oscillations. The measurement of the layer thickness in particular was an important tool in the calibration of the growth rates.

5.8 Overview of Individual Samples

A number of MOVPE-grown single- and double-layer GaSb–InAs heterostructures with different layer thicknesses were investigated for this thesis. As explained in Sec. 5.7, MOVPE is strongly dependent on growth conditions, while absolute reproducibility is not assured. As a result, there is considerable variation in the properties of nominally identical structures [22].

The DHETs are listed in Table 5.1 with their properties. Electron concentrations and, where available, hole concentrations have been quoted. The electron mobility μ_e can be estimated from the sheet resistivity $\rho_{xx}(0)$ in the absence of a perpendicular magnetic field as $\mu_\mathrm{e} \approx 1/[en_\mathrm{e}\rho_{xx}(0)]$ by setting B_z in Eq. (5.24) to zero and assuming $n_\mathrm{e}\mu_\mathrm{e} \gg n_\mathrm{h}\mu_\mathrm{h}$. (Multiple parameter fits of Eqs. (5.24) and (5.25) to the observed longitudinal and Hall resistivities as a function of perpendicular field indicate that $\mu_\mathrm{e}/\mu_\mathrm{h}$ is typically of the order 10.) As explained in Secs. 5.5.3 and 5.6.1, a more realistic estimated will be obtained from the minimal value of ρ_{xx} in a high parallel magnetic field.

As may be expected, the electronic properties will in general vary slightly between different parts of the same epitaxial structures, and this variation can become quite noticeable for large samples or ones that were grown under less than optimal conditions. No attempt was made to average over different parts of such samples; the values quoted here correspond to a single location deemed representative.

The observed properties of these samples fit the expectation put forward in Sec. 5.5.2 that the electron–hole-ratio is strongly affected by the presence of nearby surface states: The samples of the 0X37xx varying cap thickness series show a decreasing hole sheet density as the cap thickness decreases from approximately 1,200 Å to 800 Å. Even thinner cap samples

Sample	t_{well} (Å)	t_{cap} (Å)	n_{SdH} (10^{11}/cm^2)	n_h (10^{11}/cm^2)	$\rho_{xx}(0)$ (Ω/□)	$\rho_{xx}(B_\parallel)$ (Ω/□)
OX3735	250	600	9·4		220	
OX3729	250	800	7·8†	3·7†	40†	
OX3730	250	1,000	8·0†	5·2†	40†	
OX3733	250	1,200	7·9†	6·1†	56†	
OX4256	(300)	(1,000)	8·0		40	
OX4338	(300)	—	16·0		256	
OX4340	(300)	(1,200)	7·0		186	
OX4434	(300)	(700)	7·9		74	
OX4529	300	1,000	7·7		66	
OX4530	300	700	7·9		112	
OX4531	300	500	11·5		75	
OX4532	300	680	7·1		110*	47*
OX4532′	300	380	12·0		81*	59*
OX4576	300	1,500	6·7		141	124
OX4576′	300	600	7·8		187	

Table 5.1: DHET samples grown by MOVPE. Most InAs well and GaSb cap thicknesses t_{well}, and t_{cap} were estimated from SPA traces and are accurate to about 5 %; values in brackets were deduced from the growth rate and are accurate to about 20 %. The carrier concentration n_{SdH} was calculated from the frequency of the Shubnikov–de Haas-oscillations in the perpendicular field magnetoresistance and corresponds to the electron concentration n_e; it is accurate to about 2×10^{10} cm^{-2} for a specific piece of the sample. Both the zero field value and the minimal value in a parallel magnetic field are given for the sheet resistivity. Most measurements were taken at 4·2 K. Values marked with an asterisk ($*$) were measured at 1·5 K. The data marked with a dagger ($\dagger$) were measured by TAKASHINA [51] at 50 mK; the hole concentration was estimated from the compensated quantum Hall effect.

Sample	t_1 (Å)	t_{sp} (Å)	t_2 (Å)	t_{cap} (Å)	n_{SdH} (10^{11}/ cm^2)	$\rho_{xx}(0)$ (Ω/□)	$\rho_{xx}(B_\parallel)$ (Ω/□)
OX4536	200	50	200	1,500	6·2	118	35
OX4559	200	100	200	1,500	6·2	212	130
OX4538	250	25	250	3,400	6·4	232	178
OX4533	250	50	250	1,500	6·7	71	62
OX4544	250	100	250	2,600	7·6	10	(10)
OX4564	300	50	300	1,500	7·2	124	82
OX4563	300	100	300	1,500	7·3	128	30
OX4574	300	200	300	1,500	6·9	82	21
OX4575	300	300	300	1,500	6·5	82	32
OX4576	—	—	300	1,500	6·7	141	124

Table 5.2: Double well samples grown by MOVPE. The InAs well, GaSb spacer, and GaSb cap thicknesses t_1, t_2, t_{sp}, and t_{cap} were estimated from SPA traces and are accurate to about 5 %. The carrier concentration was calculated from the frequency of the Shubnikov–de Haas-oscillations in the perpendicular field magnetoresistance and is accurate to 2×10^{10} cm^{-2} for a specific piece of the sample. Both the zero field value and the minimal value in a parallel magnetic field are given for the sheet resistivity. Measurements were taken at 4·2 K.

(OX3735, OX4531, OX4532′) also have an increased electron sheet density. The behaviour of OX4532′ and OX4576′, which were created from OX4532 and 4576, respectively, by removing part of the GaSb cap by reactive ion etching (RIE), supports this picture: In both cases the electron sheet density was increased significantly.

Table 5.2 shows the double InAs well samples ordered by well and spacer thickness. These samples are discussed in detail in Chapter 6. Where available, the minimum of the longitudinal sheet resistivity $\rho_{xx}(B_\parallel)$ as a function of the in-plane magnetic field is invariably[9] significantly lower than the magnitude of the resistivity in the absence of a parallel field, substantiating the claims about the presence of a minigap made in Sec. 5.5.3 on the basis of previous work.

[9] With the exception of the double well sample OX4544, which has an unusually low $\rho_{xx}(0)$.

5.9 Summary

The *broken gap* band arrangement of GaSb–InAs heterostructures, in which the InAs conduction band lies lower in energy than the GaSb valence band, leads to charge transfer from the GaSb to the InAs layers, forming intrinsic populations of mobile electrons and holes without the need for doping. For thin InAs wells, quantization of the confinement energy becomes relevant at low temperatures, leading to quasi-two-dimensional behaviour as the carrier wave functions are restricted in the growth direction to a number of states corresponding to discrete subband energies.

Because the electrons and holes in adjacent layers interact, hybrid states appear and a minigap opens in the dispersion relation close to the Fermi level. The effect of this minigap can be seen in the strong negative parallel field magnetoresistance observed in single- and multilayer structures and in radiative transitions. It is not directly observed in experiments probing cyclotron motion because of magnetic breakdown, *i.e.*, tunnelling across the minigap once the cyclotron energy becomes comparable to the interaction energy.

A large number of GaSb–InAs heterostructures were grown in Oxford by MOVPE. The technique allows the creation of high-quality intrinsic samples, but is somewhat unpredictable, necessitating careful characterization of individual specimens. Both Landau quantization in a perpendicular magnetic field—via the analysis of Shubnikov–de Haas-oscillations—and a semiclassical transport theory can be used for this purpose. The accurate determination of hole concentrations and mobilities, however, remains elusive as the presence of the minigap severely limits the applicability of the semiclassical model.

A number of specific growth samples have been introduced in relation to the theoretical treatment; these samples will form the basis for the discussion in the following chapters.

Bibliography

[1] C. Weisbuch and B. Vintner. *Quantum Semiconductor Structures*. Academic Press, San Diego, California, 1991.

[2] N. W. Ashcroft and D. N. Mermin. *Solid State Physics*. Harcourt College Publishers, Fort Worth, Texas, college edition, 1976.

[3] H. Kroemer. The 6·1 Å family (InAs, GaSb, AlSb) and its heterostructures: a selective review. *Physica E*, 20:196–202, 2003.

[4] S. M. Sze. *Physics of Semiconductor Devices*. John Wiley & Sons, New York, second edition, 1981.

[5] C. Kittel. *Introduction to Solid State Physics*. John Wiley & Sons, New York, seventh edition, 1996.

[6] M. L. Cohen and T. K. Bergstresser. Band structures and pseudopotential form factors for fourteen semiconductors of the diamond and zinc-blende structures. *Physical Review*, 141(2):789–796, January 1966.

[7] I. Topol, H. Neumann, and E. Hess. Band structure of AlSb. *Czechoslovak Journal of Physics B*, 24:107–112, 1974.

[8] J. R. Chelikowsky and M. L. Cohen. Nonlocal pseudopotential calculations for the electronic structure of eleven diamond and zinc-blende semiconductors. *Physical Review B*, 14(2):556–582, July 1976.

[9] Z. M. Fang, K. Y. Ma, D. H. Jaw, R. M. Cohen, and G. B. Stringfellow. Photoluminescence of InSb, InAs, and InAsSb grown by organometallic vapor phase epitaxy. *Journal of Applied Physics*, 67(11):7034–7039, June 1990.

[10] M.-C Wu and C.-C. Chen. Photoluminescence of high-quality GaSb grown from Ga- and Sb-rich solutions by liquid-phase epitaxy. *Journal of Applied Physics*, 72(9):4275–4280, November 1992.

[11] I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan. Band parameters for III–V compound semiconductors and their alloys. *Journal of Applied Physics*, 89(11):5815–5875, June 2001.

[12] L. L. Chang and L. Esaki. Electronic properties of InAs–GaSb superlattices. *Surface Science*, 98:70–89, 1980.

[13] G. Tuttle, H. Kroemer, and J. H. English. Electron concentrations and mobilities in AlSb/InAs/AlSb quantum wells. *Journal of Applied Physics*, 65(12):5239–5342, June 1989.

[14] G. Tuttle, H. Kroemer, and J. H. English. Effects of interface layer sequencing on the transport properties of InAs/AlSb quantum wells: Evidence for antisite donors in the InAs/AlSb interface. *Journal of Applied Physics*, 67(6):3032–3037, March 1990.

[15] C. Nguyen, B. Brar, H. Kroemer, and J. H. English. Surface donor contribution to electron sheet concentrations in not-intentionally doped InAs–AlSb quantum wells. *Applied Physics Letters*, 60(15):1854–1856, April 1992.

[16] A. Aardvark, N. J. Mason, and P. J. Walker. The growth of antimonides by MOVPE. *Progress in Crystal Growth and Characterization*, 35(2–4):207–241, 1997.

[17] Y. Chen, D. M. Symons, M. Lakrimi, A. Salesse, G. B. Houston, R. J. Nicholas, N. J. Mason, and P. J. Walker. One dimensional transport and gating of InAs/GaSb structures. *Superlattices and Microstructures*, 15(1):41–45, 1994.

[18] M. S. Daly, K. S. H. Dalton, M. Lakrimi, N. J. Mason, R. J. Nicholas, M. van der Burgt, P. J. Walker, D. K. Maude, and J. C. Portal. Zero-Hall-resistance state in a semimetallic InAs/GaSb superlattice. *Physical Review B*, 53(16):10524–10527, April 1996.

[19] R. J. Nicholas, M. Lakrimi, B. Kardynał, S. Khym, N. J. Mason, J. Rehman, K. Takashina, P. J. Walker, D. M. Symons, D. K. Maude, and J. C. Portal. A digital quantum Hall effect. *Physica E*, 6:836–839, 2000.

[20] A. C. Jones. Metalorganic precursors for vapour phase epitaxy. *Journal of Crystal Growth*, 129:728–773, 1993.

[21] M. Behet, R. Hövel, A. Kohl, A. Mesquida Küsters, B. Opitz, and K. Heime. MOVPE growth of III-V compounds for optoelectronic and electronic applications. *Microelectronics Journal*, 27:297–334, 1996.

[22] R. M. Biefeld. The metal-organic chemical vapor deposition and properties of III–V antimony-based semiconductor materials. *Materials Science and Engineering R*, 36:105–142, 2002.

[23] M. Larkimi, R. W. Martin, N. J. Mason, R. J. Nicholas, and P. J. Walker. GaSb/InAs heterojunctions grown by MOVPE: effect of gas switching sequences on interface quality. *Journal of Crystal Growth*, 110:677–682, 1991.

[24] R. G. Dandrea, C. B. Duke, and A. Zunger. Interfacial atomic structure and band offsets at semiconductor heterojunctions. *Journal of Vacuum Science and Technology B*, 10(4):1744–1753, July/August 1991.

[25] C. Petchsingh. *Cyclotron Resonance Studies on InAs/GaSb Heterostructures*. DPhil thesis, Wolfson College, University of Oxford, 2002.

[26] J. Spitzer, A. Höpner, M. Kuball, M. Cardona, B. Jenichen, H. Neuroth, B. Brar, and H. Kroemer. Influence of the interface composition of InAs/AlSb superlattices on their optical and structural properties. *Journal of Applied Physics*, 77(2):811–820, January 1995.

[27] C. R. Bolognesi, H. Kroemer, and J. H. English. Interface roughness scattering in InAs/AlSb quantum wells. *Applied Physics Letters*, 61(2):213–215, July 1992.

[28] I. Sela, C. R. Bolognesi, L. A. Samoska, and H. Kroemer. Study of interface composition and quality in AlSb/InAs/AlSb quantum wells by Raman scattering from interface modes. *Applied Physics Letters*, 60(26):3283–3285, June 1992.

[29] B. R. Bennet, R. Magno, and N. Papanicolaou. Controlled n-type doping of antimonides and arsenides using GaTe. *Journal of Crystal Growth*, 251:532–537, 2003.

[30] C. R. Bolognesi, M. W. Dvorak, and D. H. Chow. High-transconductance delta-doped InAs/AlSb HFET's with ultrathin silicon-doped InAs quantum well donor layer. *IEEE Electron Device Letters*, 19(3):83–85, March 1998.

[31] C. R. Bolognesi, M. W. Dvorak, and D. H. Chow. InAs/AlSb heterostructure field-effect transistors using a Si-doped InAs/AlSb short-period superlattice modulation doping barrier. *Journal of Vacuum Science and Technology A*, 16(2):843–845, March/April 1998.

[32] C. R. Bolognesi, M. W. Dvorak, and D. H. Chow. Impact ionization suppression by quantum confinement: Effects on the DC and microwave performance of narrow-gap channel InAs/AlSb HFET's. *IEEE Transactions on Electron Devices*, 46(5):826–832, May 1999.

[33] S. J. Koester, C. R. Bolognesi, M. Thomas, E. L. Hu, and H. Kroemer. Determination of one-dimensional subband spacings in InAs/AlSb ballistic constrictions using magentic-field measurements. *Physical Review B*, 50(8):5710–5712, August 1994.

[34] S. J. Koester, B. Brar, C. R. Bolognesi, E. J. Caine, A. Patlach, E. L. Hu, H. Kroemer, and M. J. Rooks. Length dependence of quantized conductance in ballistic constrictions fabricated on InAs/AlSb quantum wells. *Physical Review B*, 53(19):13063–13073, May 1996.

[35] M. Drndic, M. P. Grimshaw, L. J. Cooper, and D. A. Ritchie. Tunable electron-hole gases in gated InAs/GaSb/AlSb systems. *Applied Physics Letters*, 70(4):481–483, January 1997.

[36] T. P. Marlow, L. J. Cooper, D. D. Arnone, N. K. Patel, D. M. Whittaker, E. H. Linfield, D. A. Ritchie, and M. Pepper. Ground state of a two-dimensional coupled electron-hole gas in inas/gasb narrow gap heterostructures. *Physical Review Letters*, 82(11):2362–2365, March 1999.

[37] G. R. Booker, M. Daly, P. C. Klipstein, M. Lakrimi, T. F. Kuech, J. Li S. G. Lyapin, N. J. Mason, I. J. Murgatroyd, J.-C. Portal, R. J. Nicholas, D. M. Symons, P. Vicente, and P. J. Walker. Growth of InAs/GaSb strained layer superlattices by MOVPE: III. Use of UV absorption to monitor alkyl stability in the reactor. *Journal of Crystal Growth*, 170:777–782, 1997.

[38] M. Altarelli. Electronic structure and semiconductor-semimetal transition in InAs–GaSb superlattices. *Physical Review B*, 28(2):842–845, July 1983.

[39] S. Datta, M. R. Melloch, and R. L. Gunshor. Possibility of an excitonic ground state in quantum wells. *Physical Review B*, 32(4):2607–2609, August 1985.

[40] J.-P. Cheng, J. Kono, B. D. McCombe, I. Lo, W. C. Mitchel, and C. E. Stutz. Evidence for a stable excitonic ground state in a spatially separated electron–hole system. *Physical Review Letters*, 74(3):450–453, January 1995.

[41] T. A. Vaughan. *Magneto-optics of InAs/GaSb Heterostructures*. DPhil thesis, Brasenose College, University of Oxford, 1995.

[42] M. Lakrimi, S. Khym, R. J. Nicholas, D. M. Symons, F. M. Peeters, N. J. Mason, and P. J. Walker. Minigaps and novel giant negative magnetoresistance in InAs/GaSb semimetallic superlattices. *Physical Review Letters*, 79(16):3034–3037, October 1997.

[43] M. S. Daly, D. M. Symons, M. Lakrimi, R. J. Nicholas, N. J. Mason, and P. J. Walker. Interface composition dependence of the band offset in InAs/GaSb. *Semiconductor Science and Technology*, 11:823–826, 1996.

[44] M. S. Daly, D. M. Symons, M. Lakrimi, R. J. Nicholas, N. J. Mason, and P. J. Walker. Interface composition dependence of the band offset in InAs/GaSb. *Surface Science*, 361/362:205–208, 1996.

[45] Y. Naveh and B. Laikhtman. Band-structure tailoring by electric field in a weakly coupled electron-hole system. *Applied Physics Letters*, 66(15):1980–1982, April 1995.

[46] G. L. B. Houston, Y. Chen, J. Singleton, N. J. Mason, and P. J. Walker. The interface properties of MIS structures on anodically oxidized GaSb. *Semiconductor Science and Technology*, 12:1140–1146, 1997.

[47] L. J. Cooper, N. K. Patel, V. Drouot, E. H. Linfield, D. A. Ritchie, and M. Pepper. Resistance resonance induced by electron-hole hybridization in a strongly coupled inas/gasb/alsb heterostructure. *Physical Review B*, 57(19):11915–11918, May 1998.

[48] M. J. Yang, F.-C. Wang, C. H. Yang, B. R. Bennett, and T. Q. Do. A composite quantum well field-effect transistor. *Applied Physics Letters*, 69(1):85–87, July 1996.

[49] F. Rahman, B. L. Gallagher, M. Behet, and J. De Boeck. Gating of inas/gasb quantum wells using a silicon monoxide gate insulator. *Applied Physics Letters*, 73(1):88–90, July 1998.

[50] Y. J. Rehman. *Processing and Magneto-transport Studies of InAs/GaSb Low Dimensional Structures*. DPhil thesis, The Queen's College, University of Oxford, 1999.

[51] K. Takashina. *Magneto-transport Studies of GaSb/InAs/GaSb Double-heterostructures*. DPhil thesis, University College, University of Oxford, 2002.

[52] M. Altarelli, J. C. Maan, L. L. Chang, and L. Esaki. Electronic states and quantum Hall effect in GaSb–InAs–GaSb quantum wells. *Physical Review B*, 35(18):9867–9870, June 1987.

[53] M. S. Daly, D. M. Symons, M. Lakrimi, R. J. Nicholas, N. J. Mason, and P. J. Walker. Interface composition dependence of the band offset in InAs/GaSb. *Semiconductor Science and Technology*, 11:823–826, 1996.

[54] M. Lakrimi, S. Khym, D. M. Symons, R. J. Nicholas, F. M. Peeters, N. J. Mason, and P. J. Walker. Mini-gaps and novel giant negative magnetoresistance in InAs/GaSb semimetallic superlattices. *Physica E*, 2:363–367, 1998.

[55] *Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology: New Series: Group III: Crystal and Solid State Physics*, volume 17. Springer, 1982.

[56] K. K. Choi, B. F. Levine, N. Jarosik, J. Walker, and R. Malik. Anisotropic magnetotransport in weakly coupled GaAs–$Al_xGa_{1-x}As$ multiple quantum wells. *Physical Review B*, 38(17):12362–12368, 1988.

[57] F. Stern. Transverse Hall effect in the electric quantum limit. *Physical Review Letters*, 21(25):1687–1690, December 1968.

[58] A. J. L. Poulter, M. Lakrimi, R. J. Nicholas, N. J. Mason, and P. J. Walker. Optical probing of the minigap in InAs/GaSb semimetallic superlattices. *Physica B*, 256–258:256–259, 1998.

[59] W. Beinvogl, A. Kamgar, and J. F. Koch. Influence of a magnetic field on electron subbands in a surface-charge layer. *Physical Review B*, 14(10):4274–4289, November 1976.

[60] M. J. Yang, C. H. Yang, B. R. Bennett, and B. V. Shanabrook. Evidence of a hybridization gap in "semimetallic" InAs/GaSb systems. *Physical Review Letters*, 78(24):4613–4616, June 1997.

[61] D. M. Symons, M. Lakrimi, R. J. Nicholas, D. K. Maude, J. C. Portal, N. J. Mason, and P. J. Walker. Magnetic breakdown in the semimetallic InAs/GaSb system. *Physical Review B*, 58(11):7292–7299, September 1998.

[62] T. Ando, A. B. Fowler, and F. Stern. Electronic properties of two-dimensional systems. *Reviews of Modern Physics*, 54(2):437–672, April 1982.

[63] H. M. Manasevit and W. I. Simpson. Use of metal-organics in preparation of semiconductor materials. I. Epitaxial gallium-V compounds. *Journal of The Electrochemical Society*, 116:1725–1732, 1969.

[64] S. V. Ivanov, P. D. Altukhov, T. S. Argunova, A. A. Bakun, A. A. Budza, V. V. Chaldyshev, Yu. A. Kovalenko, P. S. Kop'ev, R. N. Kutt, B. Ya. Meltser, S. S. Ruvimov, S. V. Shaposhnikov, L. M. Sorokin., and V. M. Ustinov. Molecular beam epitaxy growth and characterization of thin (< 2 µm) GaSb layers on GaAs(100) substrates. *Semiconductor Science and Technology*, 8:347–356, 1993.

[65] W. Qian, M. Skowronski, R. Kaspi, M. De Graef, and V. P. David. Nucleation of misfit and threading dislocations during epitaxial growth of GaSb on GaAs(001) substrates. *Journal of Applied Physics*, 81(11):7268–7272, June 1997.

[66] P. M. Thibado, B. R. Bennett, M. E. Twigg, B. V. Shanabrook, and L. J. Whitman. Evolution of GaSb epitaxy on GaAs(001)-c(4×4). *Journal of Vacuum Science and Technology A*, 14(3):885–890, May/June 1996.

[67] C. W. Bumby. *Thermophotovoltaic Devices based upon Heteroepitaxial GaSb grown by MOVPE*. DPhil thesis, Hertford College, University of Oxford, 2005.

[68] D. A. Allwood, N. J. Mason, and P. J. Walker. In situ characterisation of MOVPE by surface photoabsorption: I. Substrate oxide desorption. *Journal of Crystal Growth*, 195:163–167, 1998.

[69] P. C. Klipstein, S. G. Lyapin, N. J. Mason, and P. J. Walker. In situ characterisation of MOVPE by surface photoabsorption: II. Interface monitoring. *Journal of Crystal Growth*, 195:168–173, 1998.

6 Double Well Samples

Double, double toil and trouble;
Fire burn and cauldron bubble.

(William Shakespeare, Macbeth)

6.1 Introduction

Prior research on InAs–GaSb heterostructures in Oxford has considered single InAs wells embedded between GaSb substrate and cap layers [1–4] as well as superlattices consisting of many alternating InAs and GaSb layers [5–8]. For thick GaSb spacers, the behaviour of superlattices follows that of the double heterointerfaces (DHETS). On the other hand, if the GaSb layer thickness is small enough to allow for significant overlap between the electron wave functions centred at different InAs wells, there is a finite dispersion along the growth direction of the superlattice and the energy levels broaden into a series of 'mini-bands' with band widths approximately antiproportional to the layer spacing.

To study the interaction between electron states in different InAs layers more closely, Philip Shields and I prepared a series of samples following the pattern of Fig. 6.1. These samples, which are listed in Table 6.1, contained two InAs wells with thicknesses $t_1 = t_2 = t_w$ separated by a GaSb layer of thickness t_b. Both t_w and t_b were varied systematically. The cap thickness t_{cap} of at least 1,500 Å for all samples ensured that effects of the asymmetric band bending caused by the pinning of the Fermi level at the free GaSb surface were minimized. All of these heterostructures were grown in the metal-organic vapour phase epitaxy (MOVPE) reactor in Oxford within a relatively short time under similar conditions. None the less, the

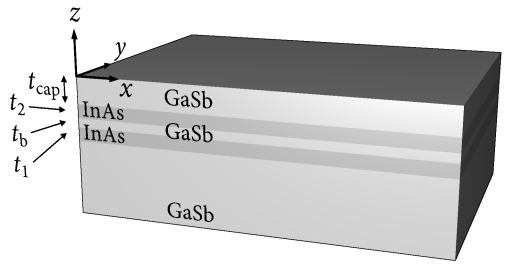


Figure 6.1: Structure of double well samples

t_b/Å	$t_w = 200$ Å	$t_w = 250$ Å	$t_w = 300$ Å
25		OX4538	
50	OX4536	OX4533	OX4564
100	OX4559	OX4544	OX4563
200			OX4574
300			OX4575

Table 6.1: Double well samples by layer dimensions

reproducibility of the epitaxy technique is not perfect and it is only approximately true that the layer thickness was the only parameter that varied between the samples. In particular, there was a measurable difference in surface roughness between different growth runs and between different areas of the same sample.

6.2 Calculation of Wave Functions from $\mathbf{k} \cdot \hat{\mathbf{p}}$ Perturbation Theory

6.2.1 Principle

The $\mathbf{k} \cdot \hat{\mathbf{p}}$ formalism is an established approximation technique based on second order perturbation theory. It allows the calculation of the energy eigenstates of a solid in a limited region of the momentum space as a linear combination of the eigenstates at a high symmetry point within the region [9,10].

In the presence of a periodic potential $V(\mathbf{r}) = V(\mathbf{r}+\mathbf{R})$, where $\mathbf{R}$ is a lattice vector, the wave function $\varphi(\mathbf{r})$ of an electron obeys the time-independent Schrödinger equation

$$\left[\frac{\hat{\mathbf{p}}^2}{2m_\mathrm{e}} + V(\mathbf{r})\right]\varphi(\mathbf{r}) = E\varphi(\mathbf{r}) \tag{6.1}$$

with $\hat{\mathbf{p}} = -i\hbar\,\nabla_{\mathbf{r}}$. Using the Bloch theorem, $\varphi_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r})\exp(i\mathbf{k}\cdot\mathbf{r})$, where $\mathbf{k}$ is the crystal momentum, Eq. (6.1) can be expressed in terms of the Bloch functions $u_{n\mathbf{k}}(\mathbf{r}) = u_{n\mathbf{k}}(\mathbf{r}+\mathbf{R})$ as

$$\left[\frac{\hat{\mathbf{p}}^2}{2m_\mathrm{e}} + \frac{\hbar}{m_\mathrm{e}}\mathbf{k}\cdot\hat{\mathbf{p}} + \frac{\hbar^2\mathbf{k}^2}{2m_\mathrm{e}} + V(\mathbf{r})\right]u_{n\mathbf{k}}(\mathbf{r}) = E_n(\mathbf{k})u_{n\mathbf{k}}(\mathbf{r}). \tag{6.2}$$

The eponymous $\mathbf{k}\cdot\hat{\mathbf{p}}$ term is then treated as a perturbation to the Hamiltonian at $\mathbf{k}_0 = \mathbf{o}$, and to lowest non-vanishing order in $|\mathbf{k}|$, the approximated Bloch wave functions and energies become

$$|u_{n\mathbf{k}}\rangle = |u_{n\mathbf{0}}\rangle + \sum_{m\neq n}\frac{\langle u_{n\mathbf{0}}|\mathbf{k}\cdot\hat{\mathbf{p}}|u_{m\mathbf{0}}\rangle}{E_n(\mathbf{o}) - E_m(\mathbf{o})}|u_{m\mathbf{0}}\rangle \qquad \text{and} \tag{6.3}$$

$$E_n(\mathbf{k}) = E_n(\mathbf{o}) + \frac{\hbar^2\mathbf{k}^2}{2m_\mathrm{e}} + \frac{\hbar^2}{m_\mathrm{e}^2}\sum_{m\neq n}\frac{\left|\langle u_{n\mathbf{0}}|\mathbf{k}\cdot\hat{\mathbf{p}}|u_{m\mathbf{0}}\rangle\right|^2}{E_n(\mathbf{o}) - E_m(\mathbf{o})}. \tag{6.4}$$

The choice $\mathbf{k}_0 = \mathbf{o}$ is appropriate for III–V materials because the extrema of both the valence and the conduction bands lie at the Γ-point (*cf.* Fig. 5.2(a)), but the method can be generalized straightforwardly to other $\mathbf{k}_0$.

6.2.2 Application to InAs–GaSb Heterostructures in a Magnetic Field

I used a programme originally developed at Oxford by Thomas Vaughan for calculating the band structure of infinite GaSb–InAs superlattices close to the Γ-point. The software has successfully been used to predict a wide range of experimental superlattice results [5, 11] and has also been applied to single-well systems by increasing the distance between InAs layers to the extent that the overlap of wave functions at different wells becomes negligible or by introducing virtual large bandgap spacer layers [4, 12]. As the implementation details are discussed extensively in Vaughan's thesis [11] and the references therein, I shall restrict myself to summarizing the crucial points of the computation.

A realistic calculation of the energy levels at high symmetry points in III–V semiconductors entails taking spin–orbit coupling into account. This leads to an additional term of the form $\hbar/(2m_e c)^2 \left[\nabla_{\mathbf{r}} V(\mathbf{r}) \times \hat{\mathbf{p}}\right] \cdot \boldsymbol{\sigma}$ in Eq. (6.1), where $\boldsymbol{\sigma}$ is a vector containing the Pauli spin matrices as elements. If an external magnetic field is present, $\hat{\mathbf{p}}$ becomes $\hat{\mathbf{p}} + e\mathbf{A}(\mathbf{r})$, where $\mathbf{A}(\mathbf{r})$ is the magnetic vector potential, and there is a further electron spin contribution $\frac{1}{2} g \mu_B \mathbf{B} \cdot \boldsymbol{\sigma}$ to the Hamiltonian, where $g \approx 2$ is the Landé g-factor and $\mathbf{B}$ is the magnetic flux density.[1] The spin–orbit and spin splitting terms can be treated perturbatively. If the coordinate system is chosen with one axis along the direction of the magnetic field, the Schrödinger equation becomes separable as outlined in Sec. 5.6.2.

Instead of expressing the wave function in the infinite basis formed by the Bloch waves at $\mathbf{k} = \mathbf{0}$, Löwdin's orthogonalization method [13] was used with a limited set of eight basis states. Following Kane [14], the basis states were chosen as linear combinations of atomic orbitals that corresponded to the conduction band, the light and heavy hole bands, and the split-off (spin-orbit) hole band (with two spin states each).

To account for the effect of the heterojunctions, the expansion coefficients in the chosen basis were allowed to vary as a function of position, so that the wave function was effectively a superposition of atomic orbitals modulated by an envelope function; it was implicitly assumed that the basis states are the same for both materials. These simplifications, known as the envelope function approximation (EFA), are valid for most material systems as long as the envelope function varies slowly on the scale of the interatomic distance [11]. The envelope function was calculated in momentum space using an approach analogous to the transfer matrix technique. As explained in Sec 5.5.2, the negative band offset causes a net charge transfer from the GaSb layers to the InAs layers if the InAs wells are sufficiently wide. The programme performed several iterations until it arrived at an effective potential that was consistent with the calculated envelope functions. In a heterostructure, the simple treatment of the magnetic field was restricted to field directions parallel to the growth direction (and, at the same time, parallel to a direction of high symmetry of the crystal lattice).

[1] This is the term that leads to Zeeman splitting of the energy levels.

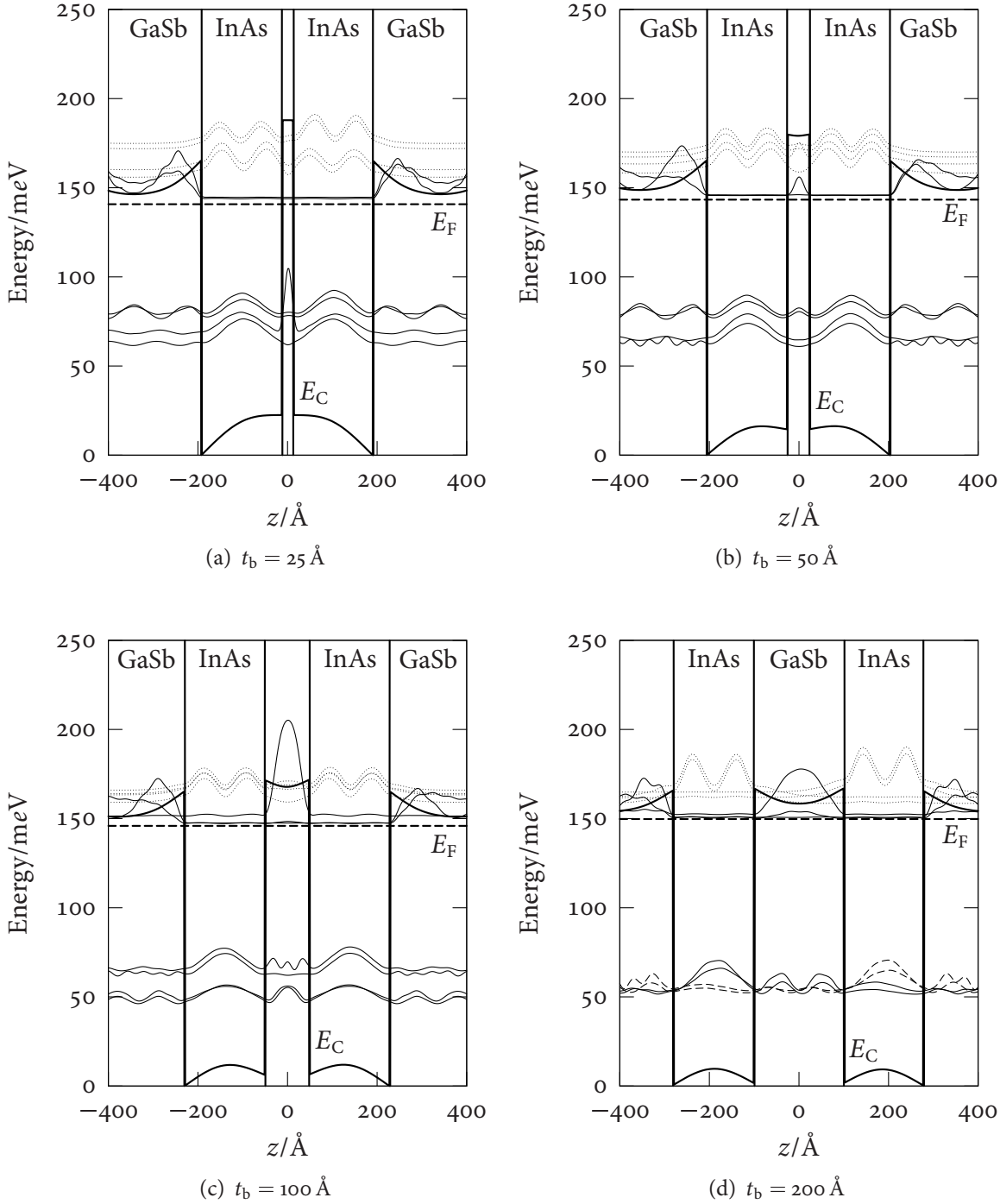
6.2.3 Double Well Calculations

Modelling DHETs and double well samples using this approach poses additional challenges. A virtual superlattice[2] must be constructed in such a way that the interaction between different periods is effectively minimized. This can be achieved by inserting either sufficiently thick bulk-like GaSb layers or additional layers of a large bandgap material such as AlSb. The addition of such fictitious layers introduces further parameters that can be adjusted to match experimental data and has successfully been used to predict Landau level diagrams of thin cap (asymmetric) DHETs that are in good agreement with observation [4, 12]. In double well samples, however, such an additional layer would either reduce the interaction between the layers or introduce a strong asymmetry between them.

The heterostructures were represented by a symmetrical model in these preliminary calculations. In real samples, this symmetry is broken by trapped charges at either the sample surface or the heterojunctions. In the structures discussed here, there was a thick GaSb cap layer and the effect of surface states could be assumed to be small.[3] The charge and the band offset at the interfaces depend on the growth conditions; previous studies have shown that the band offset is larger for InSb-like interfaces than for GaAs-like ones [12]. In our case, the interfaces were nominally unbiased; none the less the possibility of an imbalance between GaSb–InAs and InAs–GaSb interfaces could not be excluded.

For thick layers, the calculation systematically underestimated the confinement energies of both electrons and holes. Whereas experimental data has consistently shown no occupation of subbands other than the ground subband [4] for InAs wells as thick as 340 Å, the $\mathbf{k}\cdot\hat{\mathbf{p}}$ calculation for a symmetric DHET predicted occupation of the second subband at around 200 Å; a similar issue was encountered for holes in thick bulk-like GaSb layers. I calculated energy spectra for double well structures with different barrier thicknesses t_b using the $\mathbf{k}\cdot\hat{\mathbf{p}}$ method. One such calculation is shown in Fig. 6.2 for an InAs well thickness $t_w = 180$ Å; this is smaller than the well thickness in any of the experimental samples but correctly rep-

[2] Which is essential for the momentum matrix technique [11].
[3] See also the discussion in Chapter 7.



(a) $t_b = 25$ Å (b) $t_b = 50$ Å

(c) $t_b = 100$ Å (d) $t_b = 200$ Å

Figure 6.2: Electron and hole states in double well structures, calculated for a conduction band offset of 140 meV, $t_w = 180$ Å, and $B_z = 1$ T. For conduction band states below and valence band states above the Fermi level, the probability densities are shown in arbitrary units, vertically offset by the energy of the state. Dotted lines represent the unoccupied second electron subband.

resents the situation in which only one electron subband is occupied. For small barrier thicknesses, the InAs conduction band states in the two wells mixed. On increasing the barrier width from 50 to 100 Å, hole levels appeared in the barrier, which approached those in the bulk GaSb on either side of the double well structures if the barrier was made even wider. Increased localization of conduction band states to either InAs well was predicted for large t_b.

6.3 Experimental Results

6.3.1 Overview of Individual Samples

To perform magnetotransport measurements on the double well samples, I created a number of Hall bars with a width of 20 µm and an aspect ratio of 2 : 1 by means of optical lithography. I then measured the sheet resistivity components in magnetic fields up to 18 T both in a perpendicular field configuration with $B = B_z$ pointing in the growth direction and in a parallel field configuration with $B = B_\parallel$ lying in the plane of the carriers along the $\langle 100 \rangle$ family of crystallographic directions, either parallel or perpendicular to the excitation current $I_x(t)$. All data were taken at 0·5, 1·5, or 4·2 K according to the procedures laid down in Chapter 4.

Table 6.2 shows an overview of these measurements and displays a number of overall sample properties. Estimates of the sheet carrier concentration were obtained from the frequency of the low field Shubnikov-de Haas-oscillations (n_{SdH}) as well as the Hall coefficient around $B_z = 0$ (n_{R_H}); in all samples the latter value was significantly higher than the the former, with ratios $n_{R_\mathrm{H}} : n_{\mathrm{SdH}} > 2$ for samples with large barrier thicknesses t_b. Almost all samples showed a very strong negative magnetoresistance in a parallel field $B_\parallel$. The resistivity typically saturated for $B_\parallel > 12\,\mathrm{T}$; for many samples, $\rho_{\mu\mu}$ dropped to less than a fourth of its zero field value.

Fitting the low field ($B_z < 2\,\mathrm{T}$) data for the large t_b sample 0X4575 to the semiclassical Drude model of Sec. 5.6.1 indicated similar electron and hole concentrations, but did not

(a) $t_w = 200$ Å

t_b (Å)	Sample	T (K)	n_{SdH} (10^{11}/cm^2)	n_{R_H} (10^{11}/cm^2)	$\frac{n_{R_H}}{n_{SdH}}$	$\frac{\rho_{\mu\mu}(0)}{\rho_{\mu\mu}^{(min)}(B_\parallel)}$	$\frac{\rho_{\mu\mu}^{(max)}(B_\parallel)}{\rho_{\mu\mu}^{(min)}(B_\parallel)}$
50	A	4·2	6·2	10·6	1·7	3·4	4·6
	B	0·5	5·9	12·0	2·0	2·5	3·3
	C	0·5	6·4	10·5	1·6	4·5	4·7
100	A	4·2	6·0	9·8	1·6	1·6	1·9
	B	0·5	6·1	9·2	1·5		

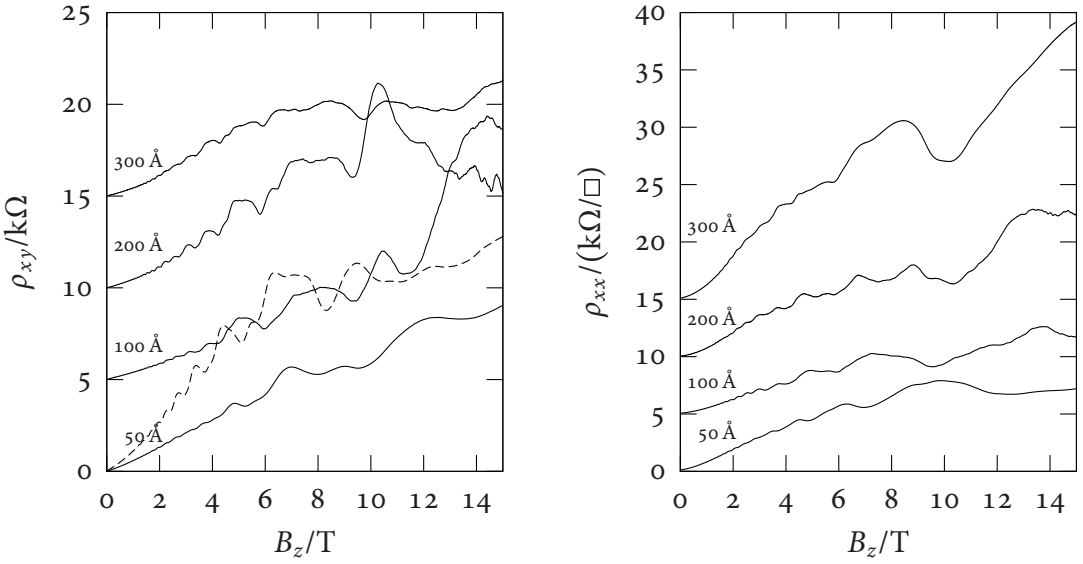
(b) $t_w = 250$ Å

t_b (Å)	Sample	T (K)	n_{SdH} (10^{11}/cm^2)	n_{R_H} (10^{11}/cm^2)	$\frac{n_{R_H}}{n_{SdH}}$	$\frac{\rho_{\mu\mu}(0)}{\rho_{\mu\mu}^{(min)}(B_\parallel)}$	$\frac{\rho_{\mu\mu}^{(max)}(B_\parallel)}{\rho_{\mu\mu}^{(min)}(B_\parallel)}$
25	A	4·2	6·0	9·4	1·6		
	B	4·2	6·1	8·7	1·4	1·3	1·6
	B	1·5	6·9	8·8	1·3	1·3	1·6
	C	0·5	6·2	9·4	1·5	3·6	4·5
50	A	4·2	6·7	9·8	1·5	1·1	1·8
	B	0·5	7·0	12·3	1·8	4·0	5·6
100	A	4·2	7·0	17·8	2·8	3·6	6·2
	B	0·5	7·6	18·4	2·4	5·5	7·1

(c) $t_w = 300$ Å

t_b (Å)	Sample	T (K)	n_{SdH} (10^{11}/cm^2)	n_{R_H} (10^{11}/cm^2)	$\frac{n_{R_H}}{n_{SdH}}$	$\frac{\rho_{\mu\mu}(0)}{\rho_{\mu\mu}^{(min)}(B_\parallel)}$	$\frac{\rho_{\mu\mu}^{(max)}(B_\parallel)}{\rho_{\mu\mu}^{(min)}(B_\parallel)}$
50	A	4·2	6·9	11·4	1·7	5·1	5·5
	B	4·2	7·5	11·5	1·5	1·5	2·0
	B	1·5	7·0	10·7	1·5	1·6	2·1
100	A	4·2	7·3	15·9	2·2	4·3	8·3
	B	0·5	7·7	16·9	2·2	3·2	4·3
	C	0·5	7·6	15·9	2·1	4·0	6·2
200	A	4·2	6·9	16·0	2·3	3·9	4·1
	B	0·5	7·4	16·3	2·2	3·7	3·9
	C	0·5	7·3	16·4	2·2	3·7	4·1
300	A	4·2	6·5	16·2	2·5	2·6	3·8
	B	0·5	7·4	16·6	2·2	3·1	3·2
	C	0·5	7·4	16·3	2·2	3·1	3·2

Table 6.2: Overview of individual samples. The concentrations n_{SdH} and n_{R_H} were calculated from the frequency of the Shubnikov-de Haas-oscillations in a low (0·4 to 1·2 T) field and from the Hall coefficient $R_H(0)$ around the origin ($B_z < 0{\cdot}5$ T).



(a) Hall resistivity (b) Magnetoresistivity

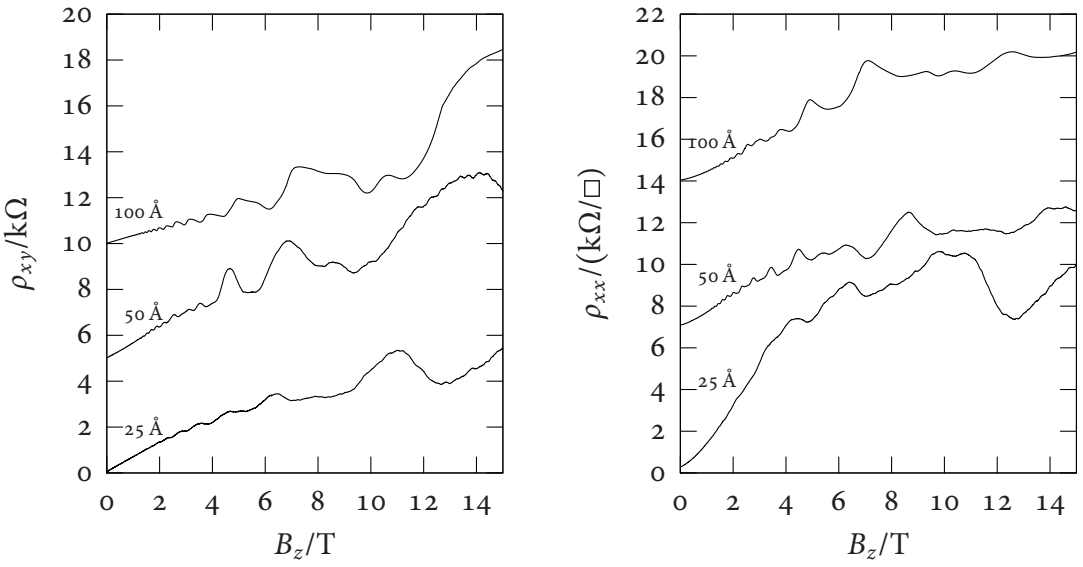
Figure 6.3: Hall resistivity and magnetoresistivity of $t_w = 300\,\text{nm}$ samples for $t_b = 50\,\text{Å}$ (bottom), 100 Å, 200 Å, and 300 Å (top). Individual curves are offset by 5 kΩ for clarity. All samples were measured at $T = 0{\cdot}5\,\text{K}$, except 0x4564 ($t_b = 50\,\text{Å}$), for which $T = 1{\cdot}5\,\text{K}$. The dashed line in (a) represents the $t_w = 300\,\text{Å}$ DHET 0x4576.

yield stable estimates for the absolute concentrations and mobilities. The resistivity is dominated by the minigap formed as a result of electron–hole hybridization (*cf.* Sec 5.5.3). Using the magnetoresistance at large $B_\parallel > 12\,\text{T}$ together with $n_e = 2n_{SdH}$ to remove the effect of the minigap, an electron mobility of 130,000 $\text{cm}^2/(\text{Vs})$ for 0x4575 and 230,000 $\text{cm}^2/(\text{Vs})$ for 0x4574 was estimated.

6.3.2 Magnetoresistance Data

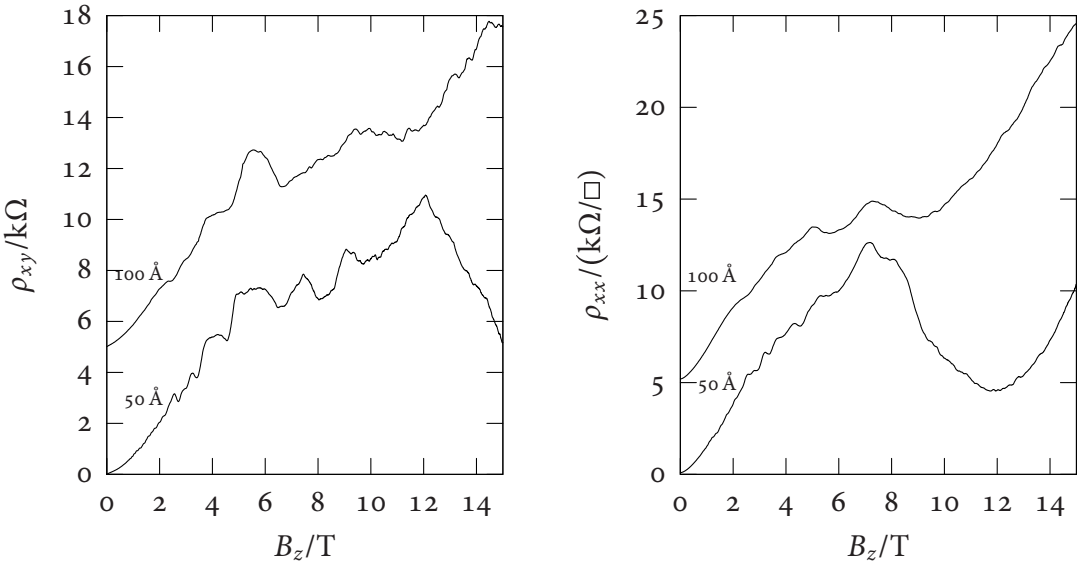
Perpendicular Magnetic Field

For large barrier thicknesses $t_b > 100\,\text{Å}$, the behaviour of the magnetoresistivity did not depend strongly on t_b. This is illustrated in Fig. 6.3, which contrasts samples with different t_b but identical well width $t_w = 300\,\text{Å}$. The intermediate magnetic field features were similar for samples with $t_b = 100\,\text{Å}$, 200 Å, and 300 Å and occurred at similar field values. While



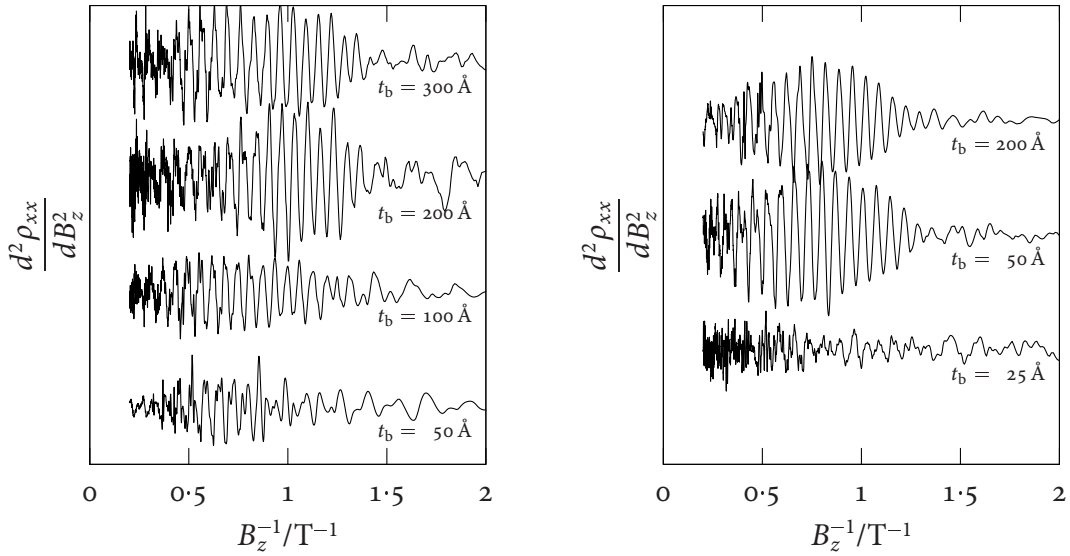
(a) Hall resistivity (b) Magnetoresistivity

Figure 6.4: Hall resistivity and magnetoresistivity of $t_w = 250$ nm samples for $t_b = 25$ Å (bottom), 50 Å, and 100 Å (top). Individual curves are offset by 5 kΩ in (a) and 7 kΩ in (b). All samples were measured at $T = 0{\cdot}5$ K, except 0x4544 ($t_b = 100$ Å), for which $T = 4{\cdot}2$ K.



(a) Hall resistivity (b) Magnetoresistivity

Figure 6.5: Hall resistivity and magnetoresistivity of $t_w = 200$ nm samples for $t_b = 50$ Å (bottom) and 100 Å (top). Individual curves are offset by 5 kΩ for clarity. Both samples were measured at $T = 0{\cdot}5$ K.



(a) $t_w = 300$ Å (b) $t_w = 250$ Å

Figure 6.6: Shubnikov–de Haas-oscillations in double well samples. The second derivative of $\rho_{xx}(B_z)$ taken from Figs. 6.3 and 6.4 is plotted against $1/B_z$. Individual curves are offset for clarity.

there are still noticeable differences between these curves, they are not significant in the sense that different pieces of the same sample showed similar disparities. Compared to the Hall resistivity of the single DHET 0X4576 with $t_w = 300$ Å and $t_{cap} = 1{,}500$ Å, overlaid on Fig. 6.3(a), the double well samples exhibited analogous features shifted to (approximately 10 %) higher field. The magnitude of ρ_{xy} was halved.

On decreasing the barrier thickness further from 100 Å, the situation became more complicated, as may be seen from the $t_b = 50$ Å trace in Fig. 6.3 and also from Figs. 6.4 and 6.5, which show similar comparisons for samples with $t_w = 250$ Å and 200 Å. The magnetoresistivity features changed qualitatively and could no longer straightforwardly be connected to comparable features observed for samples with larger t_b.

Fig. 6.6 shows the second derivative of $\rho_{xx}(B_z)$ plotted against $1/B_z$ and allows a closer look at the Shubnikov–de Haas-oscillations in these samples. The oscillations exhibited a

single dominating frequency. Splitting of individual peaks was seen in some traces, and a single sample ($t_w = 300$ Å, $t_b = 50$ Å) showed indications of a distinct second oscillation frequency, corresponding to $n_{SdH} = 3{\cdot}2 \pm 0{\cdot}2\,\mathrm{cm}^{-2}$. However, there was no compelling evidence for the beating pattern that would be associated with two incommensurate oscillation frequencies with comparable values.

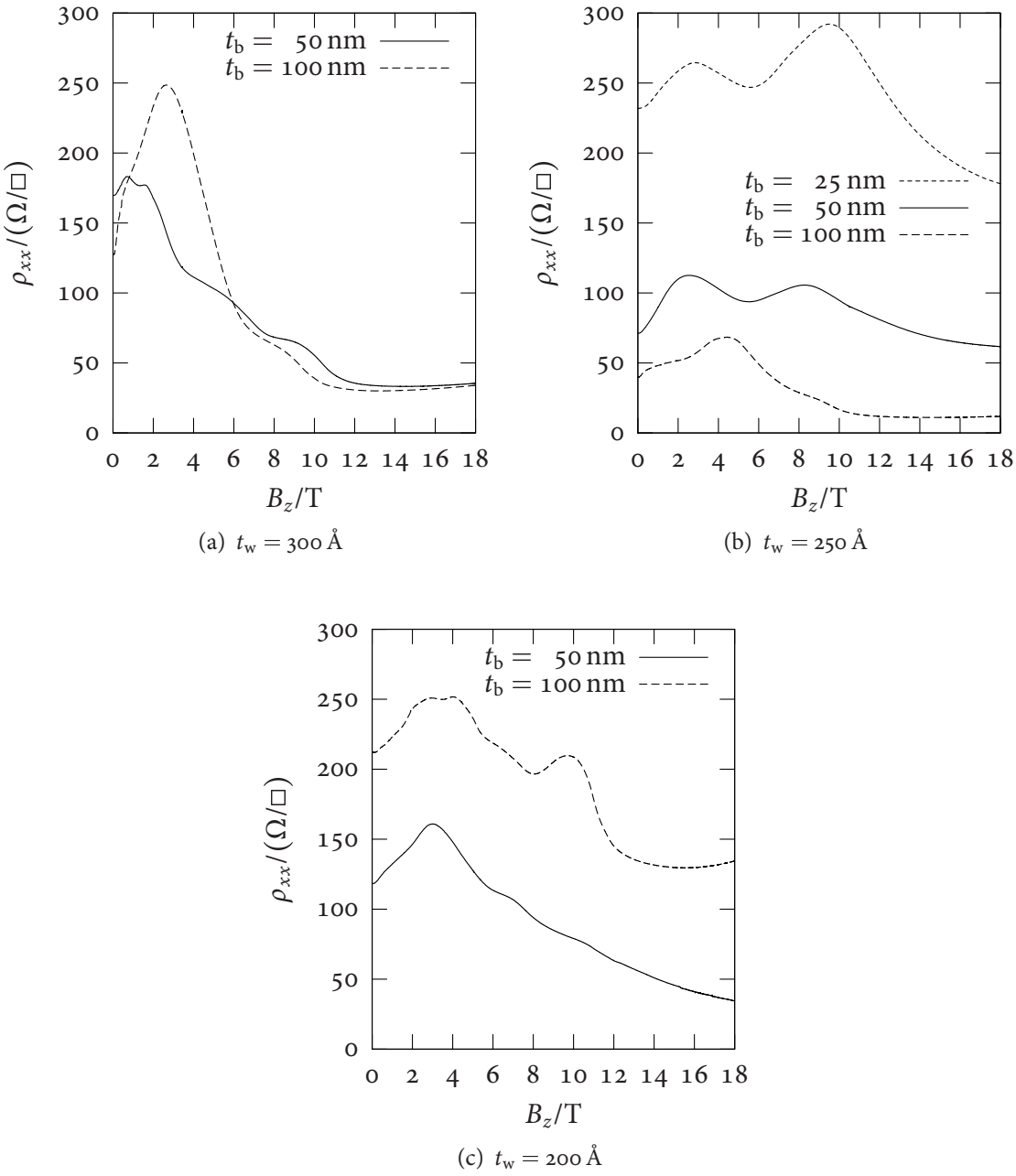
Parallel Field

There was considerable variation in the parallel field longitudinal resistivity $\rho_{\mu\mu}(B_\parallel)$ of the double well samples, shown in Fig. 6.7 for an excitation current parallel to the in-plane field. For all of these samples, $\rho_{xx}(B_\parallel)$ reached a maximum at a low field ($B_\parallel < 4$ T). A further pronounced feature occurred around 10 T. The magnetoresistance either exhibited a shoulder and dropped to a saturation value for $B_\parallel > 12$ T or it reached a second maximum—which could exceed the low field one—and continued to decrease over the investigated field region. The magnetoresistance in the direction perpendicular to the current, $\rho_{yy}(B_\parallel)$, showed similar behaviour in the limit of low and high fields and less pronounced maxima (not shown); direct comparison is problematic as these data were taken on different areas of the samples.

The thick barrier samples with $t_b = 200$ Å or 300 Å and $t_w = 300$ Å behaved consistently and always had a maximum at or close to $B_\parallel = 0$ T and a saturated resistance of a quarter to a third of the zero field value at high parallel fields.

Quantum Well Width

The comparison of the perpendicular field magnetoresistance traces for different well thicknesses t_w in Fig. 6.8 showed similar structure, which was shifted to lower magnetic field values for larger well widths. No clear trend emerged in the parallel field traces, which are collated in Fig. 6.7. In so far as features were directly comparable, they were shifted to higher fields for samples with thinner InAs wells.

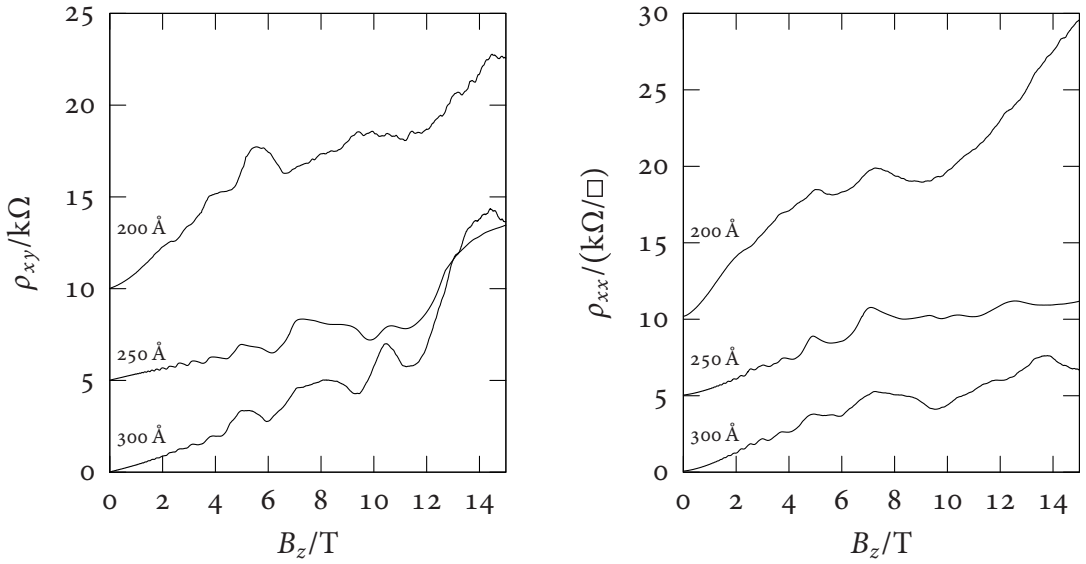


(a) $t_w = 300\,Å$

(b) $t_w = 250\,Å$

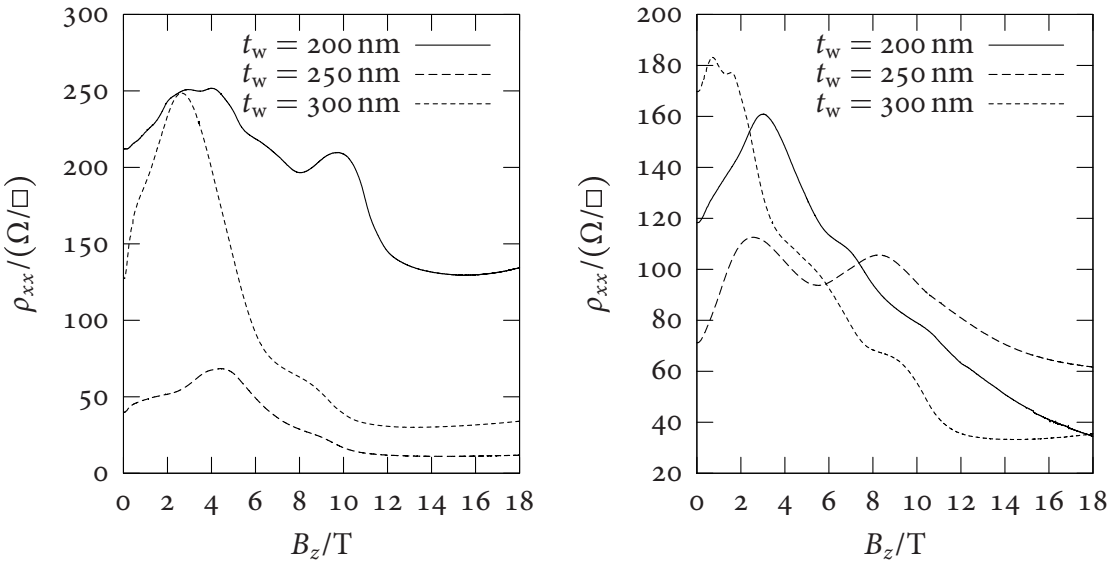
(c) $t_w = 200\,Å$

Figure 6.7: Parallel field magnetoresistance of double well samples, measured with the magnetic field parallel to the excitation current. All samples were measured at $T = 4{\cdot}2\,K$.



(a) Hall resistivity (b) Magnetoresistivity

Figure 6.8: Hall resistivity and magnetoresistivity of $t_b = 100$ nm samples for $t_w = 300$ Å (bottom), 250 Å and 200 Å (top). Individual curves are offset by 5 kΩ in (a) and 7 kΩ in (b). All samples were measured at $T = 0{\cdot}5$ K, except 0x4544 ($t_b = 100$ Å), for which $T = 4{\cdot}2$ K.



(a) $t_b = 100$ Å (b) $t_b = 50$ Å

Figure 6.9: Parallel field magnetoresistance of samples with different well widths, measured with the magnetic field parallel to the excitation current. All samples were measured at $T = 4{\cdot}2$ K.

6.4 Discussion

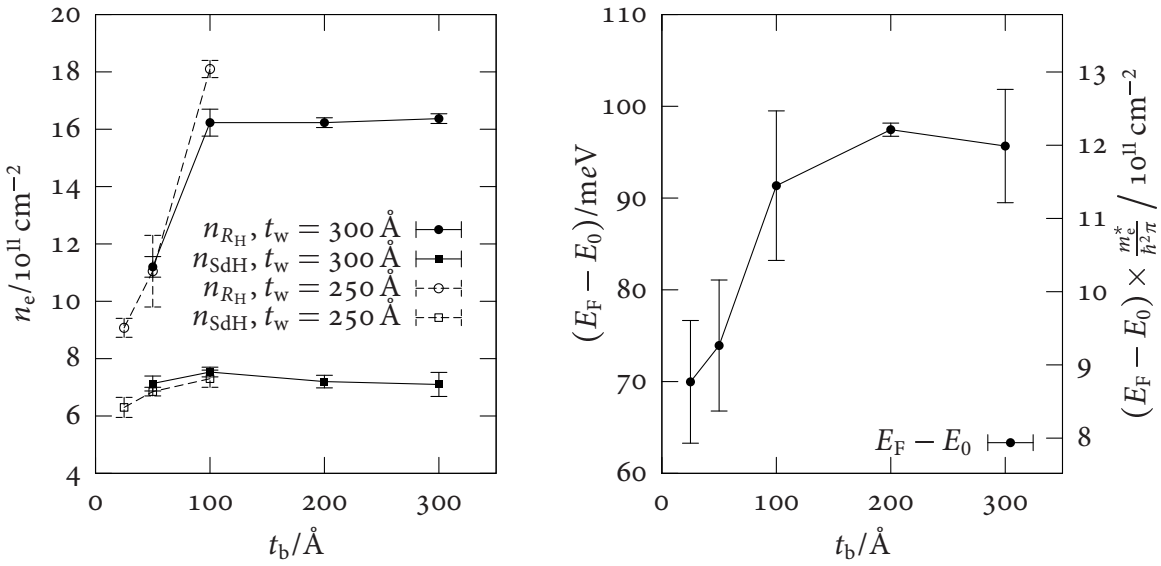
6.4.1 Carrier Concentration

In the limit of small magnetic fields, the semiclassical expression for the Hall resistance of a two dimensional gas of electrons and holes, Eq. (5.25), simplifies to

$$\rho_{xy} \approx \frac{B_z}{e} \frac{n_\mathrm{e}\mu_\mathrm{e}^2 - n_\mathrm{h}\mu_\mathrm{h}^2}{(n_\mathrm{e}\mu_\mathrm{e} + n_\mathrm{h}\mu_\mathrm{h})^2}. \tag{6.5}$$

The Hall coefficient, $R_\mathrm{H} = -\rho_{xy}/B_z$, is therefore dominated by the highest mobility carrier species and the slope of $\rho_{xy}(B_z)$ close to $B_z = 0\,\mathrm{T}$ can be used to estimate the *total* density n_{R_H} of electrons participating in charge transport. On the other hand, the concentration n_SdH estimated from the low field Shubnikov–de Haas-oscillations using Eq. (5.32) depends ultimately on the assumed Landau level degeneracy of Eq. (5.30). The observation that $n_{R_\mathrm{H}} \approx 2n_\mathrm{SdH}$ for all samples with barrier thicknesses t_b of 100 Å and larger is then explained by a doubling of the Landau level degeneracy, accounting for conduction band states in each individual InAs well. For thin barrier ($t_\mathrm{b} \leq 50\,\text{Å}$) structures, most measurements imply $n_{R_\mathrm{H}} < 2n_\mathrm{SdH}$. The origin of this systematic deviation is not clear. However, an exact agreement between n_{R_H} and n_SdH is not expected even for single DHETs as the semiclassical approximation is of limited validity in systems of strongly interacting electrons and holes.

The average electron sheet densities calculated from the experimental data of Table 6.2 are plotted in Fig. 6.10(a) against the barrier thickness t_b. It becomes apparent that the electron concentration increases significantly between $t_\mathrm{b} = 25\,\text{Å}$ and $t_\mathrm{b} = 100\,\text{Å}$ but stays approximately constant for larger barrier thicknesses. The behaviour indicates that for thin barriers the interaction between the wells strongly affects the electron energy levels, whereas the effect is small for $t_\mathrm{b} > 100\,\text{Å}$. The $\mathbf{k}\cdot\hat{\mathbf{p}}$ calculations of Sec. 6.2.3 confirm this picture. In a 2DEG, the density of states is constant and $n_\mathrm{e} = (E_\mathrm{F} - E_0) \times m^*/\hbar^2\pi$. Fig. 6.10(b) shows the bottom of the lowest electron subband E_0 relative to the Fermi energy E_F versus the well distance; the calculation sucessfully reproduces the trend of Fig. 6.10(a). However, the absolute electron densities are overestimated even though the model assumes a significantly lower well width



(a) Experimental densities from Table 6.2 (b) Calculated $E_F - E_0$ from Sec. 6.2.3

Figure 6.10: Double well electron density versus spacer thickness.

of 180 Å. The measurements of Table 6.1 also show a weak but consistent decrease in n_{SdH} with decreasing InAs well width; the average absolute values are compatible with the values reported for similar DHETS [4, 12].

As the cap thickness of the samples investigated here is well within the regime for which previous studies have established a negligible influence of the Fermi level pinning at the surface [4], any imbalance between the concentrations of mobile holes and electrons must be a result of charge trapping at the GaSb–InAs interfaces. One therefore expects the samples to show nearly intrinsic behaviour with $n_h \approx n_e$. In this case, the Fermi level lies in the minigap formed by the hybridization between electron and hole states,[4] and the conductivity is suppressed. The semiclassical Drude model then does not describe transport accurately. In a sufficiently high parallel magnetic field, the minigap is removed (*cf.* Sec. 5.5.3); comparing the conductivity in the presence of the field with the zero field value allows to estimate the suppression due to the minigap. In the majority of samples, the conductivity increases by a factor of three or greater once an in-plane magnetic field is applied. The magnitude of the ef-

[4] As the hole band is anisotropic and the position of the minigap depends on the direction, a Fermi contour (surface) none the less exists.

fect is similar to that observed by Lakrimi *et al.* [6,7] in comparable DHETs and superlattices with close to matched electron and hole concentrations.

A closer look at the behaviour at intermediate magnetic fields also exhibits evidence for the presence of a high hole concentration. The Hall resistivity $\rho_{xy}(B_z)$ of all samples shows a series of pronounced local minima which are absent from single carrier systems. In such systems, $\rho_{xy}(B_z)$ shows a step-like monotonic increase with magnetic field. At low enough temperatures, the steps eventually form plateaux with constant resistivity R_K/ν extending over the field range where the Fermi energy lies in the gap between two Landau levels; here $R_\mathrm{K} = h/e^2$ and ν is the (integer) number of occupied Landau levels [15]. This is the famous quantum Hall effect. In electron–hole systems there are separate Landau fans for electrons and holes and the quantized Hall resistivity is $\rho_{xy} = R_\mathrm{K}/(\nu_\mathrm{e} - \nu_\mathrm{h})$. If the two Landau fans overlap, as they can do in GaSb–InAs heterostructures owing to the crossed band alignment at the interface, the Fermi energy will oscillate between electron and hole levels and $\rho_{xy}(B_z)$ no longer increases monotonically [2,16,17]. The minima in the observed Hall resistivity are the remnants of this compensated quantum Hall effect after Landau level broadening.

6.4.2 Interaction

The correspondence between the Hall resistivity features in the respective traces in Fig. 6.3(b) suggests that both the position of the Fermi level and the Landau level structure are similar for the samples with 300 nm, 200 nm and 100 nm GaSb barrier thickness. One can therefore assume that in this range the interaction between the individual wells is small, an idea that is further supported by the similarity to the single well data.

The smaller barrier results show less consistency; while I cannot entirely dismiss the possibility that the differences are a result of varying sample quality caused by changes in the MOVPE growth conditions, an explanation based on the increased interaction between the individual carrier sheets is plausible. Looking back to the preliminary calculations of Sec. 6.2.3, there are several ways in which the energy spectrum might be affected: a reduction of the

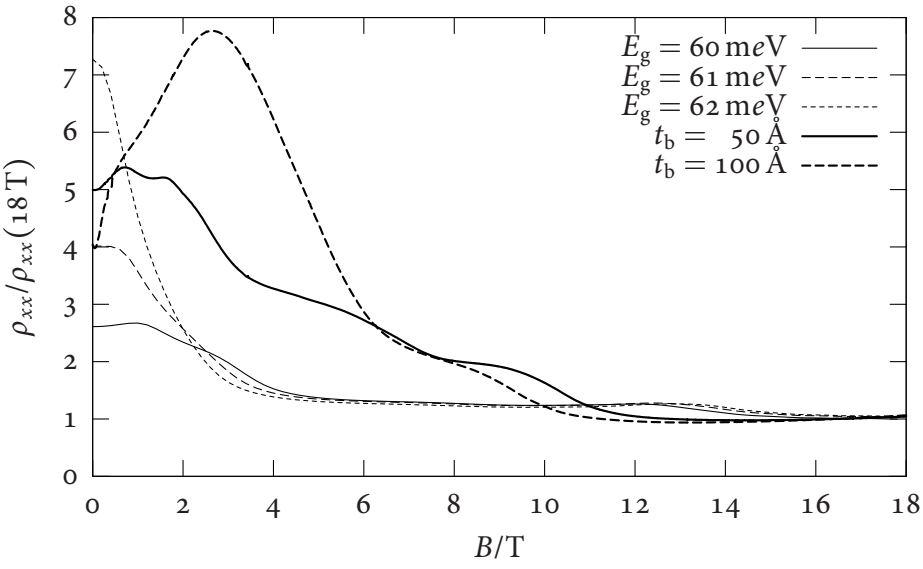


Figure 6.11: Comparison of parallel field magnetoresistivity with theory. The experimental data (thick) are taken from Fig. 6.7(a); the calculated curves assume $\Delta = 7\,meV$, $E_F = 54\,meV$, and $\delta z = 180\,Å$. The resistivity is given in terms of the resistivity at high field.

energy difference between occupied electron and hole subbands; hybridization between conduction band states in different wells; and the appearance of distinct hole states in the barrier that are no longer degenerate with hole states in the cap or buffer layer. The latter effect in particular would lead to a more complex level structure and affect the magnetic field dependence of the resistivity tensor. However, as there is still considerable broadening in the experimental data even at 0·5 K and a realistic model that correctly predicts confinement energies is still missing, a detailed test of these ideas is presently not possible.

The two-band model of Sec. 5.5.3 partially explains the basic features of most parallel field magnetoresistance measurements. Fig. 6.11 compares the experimental data of Fig. 6.7(a) to the prediction calculated from Eq. (5.18) for $E_F - E_0 = 54\,meV$, corresponding to $n_e \approx 6{\cdot}8 \times 10^{11}\,cm^{-2}$, $\delta z = 180\,Å$ and taking $\Delta = 7\,meV$ [6,18]. As explained in Sec. 5.5.3, the two prominent features correspond to parallel field values at which anticrossing occurs at the Fermi energy, *i.e.*, the sides of the (hypothetical non-interacting) hole Fermi contour pass the electron Fermi circle. For realistic values of E_F, Δ, and δz, the quantitative agreement is poor

and the model does not predict either a strong second maximum or a magnetoresistance that does not saturate at parallel fields up to 18 T. Similar pronounced double peak structures have previously appeared in superlattice measurements [7]. In both superlattices and the present double well structures, more than two carrier gases are present, and further hybridization gaps, which are not accounted for in the simplistic model, can form. Effects of the in-plane field other than the relative shift of the dispersion relations are also ignored; in particular, the scattering time τ is assumed to be independent of $B_\parallel$.

6.4.3 Symmetry

The double well structures are nominally symmetrical about the GaSb barrier, and the electron states in both InAs wells as well as the hole states in the GaSb buffer and cap layers should be equivalent. In reality, this symmetry can be broken by the Fermi level pinning at the free GaSb surface or by trapped charges at the GaSb–InAs interfaces. In fact, the electronic properties of single well DHETs grown under similar conditions are best explained by models containing only two conduction channels representing electrons in the InAs well and holes in the GaSb buffer, with no hole occupancy in the GaSb cap [4, 12].

As reasoned in Sec. 6.4.1, both wells contribute to the overall conductivity with individual electron densities similar to those of comparable DHETs. Parallel conduction through two independent electronic channels with slightly different sheet concentrations would lead to an interference between the Shubnikov–de Haas-oscillations corresponding to each channel and the emergence of a beating pattern. The oscillations of Fig. 6.6 exhibit only a single dominating frequency and no indication of beating. By taking the inverse field range over which oscillations are observable (around $1\,\mathrm{T}^{-1}$) as the minimum distance between nodes of the envelope function, one estimates an conservative upper limit on the difference between the carrier concentrations of $5 \times 10^{10}\,\mathrm{cm}^{-2}$. This translates to a maximal difference in the position of the Fermi level relative to the bottom of the conduction band of approximately 5 meV across the entire double well structure. A comparison between the behaviour of the

two wells is not sensitive to an asymmetry between the top and bottom interfaces of each individual InAs layer, and the present analysis cannot address this question.

6.5 Conclusions

I have investigated a range of samples containing two adjacent InAs wells using only magnetoresistance measurements at moderately low temperatures from 0·5 to 4·2 K and high fields up to 18 T. The samples were designed with thick cap layers of at least 1500 Å to minimize the effect of surface states and there is considerable evidence that most samples are nearly intrinsic. However, the low mobility of the hole gas and the hybridization between InAs conduction band states and GaSb valence band states prevent an accurate determination of the hole concentration.

For distances between the wells of 100 Å and above, the structures behave essentially like two DHETs conducting in parallel. Within the accuracy of the measurements, the wells behave identically, putting an upper limit of at most 5 meV on the variation of the Fermi level relative to the bottom of the conduction band throughout the double well system. At smaller well distances, qualitatively different behaviour is observed. Based on model calculations, which, however, do not accurately reflect the energy spectrum of actual samples, it is suggested that the change in hole energy levels in the GaSb barrier and the resulting change in charge distribution become important in this regime.

The magnetoresistance in a magnetic field in the plane of the InAs wells shows large and unsystematic changes between different samples and even between different parts of the same structure. The results have been discussed in the light of the two band hybridization model introduced in the previous Chapter. The theoretical model predicts a strong dependence on the position of the Fermi energy and the band overlap, which likely plays a part in the observed variation. Individual features attributed to the crossing of the electron and hole Fermi contours, which are also seen in DHETs, can be recognized in the double well data, but the quantitative agreement between theory and experiment is poor. In particular, the rel-

ative height of the peaks around 10 T that are seen in many experimental traces cannot be explained in a simple two-band model.

Bibliography

[1] D. M. Symons, M. Lakrimi, R. J. Nicholas, D. K. Maude, J. C. Portal, N. J. Mason, and P. J. Walker. Magnetic breakdown in the semimetallic InAs/GaSb system. *Physical Review B*, 58(11):7292–7299, September 1998.

[2] R. J. Nicholas, K. Takashina, M. Lakrimi, B. Kardynał, S. Khym, and N. J. Mason. Metal-insulator oscillations in a two-dimensional electron-hole system. *Physical Review Letters*, 85(11):2364–2367, September 2000.

[3] R. J. Nicholas, M. Lakrimi, B. Kardynał, S. Khym, N. J. Mason, J. Rehman, K. Takashina, P. J. Walker, D. M. Symons, D. K. Maude, and J. C. Portal. A digital quantum Hall effect. *Physica E*, 6:836–839, 2000.

[4] K. Takashina. *Magneto-transport Studies of GaSb/InAs/GaSb Double-heterostructures*. DPhil thesis, University College, University of Oxford, 2002.

[5] M. S. Daly, K. S. H. Dalton, M. Lakrimi, N. J. Mason, R. J. Nicholas, M. van der Burgt, P. J. Walker, D. K. Maude, and J. C. Portal. Zero-Hall-resistance state in a semimetallic InAs/GaSb superlattice. *Physical Review B*, 53(16):10524–10527, April 1996.

[6] M. Lakrimi, S. Khym, R. J. Nicholas, D. M. Symons, F. M. Peeters, N. J. Mason, and P. J. Walker. Minigaps and novel giant negative magnetoresistance in InAs/GaSb semimetallic superlattices. *Physical Review Letters*, 79(16):3034–3037, October 1997.

[7] M. Lakrimi, S. Khym, D. M. Symons, R. J. Nicholas, F. M. Peeters, N. J. Mason, and P. J. Walker. Mini-gaps and novel giant negative magnetoresistance in InAs/GaSb semimetallic superlattices. *Physica E*, 2:363–367, 1998.

[8] A. J. L. Poulter, M. Lakrimi, R. J. Nicholas, N. J. Mason, and P. J. Walker. Optical probing of the minigap in InAs/GaSb semimetallic superlattices. *Physica B*, 256–258:256–259, 1998.

[9] J. M. Luttinger and W. Kohn. Motion of electrons and holes in perturbed period fields. *Physical Review*, 97(4):869–883, February 1955.

[10] J. M. Luttinger. Quantum theory of cyclotron resonance in semiconductors: General theory. *Physical Review*, 102(4):1030–1041, May 1956.

[11] T. A. Vaughan. *Magneto-optics of InAs/GaSb Heterostructures*. DPhil thesis, Brasenose College, University of Oxford, 1995.

[12] C. Petchsingh. *Cyclotron Resonance Studies on InAs/GaSb Heterostructures*. DPhil thesis, Wolfson College, University of Oxford, 2002.

[13] P.-O. Löwdin. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystal. *Journal of Chemical Physics*, 18(3):365–375, March 1950.

[14] E. O. Kane. Band structure of indium antimonide. *Journal of Physics and Chemistry of Solids*, 1(4):249–261, January 1957.

[15] K. von Klitzing, G. Dorda, and M. Pepper. New method for high-accuracy determination of the fine-structure constant based on quantized Hall resistance. *Physical Review Letters*, 45(6):494–497, August 1980.

[16] E. E. Mendez, L. Esaki, and L. L. Chang. Quantum Hall effect in a two-dimensional electron-hole gas. *Physical Review Letters*, 55(20):2216–2219, November 1985.

[17] M. S. Daly, D. M. Symons, M. Lakrimi, R. J. Nicholas, N. J. Mason, and P. J. Walker. Interface composition dependence of the band offset in InAs/GaSb. *Surface Science*, 361/362:205–208, 1996.

[18] J.-C. Chiang, S.-F. Tsay, Z. M. Chau, and I. Lo. Conduction-valence Landau level mixing effect. *Physical Review Letters*, 77(10):2053–2056, September 1996.

7 Antidot Samples

> The world changed from having the determinism of a clock to having the contingency of a pinball machine.
>
> *(Heinz Rudolf Pagels, The Cosmic Code)*

7.1 Introduction

7.1.1 Overview

Lateral potential modulations of quasi-two-dimensional carrier gases have been a main focus of the research leading to this thesis. In the present chapter I shall finally report magnetotransport measurements taken on InAs–GaSb double heterostructures (DHETS) that have been modified by imposing a two-dimensional periodic superlattice potential with a lattice period a. All samples investigated in this chapter fall in the range $\lambda_F < a < \ell_f$, with electron concentrations n_e between 7·1 and 12·0 $\times$ 10^{15} cm^{-2}, corresponding to a Fermi wave length λ_F of 30 to 22 nm and an estimated mean free path ℓ_f between 1·5 and 2·7 µm at 4·2 K.[1] For this reason, the exposition of the experimental and theoretical background in Sec. 7.2 concentrates on the features of essentially classical origin resulting from the commensurability of the cyclotron radius R_c and the lattice period a, which are expected in this regime. For the quoted electron concentrations and modulation periods

[1] The electron mobility is difficult to measure because of the effect of the minigap (see Sec. 5.5.3), and the given values, which were obtained from the resistivity in a high magnetic field parallel to the plane of the sample, should be regarded as lower limits.

from 100 to 800 nm, the commensurability condition $2R_c \approx a$ is fulfilled at perpendicular magnetic fields B_z between 0·4 and 1·4 T. This field range is accessible to experiments with comparable ease and there is a temperature window in which potential commensurability peaks will not be obscured by quantum oscillations.

I had originally hoped to employ the high resolution of the local anodic oxidation (LAO) technique to create potential modulations with small periods of only a few λ_F. However, it became clear during the experiments on LAO-modified samples, which are summarized in Sec. 7.4, that the shallow surface modification that could be achieved with this method would be insufficient to generate a sufficiently strong modulation to investigate commensurability effects in this regime. Instead of further pursuing this approach, I chose to revisit the creation of antidot patterns by means of a multistep process involving deep anisotropic plasma etching,[2] and Sec. 7.5 relates comprehensive magnetotransport measurements on electron-rich DHETs patterned by this method.

7.1.2 A Note on Background Subtraction

The DHETs discussed in this thesis exhibit a strong positive magnetoresistance in a magnetic field perpendicular to the plane of the carrier sheets. This is typical for bipolar systems, and may in general terms be understood with the help of the Drude model introduced in Sec. 5.6.1. Consequently, modulations in the magnetoresistance due to the imposed potential can become obscured.[3] It is therefore sometimes desirable to subtract a smooth reference curve for clarity.

While the background in the absence of a potential modulation can be determined by measuring the magnetoresistance of a suitably chosen control sample, subtracting it is usually not helpful. Apart from any commensurability peaks that may appear, the imposed pattern typically affects the shape of the magnetoresistance curve around zero field and may

[2] See Secs. 4.2.6, 4.2.2, and 4.2.4.

[3] This is different from the situation in pure electron gases, in which the perpendicular field magnetoresistance in the absence of a potential modulation is weak and changes become immediately apparent.

change the average electron–hole ratio. The difference then still has a strong field dependence that makes the commensurability features difficult to assess.

Instead, we subtract a smooth curve that has been determined by a standard least-squares fit of an appropriate fit function to the area of interest. Since there is no known analytical model taking into account all contributions to the background,[4] we employ a simple second-order polynomial as the fit function. Such a parabola is the simplest fit function that yields a satisfactory result for background subtraction and has the added advantage that it does not have any inflection points that may lead to spurious peaks in the difference. All magnetoresistance traces in this chapter that have had a background subtracted have been prepared in this way; only the positive field part is used for fitting, and even though the magnetoresistance is symmetric, the fitted parabola is hence not necessarily centred around zero.

7.2 Background

7.2.1 Overview

Several approaches to the periodic structuring of two-dimensional carrier gases are possible, and most of them have been pursued to some extent. The carriers can be subjected to a modulated electric or magnetic field, and the modulation can be periodic in one or two dimensions. The electric potential can be varied directly, by modulating the charge density at a nearby surface—*e.g.*, by creating local variations of surface states or by using a patterned gate electrode [1]—or indirectly, as a piezoelectric potential in a strained material [2]. In both cases, the pattern has to be defined lithographically; in systems such as GaAs–$Al_xGa_{1-x}As$ heterostructures, in which carriers can be created by photoabsorption, a periodic modulation of the carrier concentration can also be achieved by holographic means. The magnetic

[4] It is tempting to use a function of the form $y(x) = a(1 + bx^2)/(1 + cx^2)$ derived from the simple model of Eq. (5.25), as this gives a good apparent fit for most experimental curves. However, the Drude model ignores electron–hole interaction and can in any case not account for the effect of a periodic modulation on the low-field magnetoresistance—the good agreement is merely a result of the large number of parameters. Subtracting such a background usually works fairly well, but may introduce additional artificial peaks.

field can be modulated by bringing the carrier sheet close to superconducting or ferromagnetic structures exhibiting the desired periodicity.

The theoretical treatment of the transport properties in the presence of such modulations has been centred around two complementary approaches: perturbation theory, which is valid for weak periodic components $V(\mathbf{r}) \ll E_\mathrm{F}/e$, and numerical simulations of individual carriers, which are also applicable to antidots and the intermediate range. The pertinent features of modulated quasi-two-dimensional electron gases (2DEGs) are now well understood on the basis of such calculations. In the case of strong modulations the carrier dynamics become fully chaotic.

7.2.2 Weak Modulations

Evidence

The effect of weak lateral modulations on the transport in a 2DEG was first studied in 1989 by Weiss *et al.* [3–5]. They used the persistent photoconductivity effect of GaAs–$Al_xGa_{1-x}As$ heterojunctions to realize, by means of holographic illumination, one-dimensional modulations of the electron density of a 2DEG with a sub-micron period a. At liquid helium temperatures, the magnetoresistance was measured with the current flowing either parallel (ρ_{xx}) or perpendicular (ρ_{yy}) to the direction in which the potential varied. Low-field oscillations appeared, which were periodic in $1/B_z$; the position of the minima of $\rho_{xx}(B_z)$ could be described by the commensurability condition

$$2R_\mathrm{c}(B_z) = (\lambda - \varphi)a; \qquad \lambda \in \mathbb{Z}^+, \tag{7.1}$$

where $R_\mathrm{c}(B_z)$ is the cyclotron radius of the two-dimensional electrons at the Fermi level E_F and $\varphi \approx \frac{1}{4}$ is a phase shift. In ρ_{yy}, much weaker oscillations, which were shifted by half a period, could be observed, while no oscillations were detected in the Hall resistivities ρ_{xy} and ρ_{yx}. The magnetotransport features were later confirmed in similar experiments [1] and have since become known as *Weiss oscillations*.

These studies were soon extended to two-dimensional modulations, for which weaker oscillations were seen in both $\rho_{xx}(B_z)$ and $\rho_{yy}(B_z)$; their phase corresponded to that of the $\rho_{yy}(B_z)$ oscillations in the one-dimensional case [6]. Magnetoresistance oscillations in a periodic *magnetic* field were first reported by Ye *et al.* [2] in 1995; compared to experiments imposing electric field modulations, the minima of the $\rho_{xx}(B_z)$ oscillations were shifted by half a period, corresponding to $\varphi \approx -\frac{1}{4}$ in Eq. (7.1).

Quantum Model

Following the original approach by Gerhardts [5], the magnetoresistance features in the presence of a one-dimensional modulation can be understood by assuming a harmonic modulation of the electric potential $V(x) = V_0 \cos Kx$, where $K \stackrel{\text{def}}{=} 2\pi/a$, and modifying Eq. (5.28) accordingly. The one-dimensional Hamiltonian of the modulated system then becomes

$$\hat{H} = \hat{H}_0 + V(x) = -\frac{\hbar^2}{2m^*}\frac{\partial^2}{\partial x^2} + \frac{e^2 B_z^2 (x - x_0)^2}{2m^*} + V(x), \tag{7.2}$$

where $\hat{H}_0$ is the Hamiltonian in the absence of a modulation and $x_0 = -\hbar k_y / eB_z$ the orbit centre coordinate.[5]

Eq. (7.2) can be diagonalized numerically, or $V(x)$ may be treated as a perturbation to $\hat{H}_0$. In this case, the first-order correction to the energy is

$$\langle \psi_N | V(x) | \psi_N \rangle = V_0 \cos(Kx_0) e^{-X/2} L_N(X), \tag{7.3}$$

where ψ_N is the unperturbed wave function in Landau level N, $X \stackrel{\text{def}}{=} K^2\hbar/2eB_z$, and $L_N(X)$ the N^{th} Laguerre polynomial. For fixed X, $L_N(X)$ oscillates as a function of the index N, leading to a broadening of the individual Landau bands that oscillates with the Landau level index; for $L_N(X) = 0$, the energy correction vanishes and there is no broadening. As the λ^{th} zero of $L_N(X)$ is approximately $X_\lambda^{(N)} \approx \left[(\pi/2)(\lambda - \frac{1}{4})\right]^2 \big/ (N + \frac{1}{2})$, the flat-band condition

[5] In Sec. 5.6.2, $\xi = x - x_0$. This substitution is no longer useful, as the eigenvalues are no longer independent of the orbit position if a modulation is present.

in terms of the cyclotron radius $R_c^{(N)} = (2/K)\left[X_\lambda^{(N)}(N+\frac{1}{2})\right]^{1/2}$ of the N^{th} Landau Level is $2R_c^{(N)} = (\lambda - \frac{1}{4})a$. If $R_c^{(N)}$ is identified with the cyclotron radius at E_F, Eq. (7.1) is recovered.

The eigenstates $\psi_N^{(x_0)}$ of $\hat{H}$ have a finite group velocity and carry current in the y-direction [5]:

$$\left\langle \psi_N^{(x_0)} \middle| \hat{v}_y \middle| \psi_N^{(x_0)} \right\rangle = -\frac{1}{\omega_c}\frac{\partial E_N(x_0)}{\partial x_0}, \tag{7.4}$$

where $E_N(x_0)$ is the energy eigenvalue in the N^{th} Landau level. Using Kubo's linear response theory [7] (see Appendix B) with the assumption of a constant relaxation time τ, the contribution to the conductivity σ_{yy} due to such current carrying states can be estimated in terms of the velocity matrix elements as [5]

$$\Delta\sigma_{yy} = -\frac{2e^2}{2\pi}\frac{eB_z}{\hbar}\int_0^a \frac{dx_0}{a}\sum_N \tau f'\left[E_N(x_0)\right]\left|\left\langle \psi_N^{(x_0)} \middle| \hat{v}_y \middle| \psi_N^{(x_0)} \right\rangle\right|^2, \tag{7.5}$$

where $f'(E)$ is the derivative of the Fermi-Dirac distribution. As a consequence of Eqs. (7.5), (7.4), and (7.3), $\sigma_{yy}(B_z)$ exhibits oscillations with minima at magnetic fields for which E_F lies in a flat Landau band. If $\sigma_{xy}^2 \gg \sigma_{xx}\sigma_{xy}$, which is realistic for the 2DEGs originally investigated, $\rho_{xx} \approx \sigma_{yy}/\sigma_{xy}^2$, and the minima in $\rho_{xx}(B_z)$ occur at the same positions, which are given by Eq. (7.1) as explained above.

The remainder of the conductivity is due to scattering between Landau levels; in the unmodulated system, this is the only contribution. The oscillations in ρ_{yy} arise from this term and can be explained by abandoning the assumption of a constant relaxation time [6, 8, 9]. The effect of the impurity potential can then be described with the help of Green's functions in the self-consistent Born approximation. This results in a modulation of the peak height of the density of states oscillations, and the height becomes maximal for those peaks corresponding to flat Landau bands. The components of the conductivity tensor, when calculated using the Kubo formula, are essentially proportional to the square of the density of states, leading to maxima at the flat band condition, Eq. (7.1). The Landau band conductivity discussed in the previous paragraph, which has minima at the same magnetic field values, dominates σ_{yy} (and hence ρ_{xx}) while oscillations with maxima at these positions appear in σ_{xx} (and ρ_{yy})—in agreement with experiment.

On extending the treatment to two-dimensional potentials [6, 8, 9], the bandwidth oscillations for each direction are individually found to be the same as in the one-dimensional case. However, as discussed in the description of the Hofstadter spectrum (see Sec. 1.2), each Landau level is split into subbands; if the electron mobility is high enough that the splitting between a given pair of subband groups is resolved, the matrix element between corresponding states does not contribute to $\Delta\sigma_{yy}$ and the band conductivity is accordingly strongly suppressed for high mobility samples. In this case, the observable oscillations are dominated by the density of states and ρ_{xx}, like ρ_{yy}, has maxima at the fields given by Eq. (7.1).

Vasilopoulos and Peeters [10, 11] as well as Xue *et al.* [12] have used the same *Ansatz* for weak magnetic modulations $B_z^{(\mathrm{m})}(x) = B_0 \cos Kx$. The magnetic vector potential in the Landau gauge is $\mathbf{A} = \big[0, B_z x + (B_0/K)\sin Kx, 0\big]$ and the Hamiltonian becomes $\hat{H} = \hat{H}_0 + \hat{H}_1$, where

$$\hat{H}_1 = \frac{\omega_1}{K}(\hat{p}_y + eB_z x)\sin(Kx) + \frac{m^*\omega_1^2}{4K^2}\big[1 - \cos(2Kx)\big] \tag{7.6}$$

can be treated as a perturbation if $\omega_1 \stackrel{\mathrm{def}}{=} eB_0/m^* \ll \omega_\mathrm{c}$. Ignoring terms quadratic in ω_1, the first-order correction to the energy is then

$$\big\langle \psi_N \big| \hat{H}_1 \big| \psi_N \big\rangle = \hbar\omega_1 \cos(Kx_0) e^{-X/2}\left[L_{N-1}^{(1)}(X) + \frac{1}{2}L_N(X)\right], \tag{7.7}$$

where $L_N^{(1)}$ is a generalized Laguerre polynomial. For large Landau level indices N, the energy correction, and hence the Landau band width, is approximately proportional to $\sin(KR_\mathrm{c} - \pi/4)$, so that flat bands occur for $2R_\mathrm{c} \approx (\lambda + \frac{1}{4})$. Compared to the situation for an electric modulation, the magnetoresistance oscillations are shifted by a quarter period, as observed in experiments.

Semiclassical Model

Since the commensurability oscillations—unlike the Shubnikov–de Haas-oscillations described in Sec. 5.6.3—are not a quantization effect, they can also be described by the classical motion of electrons at the Fermi energy. In such a semiclassical model, one implicitly assumes that transport is ballistic on the scale of the cyclotron orbits, *i.e.*, $\tau \gg 1/\omega_\mathrm{c}$. Beenak-

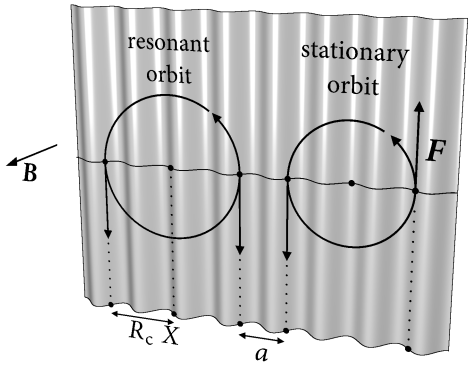


Figure 7.1: One-dimensional guiding centre model

KER [13] has suggested an explanation based on the drift of the *guiding centres* of individual electrons. He defines the guiding centre of a (nearly) free electron at $\mathbf{r}$ having velocity $\mathbf{v}$ as $\mathbf{R} = \mathbf{r} + \hat{\mathbf{z}} \times \mathbf{v}/\omega_c$, where $\hat{\mathbf{z}}$ is the unit vector in the z-direction. In the presence of a potential modulation $V(\mathbf{r})$ giving rise to an electric field $\mathbf{E}(\mathbf{r}) = \nabla_{\mathbf{r}} V(\mathbf{r})/e$, the drift of the guiding centre is $\mathbf{v_R} = \dot{\mathbf{R}} = \hat{\mathbf{z}} \times \mathbf{E}(\mathbf{r})/B_z$. For a weak potential, the cyclotron orbits remain approximately circular and the time average of $\mathbf{v_R}$ can be obtained by integrating the electric field along such orbits:

$$\overline{\mathbf{v}}_{\mathbf{R}}(\mathbf{R}) \approx \frac{1}{2\pi} \int_0^{2\pi} \mathbf{v}_{\mathbf{R}}(\mathbf{R}, \varphi)\, d\varphi = -\frac{1}{2\pi B_z} \int_0^{2\pi} \hat{\mathbf{z}} \times \mathbf{E}\left[\mathbf{R} + \hat{\mathbf{z}} \times \frac{\mathbf{v}(\varphi)}{\omega_c}\right] d\varphi. \tag{7.8}$$

BEENAKKER [13] notes that for a one-dimensional potential with period $a \ll R_c$ the contributions to $\overline{v}_Y(X)$ average out for most of the cyclotron orbit and the integral is dominated by the drift at the extremal points $X \pm R_c$ as shown in Fig. 7.1; here X and Y denote the components of $\mathbf{R}$. He approximates Eq. (7.8) for a harmonic potential of the form $V(\mathbf{r}) = V(x) = V_0 \cos Kx$, and, on averaging over X, finds $\sqrt{\langle \overline{v}_Y^2 \rangle_X} \approx (V_0/m^*\omega_c)\sqrt{K/\pi R_c}\cos(KR_c - \pi/4)$. This drift causes diffusion of the electrons with a diffusion coefficient $\Delta D_{yy} \approx \tau \langle \overline{v}_Y^2 \rangle_X$. One can use the Einstein-Smoluchowski relation $\boldsymbol{\sigma} = e^2 g(E)\mathbf{D}$, where $g(E)$ is the density of states, to calculate the effect of $V(x)$ on σ_{yy}. Since $g(E) = m^*/\pi\hbar^2$ for a 2DEG,

$$\Delta\sigma_{yy} = \frac{e^2 m^* \tau}{\pi\hbar^2} \left\langle \overline{v}_Y^2 \right\rangle_X \approx \frac{e^2\tau}{\pi^2 m^*} \frac{V_0^2}{(\hbar\omega_c)^2} \frac{K}{R_c} \cos^2\left(KR_c - \frac{\pi}{4}\right). \tag{7.9}$$

Eq. (7.9) evidently has minima at the values of $R_c(B_z)$ given by Eq. (7.1), and if $\sigma_{xy}^2 \gg \sigma_{xx}\sigma_{xy}$, these correspond to minima in $\rho_{xx} \approx \sigma_{yy}/\sigma_{xy}^2$. Beenakker is thus able to successfully explain the ρ_{xx} oscillations observed experimentally. While the approximation of the diffusion tensor leads to a small effect on ρ_{yy} and ρ_{xy}, he points out that in a more rigorous calculation of $\mathbf{D}$ based on the Boltzmann equation oscillations are predicted in ρ_{xx} only.

Gerhardts has extended the guiding centre model to arbitrary two-dimensional periodic potentials $V(\mathbf{r}) = \sum_{\mathbf{K}} V_{\mathbf{K}} \exp(i\mathbf{K}\cdot\mathbf{r})$ [14] and commensurate magnetic modulations $B_z^{(\mathrm{m})}(\mathbf{r}) = \sum_{\mathbf{K}} B_{\mathbf{K}} \exp(i\mathbf{K}\cdot\mathbf{r})$ [15,16]. Here $\mathbf{K} \stackrel{\text{def}}{=} 2\pi(n_x/a_x, n_y/a_y)$, a_x and a_y are the periods, and $n_x, n_y \in \mathbb{Z}^+$. $V_{\mathbf{K}}$ and $B_{\mathbf{K}}$ are the Fourier coefficients of the electric and magnetic modulations, respectively. Using $\Delta D_{\mu\nu} \approx \tau\langle \overline{v}_M \overline{v}_N \rangle_{XY}$ (averaging over both X and Y), he arrives at a generalized form of Eq. (7.9):

$$\Delta\sigma_{\mu\nu} \approx \frac{e^2\tau}{\pi m^*(\hbar\omega_c)^2} \sum_{\mathbf{K}} \kappa_{\mu\nu} \left| \operatorname{sgn}(B_z) V_{\mathbf{K}} J_0\big(|\mathbf{K}|R_c\big) + \frac{k_F}{|\mathbf{K}|} \frac{\hbar e B_{\mathbf{K}}}{m^*} J_1\big(|\mathbf{K}|R_c\big) \right|^2, \tag{7.10}$$

where $\kappa_{xx} \stackrel{\text{def}}{=} K_y^2$, $\kappa_{yy} \stackrel{\text{def}}{=} K_x^2$, $\kappa_{xy} \stackrel{\text{def}}{=} -K_x K_y$, k_F is the Fermi wave number, and J_n is the n^{th} order Bessel function of the first kind. While Eq. (7.10) correctly predicts the effect of realistic one-dimensional potential modulations, it cannot fully account for the suppression of the Weiss oscillations seen with two-dimensional potentials.

This suppression can none the less be explained in the semiclassical guiding centre picture [17]: In a two-dimensional superlattice, a large proportion of the *guiding centre* orbits form closed loops, and if the scattering time is sufficiently large, the drift velocity of these orbits averages to zero. One can calculate this effect quantitatively by using a classical Kubo formula[6] (see Appendix B) to derive a more accurate expression for $\boldsymbol{\sigma}$ in terms of the average of the autocorrelation function of the guiding centre velocity. Assuming a constant transport relaxation time τ and negligible broadening of the Fermi contour,

$$\Delta\sigma_{\mu\nu} = \frac{m^* e^2}{\pi\hbar^2} \int_0^\infty e^{-t/\tau} \big\langle v_M(t) v_N(0) \big\rangle^{(E_F)} \, dt, \tag{7.11}$$

[6] Which is equivalent to Chambers' formula under the assumptions made here.

where averaging is now over the phase space at fixed energy E_F [15, 17]. Eq. 7.11 can then be evaluated numerically by simulating a set of guiding centre trajectories corresponding to the required section of the phase space and computing $\left\langle v_N(0)v_M(t)\right\rangle^{(E_F)}$ from it.

7.2.3 Strong Modulations—Antidots

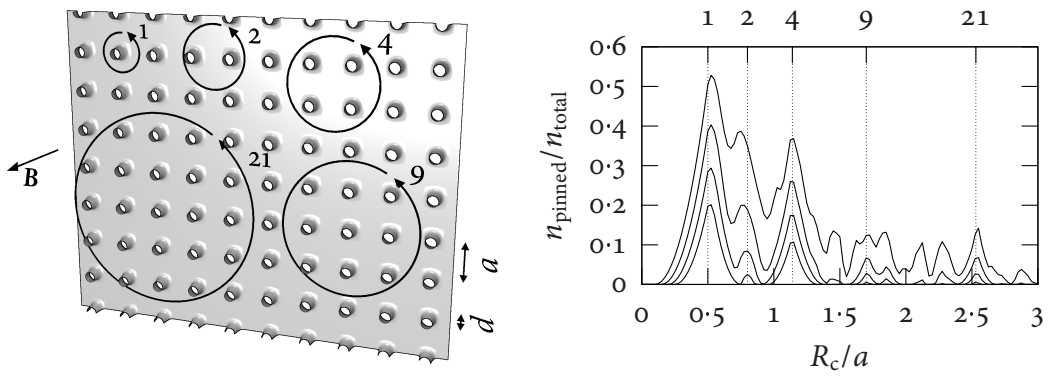
Evidence

Antidot patterns, in which the potential modulation is strong enough to entirely deplete the carrier system in well-defined areas, were first reported by Weiss *et al.* [18] in 1991. They used electron beam lithograph (EBL) and reactive ion etching (RIE) to create a square array of holes with diameter d and sub-micron period a in a GaAs–$Al_xGa_{1-x}As$ heterostructure; the high mobility 2DEG formed at the junction had a mean free path of several micron at liquid helium temperatures. The magnetoresistance ρ_{xx} of such structures showed a number of pronounced peaks in the low field region and a strong increase near $B_z = 0$ compared to an unpatterned sample; in the Hall resistance ρ_{xy}, plateau-like features were seen at the same field values. In contrast to the situation for weak potential modulations, the peaks are not periodic in $1/B_z$ and cannot be described by a simple formula like Eq. (7.1).

Ballistic transport in antidot arrays has since been studied in considerable detail in GaAs–$Al_xGa_{1-x}As$ heterostructures [19, 20]. The effects of large [21] and non-circular [22–24] antidots, as well as generalized rectangular lattices [25] have been investigated. A number of experiments were performed on InAs–GaSb heterostructures [21], which were chosen for the absence of a depletion length at exposed InAs surfaces; in these studies, the GaSb cap layer was very thin and no mobile holes were assumed to be present.

Models

The magnetoresistance peaks in antidot lattices were originally explained by Weiss *et al.* [18] within the semiclassical *pinned orbit* model illustrated in Fig. 7.2. In this model, the cyclotron orbits with radius $R_c(B_z)$ are divided into *scattering* orbits, which collide with antidots,



(a) Pinned orbits around 1, 2, 4, 9, and 21 antidots, corresponding to $R_c/a = 0{\cdot}50, 0{\cdot}80, 1{\cdot}14, 1{\cdot}70$, and $2{\cdot}53$.

(b) Proportion of pinned orbits as a function of R_c/a for different values of d/a ranging from 0·5 (bottom) to 0·2 (top) in steps of 0·1. The calculation follows Weiss *et al.* [18].

Figure 7.2: Pinned orbit model.

drifting orbits, which drift without encountering antidots at all, and *pinned* orbits, which encircle a number of antidots without colliding with them. It is assumed that pinned orbits become localized and do not contribute to conduction since the potential gradient close to the antidots guides the electrons and counteracts their drift in the Hall field $E_y^{(H)}$. The proportion of the different types of orbits depends on $R_c(B_z)/a$ and d/a; the fraction of pinned orbits is high if $R_c(B_z)$ has values that allow the cyclotron orbits to fit exactly around a specific number of antidots. For smaller (rescaled) dot diameters d/a, more pinned orbits become possible, yet the peak positions stay approximately constant except for the resonance corresponding to cyclotron orbits encircling two antidots.[7] The magnetoresistance $\rho_{xx}(B_z)$ exhibits peaks at the values of B_z for which the proportion of pinned orbits attains a local maximum. The pinned orbit model predicts peak positions that are in good agreement with a large number of experimental results but makes several *ad hoc* assumptions.

A more robust approach consists in directly calculating electron trajectories by numerical integration of the equations of motion for a suitable model potential (see Appendix C). A chaotic orbit near a commensurability resonance that has been computed in such a way appears in Fig. 7.3. Provided a set of orbits with an appropriate phase space distribution has

[7] The stronger dependence of this peak on d/a is readily understood from geometry, *cf.* Fig. 7.2(a).

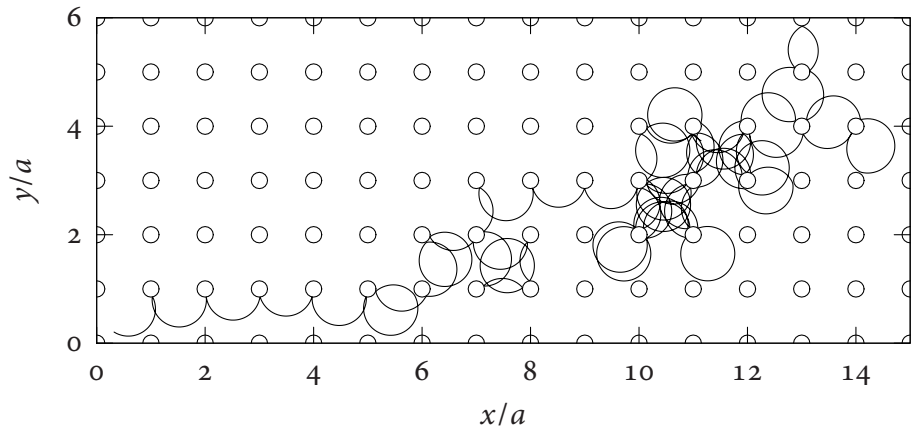


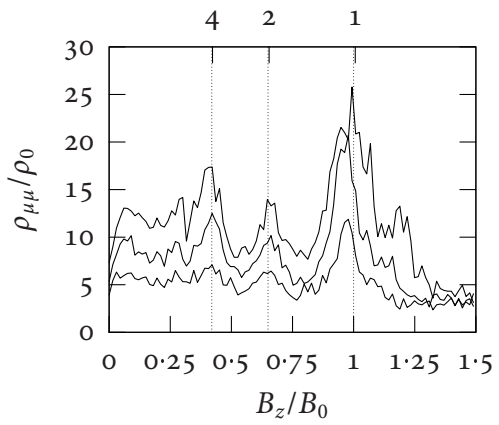
Figure 7.3: Chaotic electron trajectory in an antidot lattice with lattice constant a for $2R_c \approx a$.

been simulated, the conductivity $\boldsymbol{\sigma}$ (and hence the resistivity $\boldsymbol{\rho} = \boldsymbol{\sigma}^{-1}$) can be calculated from the the velocity autocorrelation using a classical Kubo formula [7] in the same way as explained in Sec. 7.2.2 for the motion of guiding centres: assuming a constant phenomenological relaxation time τ, the static conductivity at zero temperature is

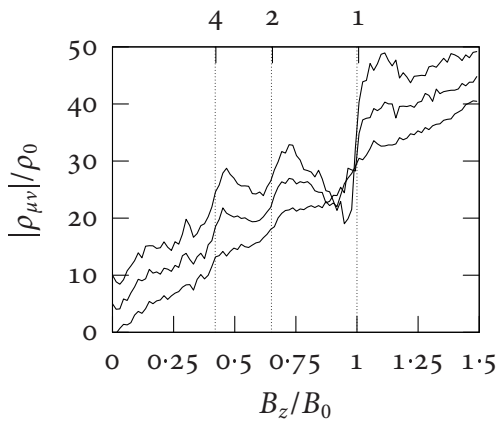
$$\sigma_{\mu\nu} = \frac{m^* e^2}{\pi\hbar^2} \int_0^\infty e^{-t/\tau} \left\langle v_\mu(t) v_\nu(0) \right\rangle^{(E_F)} dt, \tag{7.12}$$

where m^* is the effective mass of the electrons and averaging is over the phase space at constant energy E_F (see Appendix B). Such calculations were originally proposed by Fleischmann *et al.* [26] and have since been used extensively by other authors [21, 23, 27–29]. They are easily adapted to different potentials and can be justified whenever the semiclassical approximation is valid, *i.e.*, carriers that participate in scattering can be regarded as moving ballistically with an energy determined by the Fermi level E_F.

Fig. 7.4 shows the resistivity simulated for a steep antidot potential. The positions of the peaks in the diagonal components of $\boldsymbol{\rho}$ (corresponding to the longitudinal resistance) closely mirror those predicted by the pinned orbit model. The off-diagonal components (corresponding to the Hall resistance) exhibit corresponding derivations from linearity. For small dot diameters d/a, step structures similar to those observed in the original experiments [18] appear, which gradually change to dips for large d/a. A careful consideration of phase-space



(a) Diagonal components of the resistivity tensor as a function of B_z for $d/a = 0{\cdot}2$ (bottom), 0·3, and 0·4.



(b) Off-diagonal (Hall) components of the resistivity tensor as a function of B_z for $d/a = 0{\cdot}2$ (bottom), 0·3, and 0·4. The curves are offset by $5\rho_0$ for clarity.

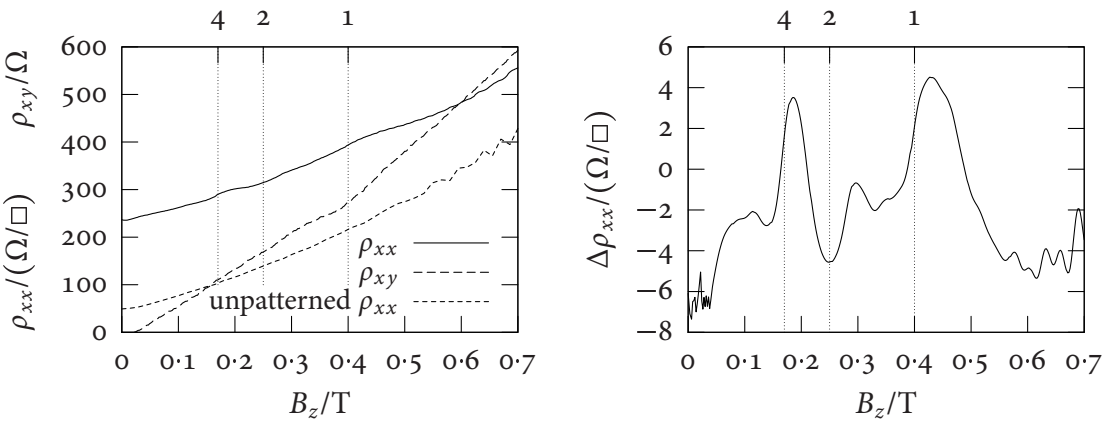
Figure 7.4: Calculating the resistivity via semiclassical simulations. $B_0 = 2m^* v_F/ea$ is the field for which $2R_c = a$ and $\rho_0 = (2\pi\hbar^2)/(m^* e^2 \tau v_F^2)$ is the Drude resistivity.

maps reveals that the contribution to the resistance due to pinned orbits is actually comparatively small. Even if those orbits are disregarded, strong commensurability peaks remain in the calculated resistivity; they can be attributed to chaotic orbits that lie close to the islands of stability that are formed by pinned and precessing orbits and have similar dynamics [26]. Further insight into the physical mechanisms underlying the structure of the classical velocity autocorrelation function may be gained by classifying chaotic skipping orbits according to the direction taken after a small number of hops [27].

7.3 Initial Results

7.3.1 Commensurability Features in the GaSb–InAs–GaSb System

Lateral modulations were first realized in hole-rich InAs–GaSb DHETs by Javed Rehman and Beata Kardynał in Oxford [30]. A number of DHET samples with 300 Å InAs wells and 900 Å GaSb cap layers were patterned with square antidot lattices by electron beam



(a) Magnetoresistance and Hall resistance

(b) Magnetoresistance after background subtraction

Figure 7.5: 700 nm antidot lattice on an InAs–GaSb DHET (0X3513) with a 300 Å InAs well and a 900Å GaSb cap [30]. The pattern was created by EBL and transferred by RIE, creating holes with an estimated depth of 140 nm. The dotted vertical lines show the expected commensurability peaks for pinned electron orbits around 1, 2, and 4 antidots calculated using the electron density deduced from the Shubnikov–de Haas-oscillations [18].

lithography. The patterns were transferred by either RIE or wet chemical etching using the tartrate-based etch described in Sec. 4.2.5; it was estimated that the RIE patterns reached down to the InAs layer, while the wet etched patterns took away approximately 60 % of the GaSb cap layer. Modulation periods between 400 and 800 nm were attempted. Given the limited performance of the available electron beam equipment, periods of 650 nm and above could reliably be achieved with this method.

The perpendicular field magnetoresistance of these patterned samples, when measured in a Hall bar geometry at low temperatures (0·3 to 20 K), typically exhibits an increased resistance around zero field and at least two clearly resolved peaks at low magnetic field. At the magnetic field values of the strongest magnetoresistance peaks, the Hall resistance also deviates from its behaviour in unpatterned samples. This behaviour is illustrated in Fig. 7.5 for a typical sample; the features show little reduction in amplitude for temperatures up to 20 K. The position of these peaks could be explained by assuming that they result from the commensurability of electron-like orbits with the imposed modulation according to the pinned

orbit model. The cyclotron radius R_c is calculated from the electron concentration n_e by the free electron expression

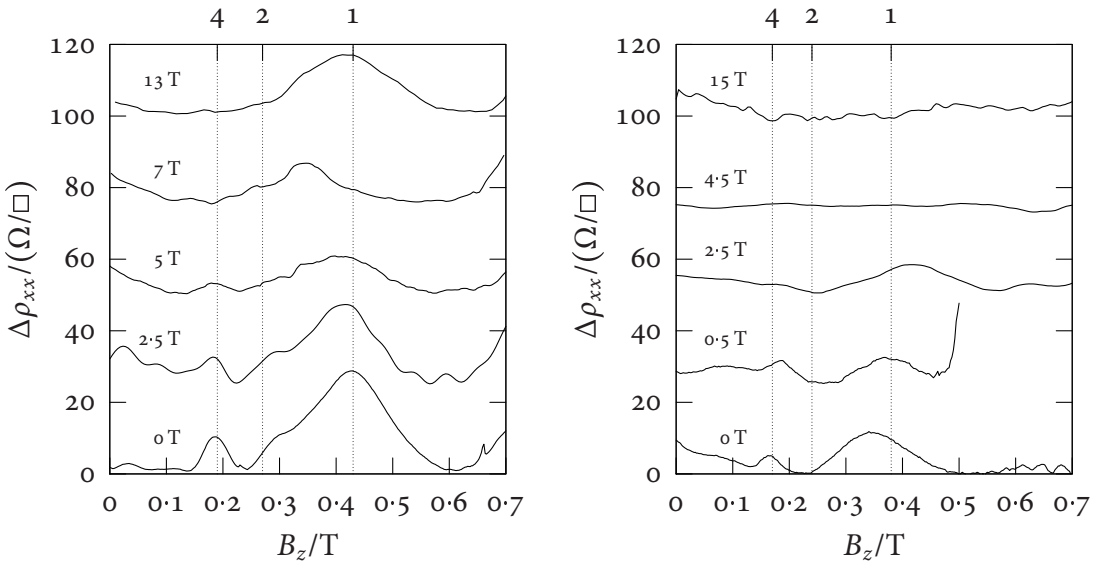
$$R_c = \ell_m^2 k_F = \frac{\hbar\sqrt{2\pi n_e}}{eB_z}. \tag{7.13}$$

As explained in Sec 5.5.4, such orbits become possible via magnetic breakdown even though the Fermi contours in the absence of a magnetic field have a more complicated shape owing to electron–hole mixing.

7.3.2 Behaviour in the Parallel Field

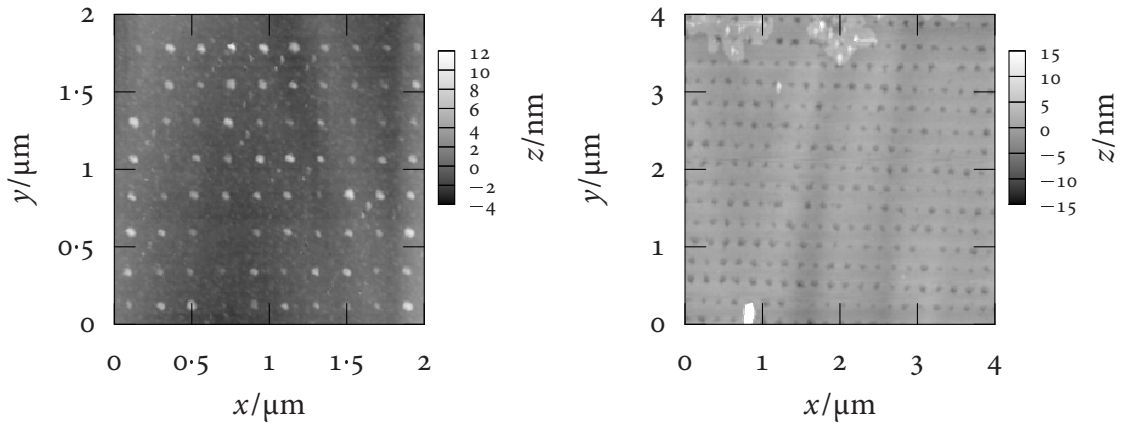
Beata Kardynał and I [31] further investigated the behaviour of these samples in the presence of an in-plane magnetic field. According to the reasoning of Sec. 5.5.3, such a magnetic field was expected to decouple the strong interaction of the electrons in the InAs layer with the mobile holes in the GaSb layers close to the interface while having only a small effect on the envelope wave functions in the direction perpendicular to the carrier sheets.

As shown in Fig. 7.6, the behaviour differed significantly between deep dry etched and shallow wet etched samples, which both exhibited similar commensurability features in the absence of an in-plane field. While the commensurability peaks disappeared completely in the wet etched samples for $B_\parallel \geq 4{\cdot}5\,\mathrm{T}$, they persisted in the dry etched samples even if a strong in-plane field was present. Both classes of samples showed small shifts in the peak positions as a function of $B_\parallel$. We tried to explain this behaviour by assuming that the wet-etched pits, being shallower, only represented an antidot potential for the mobile holes in the GaSb; the electrons in the InAs, being of opposite charge, would experience the antidot sites as attractive. While the commensurability features could be assigned to magnetic breakdown orbits closely following the electron cyclotron orbits, the interaction the electrons and holes in the absence of an in-plane field would cause increased scattering of the electrons at the antidots. In a large in-plane field, removal of the interaction would prevent such a scattering mechanism and consequently the appearance of commensurability peaks.



(a) RIE sample (b) Wet etched sample

Figure 7.6: Dry and wet etched antidot samples in a parallel field [31]. The magnetoresistance after background subtraction is shown for several values of the in-plane field $B_\parallel$; the individual curves are shifted by 25 Ω/□. Both samples were produced from InAs–GaSb DHETs with a 300 Å InAs well and a 900 Å GaSb cap and were patterned with a 650 nm period square antidot lattice using EBL. For sample (a), the pattern was transferred into the DHET (OX3251) using RIE and it extends through the InAs layer; sample (b) was created using isotropic wet chemical etching. The dotted vertical lines show the expected commensurability peaks for pinned electron orbits around 1, 2, and 4 antidots calculated using the electron density deduced from the Shubnikov–de Haas-oscillations [18].



(a) Before removal of oxide

(b) After removal of oxide

Figure 7.7: AFM micrographs of sample OX4531B before and after deoxidation

7.4 Samples Created by Direct Surface Modification

7.4.1 Overview

Using the local anodization technique described in Chapter 3, I prepared a substantial number of samples by patterning the surfaces of Hall bars that were created from substrates containing InAs–GaSb heterostructures. All heterostructures were grown in Oxford using metal-organic vapour phase epitaxy (MOVPE) as explained in Chapter 5. The modifications, which had an aspect ratio of 3 : 1, entirely covered Hall bars with a width of 5 or 10 µm and an aspect ratio of 2 : 1. For the samples discussed in this section, the patterns took the shape of square arrays of dots with a lattice constant a between 100 and 400 nm. The oxide dots had diameters ranging between 50 and 100 nm and were approximately 5 to 10 nm high. This height corresponded to the oxidation of 10 to 20 nm of a GaSb surface layer and was the largest vertical extent of the modification that could be achieved with sufficient reproducibility. I further modified a number of samples by chemically dissolving the oxide (see Chapter 3), leaving pits in place of the dots as seen in Fig. 7.7. Samples of the type discussed here occasionally exhibited a change in behaviour after removal from and reintroduction into a cryogenic experiment. Such degradation was *not* always seen and was not investig-

ated systematically in great detail. Given the sensitivity of the electrical properties to the surface states and the solubility of the GaAs oxide mixture in water, I consider the exposure to condensation, which could not be avoided entirely with some inserts, a likely cause of the problem. Another possible issue that is consistent with measurements is physical damage to the voltage probes.

7.4.2 Double Heterostructures

Patterned double heterostructures comprised both thick (800 to 1,200 Å) cap samples with a high concentration of mobile holes approaching that of the conduction electrons (0x3729, 0x3730, 0x3733, 0x4256, 0x4340) and thin (500 to 700 Å) cap samples that lay in the region of the parameter space for which we previously demonstrated that a change of the cap layer thickness has a large effect on the hole concentration (0x3735, 0x4434, 0x4530, 0x4531, 0x4532).[8] While it was in doubt whether the surface modification with the atomic force microscope (AFM) would create a sufficient potential modulation at the InAs layer to produce a measurable effect in the former set of samples, the previous success with partial removal of the cap layer reported in Sec. 7.3 and the sensitivity of the carrier concentration to the cap thickness led us to expect substantial modulation effects in the latter set—after all, up to 30 % of the cap layer could be affected by the anodization.

None the less, magnetotransport measurements at 0·5 to 4·2 K following the procedure laid down in Chapter 4 revealed no compelling evidence for commensurability features in any of these samples. While various background substitution approaches showed small reproducible resistivity fluctuations above the noise threshold, these were not correlated with the surface modification; the behaviour of the magnetoresistance and Hall resistance was qualitatively similar for both patterned and control Hall bars. The outcome of these experiments is summarized in Table 7.1, which gives a limited number of quantitative results in a concise form. Particular attention is given to the electron concentration, as this quantity is

[8] Chapter 5 explains the influence of the surface states and the cap layer thickness on the electronic properties of the DHET in detail.

Substrate	t_{cap} (Å)	Sample	a (nm)	T (K)	Control Area n_e (10^{11}/cm^2)	Control Area $\rho_{xx}(0)$ (Ω/□)	Patterned Area n_e (10^{11}/cm^2)	Patterned Area $\rho_{xx}(0)$ (Ω/□)
OX4256	1,000	A	400	4·2	7·7	37	7·9	69
		B	400	4·2	7·7	45	8·0	46
		C	400	4·2	7·9	48	8·1	41
OX4434	700	A	200	4·2	8·1	76	7·9	89
OX4340	1,200	A	200	4·2	7·0	186	7·3	173
		B	200	4·2	7·5	186	7·3	210
OX3735	600	A	200	4·2	8·9	214	9·1	221
OX3733	1,200	A	200	4·2	7·6	179	7·6	206
OX3729	800	A	200	4·2	7·6	84	7·5	80
OX3730	1,000	A	200	4·2	7·4	159	7·4	146
		B	200	4·2	7·5	148	7·6	109
		C, deox.	200	4·2	7·0	46	7·0	39
OX4530	700	A	200	4·2	6·8	102	6·8	102
		A	200	10	6·7	113	6·7	113
		A	200	15	6·9	146	6·8	132
		A	200	20	—	197	—	155
		B	400	4·2	7·1	112	7·1	110
		B	400	10	7·3	118	7·2	119
		B	400	15	7·2	143	7·1	142
		B	400	20	—	186	—	183
		C	200	0·5	7·3	88	7·4	94
		D	400	0·5	7·6	81	7·4	89
OX4531	500	A, deox.	100	4·2	7·7	73	7·7	85
		B, deox.	200	4·2	7·9	76	7·9	83
		C, deox.	400	4·2	7·6	79	7·6	73
		C, deox.	400	1·5	7·5	74	7·5	75
OX4532	680	A	200	0·5	7·3	101	7·3	121
		B	400	0·5	7·3	83	7·3	83

Table 7.1: Direct surface modification. The electron concentration n_e was calculated from the frequency of the Shubnikov-de Haas-oscillations in the perpendicular field magnetoresistance; it is accurate to about 2×10^{10} cm^{-2}. The absolute error in the resistivity is determined by the uncertainty in the precise shape and effective aspect ratio of the Hall bar (around 10 %); the error in the resistance measurement is negligible in comparison.

susceptible to the presence of an electrostatic potential and can be extracted from the period of the low-field Shubnikov–de Haas-oscillations directly. In contrast, the resistivity depends on the position of the Fermi energy with respect to the minigap as well as on the carrier concentrations and mobilities.

To investigate the possibility of an *average* effect of the surface modification—such as a net change in the Fermi level pinning or reduced mobilities resulting from increased random scattering—I have collected the number of samples showing a specific qualitative change in the observed quantities in Table 7.2. Changes in the electron concentration, the zero field sheet resistivity, and the amplitude of the Shubnikov–de Haas-oscillations are listed for various groups of experiments. These results confirm the impression given by the data of Table 7.1 in that they do not support any hypothetical trends.

7.4.3 InAs Surface

It is well known that in many circumstances a surface accumulation layer forms at InAs surfaces covered with native oxide. We have created a InAs–GaSb single heterostructure with an exposed InAs top surface (0x4338) by omitting the GaSb cap layer from the MOVPE growth process described in Chapter 5. In such a structure, a sheet of mobile electrons appears that is confined to the InAs layer. Since it is located directly at the surface, this quasi-two-dimensional electron gas (2DEG) may be locally depleted by a shallow surface modification technique such as LAO. Indeed, this has been demonstrated by SASA *et al.* [32, 33] in earlier work on samples of a similar type.

Magnetotransport traces of sample 0x4338 showed spin-split Shubnikov–de Haas-oscillation peaks and no strong positive magnetoresistance or curvature in the Hall resistance, as expected for a 2DEG. The two-dimensional electron density was $13{\cdot}5 \pm 0{\cdot}2 \times 10^{11}\,\mathrm{cm}^{-2}$, while the mobility was $18{,}000 \pm 1{,}000\,\mathrm{cm}^2/(\mathrm{V\,s})$. From this substrate, I prepared and measured at 4·2 K two samples containing square antidot patterns with a lattice period of 200 nm following the approach detailed in Sec. 7.4.2. The magnetoresistance of both the patterned

(a) Carrier concentration

	$n_e^{(pat)} > n_e^{(ctrl)}$	$n_e^{(pat)} \approx n_e^{(ctrl)}$	$n_e^{(pat)} < n_e^{(ctrl)}$
$t_{cap} \leq 700\,Å$	2	9	5
$t_{cap} \geq 800\,Å$	5	3	2
$a \leq 200\,nm$	4	8	4
$a = 400\,nm$	3	4	3
$T = 0{\cdot}5\,K$	1	2	1
$T = 4{\cdot}2\,K$	6	8	3
$T \geq 10\,K$	0	1	3
all samples	7	12	7

(b) Resistivity

	$\rho_{xx}^{(pat)} > \rho_{xx}^{(ctrl)}$	$\rho_{xx}^{(pat)} \approx \rho_{xx}^{(ctrl)}$	$\rho_{xx}^{(pat)} < \rho_{xx}^{(ctrl)}$
$t_{cap} \leq 700\,Å$	8	9	5
$t_{cap} \geq 800\,Å$	3	2	5
$a \leq 200\,nm$	8	3	6
$a = 400\,nm$	2	7	2
$T = 0{\cdot}5\,K$	2	2	0
$T = 4{\cdot}2\,K$	7	4	6
$T \geq 10\,K$	0	3	3
all samples	10	9	9

(c) Amplitude of Shubnikov–de Haas-oscillations

	$A_{SdH}^{(pat)} > A_{SdH}^{(ctrl)}$	$A_{SdH}^{(pat)} \approx A_{SdH}^{(ctrl)}$	$A_{SdH}^{(pat)} < A_{SdH}^{(ctrl)}$
$t_{cap} \leq 700\,Å$	7	6	3
$t_{cap} \geq 800\,Å$	5	0	5
$a \leq 200\,nm$	9	2	5
$a = 400\,nm$	2	5	3
$T = 0{\cdot}5\,K$	1	1	2
$T = 4{\cdot}2\,K$	7	4	6
$T \geq 10\,K$	2	2	0
all samples	12	6	8

Table 7.2: Trends in LAO-modified samples, in numbers of samples satisfying the given conditions. Differences as small as $10^{10}/cm^2$ and $1\,\Omega/\square$ have been taken into account, as small changes in the period of the Shubnikov–de Haas-oscillations can be deduced qualitatively by comparing plots and the effective aspect ratio of two identically manufactured Hall bars on the same sample is consistent. If conservative estimates for the absolute errors are used, there is *no* significant change in any of the experiments.

and the control area exhibited a small (6 ‰) peak at $B_z = 0$ and reproducible fluctuations. However, no correlation between such fluctuations and the presence or absence of the lateral dot superlattice was observed.

7.5 Samples Created by Reactive Ion Etching

7.5.1 Overview

Thanks to the help of Geb Jones, who performed the electron beam exposure in Cambridge, I was also able to create a limited number of deep RIE etched samples which improve on the resolution previously attained for such structures in our group in Oxford. A number of square antidot patterns with a nominal lattice constant between 80 and 400 nm were defined on Hall bars prepared from different substrates and transferred into the semiconductors by means of plasma etching; the exact fabrication method is laid down in Chapter 4. Owing to partial underexposure and peeling of the masks in the RIE, the pattern transfer regrettably did not succeed for the smaller pitch patterns. In the following I shall describe two samples with a period of 400 nm and a physical antidot diameter d/a of $0{\cdot}25 \pm 0{\cdot}05$ determined by AFM measurements, which exhibit notable magnetoresistance features.

Sample OX4531R was created from the DHET OX4531 with a 300 Å InAs well and a 500 Å GaSb cap. The masked sample was placed into a plasma etcher for a sufficient time to remove approximately 140 nm of the semiconductor as estimated from the previously determined etch rates. The pattern transfer was verified with the help of an AFM, establishing a minimum hole etch depth of 30 ± 5 nm and confirming the larger estimated etch depth at the edges of macroscopic mask features. While the observed antidot depth was affected by finite tip size effects, there was considerable evidence from other samples of the same series that the etch depth of holes was less than that of larger features; the difference was attributed to underexposure resulting in incomplete development of the antidots. As a consequence, an

uncertainty remains regarding the actual etch depth, which must lie between the given limits. After processing, the unpatterned control area at 4·2 K had a sheet resistivity of $71 \pm 7\,\Omega/\square$ and an electron concentration determined from the period of the Shubnikov–de Haas-oscillations as $9{\cdot}8 \pm 0{\cdot}2 \times 10^{11}\,\mathrm{cm}^{-2}$. In the presence of a magnetic field $B_\parallel = 18\,\mathrm{T}$ parallel to the surface of the sample, the resistivity dropped to approximately 80 % of its zero field value. Using the high parallel field value, an electron mobility of at least $120{,}000\,\mathrm{cm}^2/(\mathrm{V\,s})$ and a mean free path of at least 2·0 µm were estimated. The electron concentration was significantly increased compared to samples with a similar InAs thickness that were known to have comparable electron and hole concentrations [34]. None the less, the positive perpendicular field magnetoresistance indicated the presence of mobile electron holes; comparing the curvature of the magnetoresistance to the Drude model of Sec. 5.6.1 led to an effective[9] hole concentration of $1{\cdot}6 \times 10^{11}\,\mathrm{cm}^{-2}$.

Sample 0X4532′R was based on the substrate 0X4532 with a 300 Å InAs well and a 680 Å GaSb cap. It was initially etched under the same conditions as 0X4531; there was no significant change in the magnetoresistance as a result of the pattern transfer. The sample was then subjected to another etch step designed to remove an additional 30 nm of the semiconductor; in the absence of an etch mask, one expects the same amount of material to be removed from the flat surface as from the bottom of the antidot pits. Further observation in the AFM confirmed that the antidot pattern was still well resolved. At 4·2 K, the unpatterned control area of the 380 Å cap sample 0X4532′ created by this procedure had a sheet resistivity of $82 \pm 8\,\Omega/\square$ and an electron concentration of $12{\cdot}0 \pm 0{\cdot}2 \times 10^{11}\,\mathrm{cm}^{-2}$. In the presence of a parallel magnetic field $B_\parallel = 18\,\mathrm{T}$, the resistivity dropped to approximately 73 % of its zero field value. The electron mobility and the mean free path estimated from the high parallel field value were at least $85{,}000\,\mathrm{cm}^2/(\mathrm{V\,s})$ and 1·5 µm, respectively. Despite the higher electron density reflecting the reduced cap thickness, the DHET still behaved as a two-carrier system; the effective hole concentration from the Drude model was $1{\cdot}6 \times 10^{11}\,\mathrm{cm}^{-2}$.

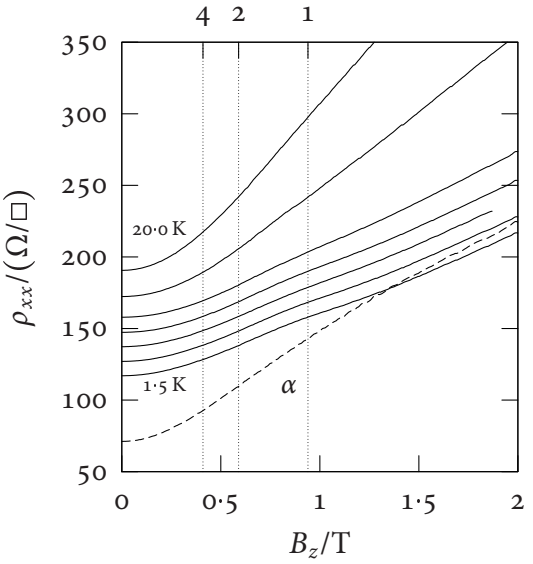
[9] As the minigap is not taken into account, this value does not necessarily agree with the actual hole density.

7.5.2 Low Field Commensurability Peaks

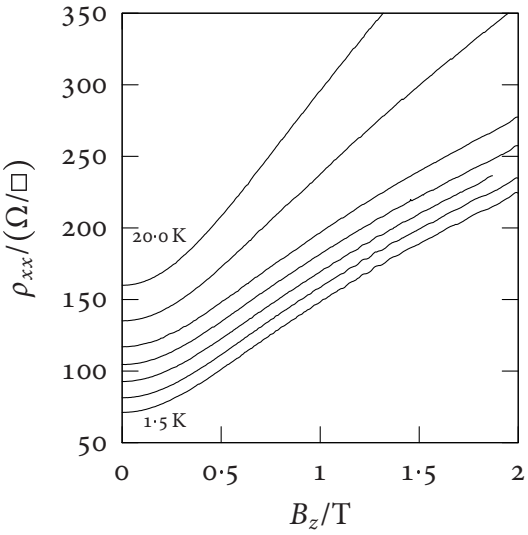
Temperature Dependence

The low field magnetoresistivity of sample 0x4531r is shown in Fig. 7.8 for a number of temperatures between 0·5 and 20·0 K. Compared to the control area covered in Fig. 7.8(b), the modulated DHET of Fig. 7.8(a) had a significantly increased zero field resistivity (117 Ω/□ compared to 71 Ω/□). The Shubnikov–de Haas-oscillation amplitude was reduced and the electron concentration increased from $9{\cdot}8 \pm 0{\cdot}2 \times 10^{11}\,\mathrm{cm}^{-2}$ to $13{\cdot}0 \pm 0{\cdot}5 \times 10^{11}\,\mathrm{cm}^{-2}$. The shape of $\rho_{xx}(B_z)$ as a function of the magnetic field differed substantially, being flatter around the origin and exhibiting a single broad peak (α) around $B_z \approx 0{\cdot}8$ T for temperatures up to 10 K. The peak is more readily seen in Fig. 7.8(c), which displays the magnetoresistivity after background subtraction. Its position is consistent with the commensurability between the electron cyclotron orbit of a nearly-free electron gas with the observed electron concentration and the period of the artificial superlattice. In the pinned orbit model, this corresponds to an orbit trapped at a single antidot site; at the magnetic field values corresponding to similar orbits encircling larger groups of antidots, no features were observed in the magnetoresistance.

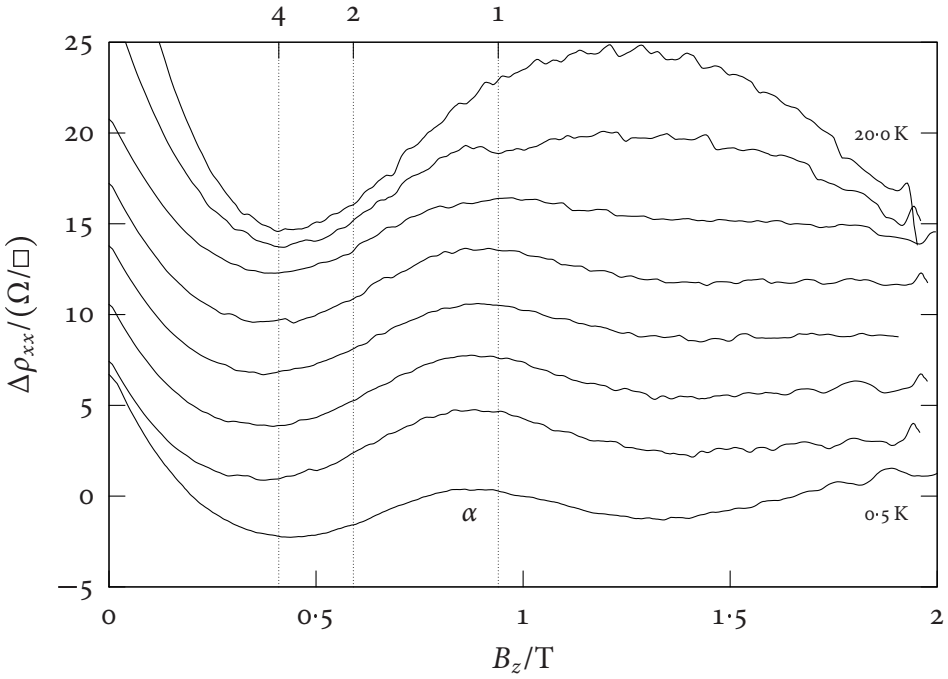
The corresponding data for 0x4532′r is presented in Fig. 7.9. Comparing patterned and control Hall bars, the electron concentration increased from $12{\cdot}0 \pm 0{\cdot}2 \times 10^{11}\,\mathrm{cm}^{-2}$ to $13{\cdot}4 \pm 0{\cdot}5 \times 10^{11}\,\mathrm{cm}^{-2}$ and the amplitude of the Shubnikov–de Haas-oscillations was reduced. In this sample, the absolute zero field resistivity was similar for the patterned (74 Ω/□ at 4·2 K) and the control region (82 Ω/□ at 4·2 K). Even so, the magnetoresistivity changed as a result of the modulation in a way similar to that seen in 0x4531r, becoming significantly flatter around zero field and exhibiting additional features absent from the control measurements of Fig. 7.9(b). A more complicated peak structure emerged, with a main peak (α) persisting up to 10 K at approximately 0·7 T and an additional higher field peak (β) disappearing at 4·2 K around 2 T. A third peak (γ) might tentatively be assigned to the area between those two features, which exhibited a small resistivity rise at low temperatures. The peaks do not



(a) Patterned area. The dashed line corresponds to the unmodified DHET at 1·5 K.

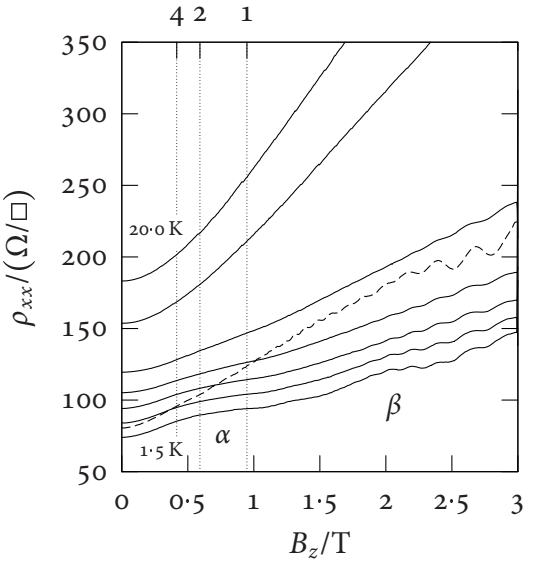


(b) Control area

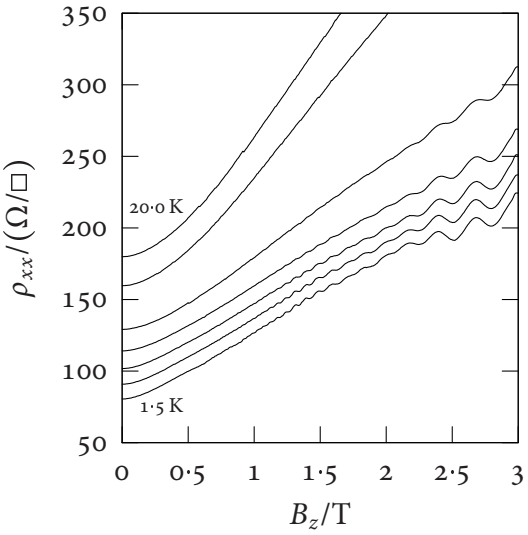


(c) Patterned area after background subtraction

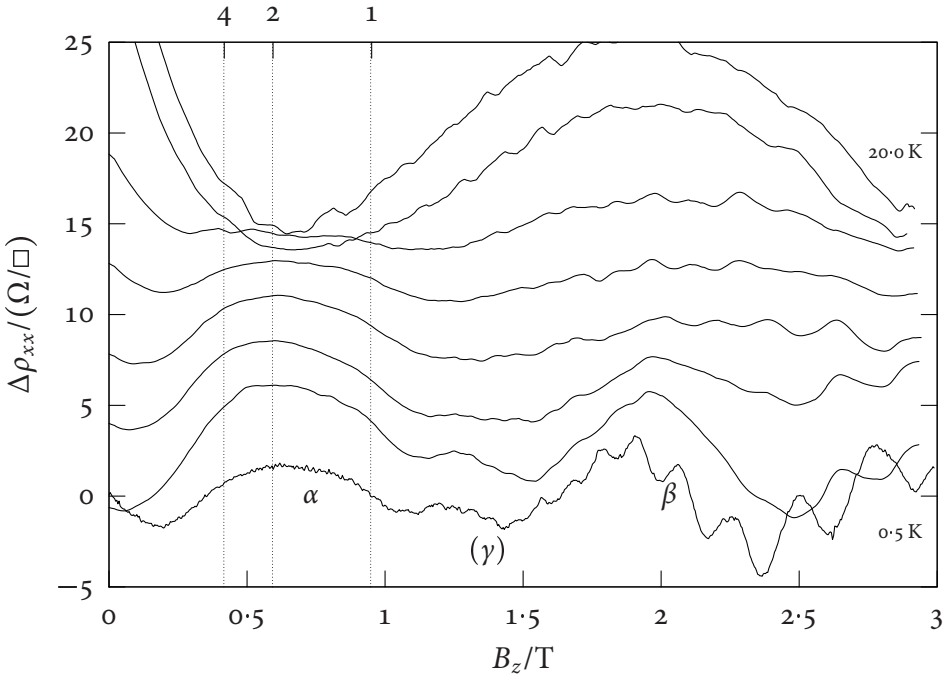
Figure 7.8: Temperature dependence of the low field magnetoresistivity of sample OX4531R. The symmetric part of the measured longitudinal resistivity is shown for (from bottom to top) 0·5, 1·5, 2·5, 4·2, 7·0, 10·0, 15·0, and 20·0 K; the individual curves are offset by 10 Ω/□ in (a) and (b) and 3 Ω/□ in (c) for clarity. Dotted vertical lines show the commensurability condition for pinned electron orbits around 1, 2, and 4 antidots calculated using an electron concentration of $13{\cdot}0 \times 10^{11}\,\mathrm{cm}^{-2}$.



(a) Patterned area. The dashed line corresponds to the unmodified DHET at 1·5 K.



(b) Control area



(c) Patterned area after background subtraction

Figure 7.9: Temperature dependence of the low field magnetoresistivity of sample OX4532′R. The symmetric part of the measured longitudinal resistivity is shown for (from bottom to top) 0·5, 1·5, 2·5, 4·2, 7·0, 10·0, 15·0, and 20·0 K; the individual curves are offset by 10 Ω/□ in (a) and (b) and 3 Ω/□ in (c) for clarity. Dotted vertical lines show the commensurability condition for pinned electron orbits around 1, 2, and 4 antidots calculated using an electron concentration of $13{\cdot}2 \times 10^{11}$ cm^{-2}.

Sample	$n_e^{(SdH)}$ ($10^{11}/cm^2$)	R_H (Ω/T)	$n_e^{(R_H)}$ ($10^{11}/cm^2$)
OX4531R	13·0	411	14·2
OX4531 (control)	9·8	630	9·9
OX4532′R	13·4	356	17·1
OX4532′ (control)	12·0	462	13·5

Table 7.3: Hall effect in OX4531R and OX4532′R

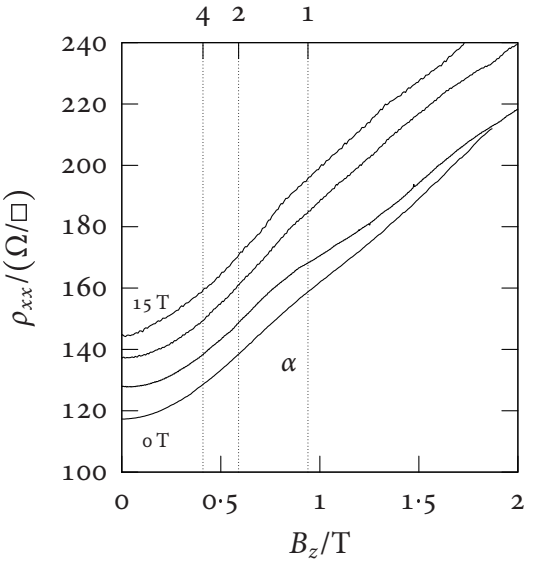
directly correspond to commensurability conditions for the electron cyclotron radius R_c, although the primary peak (α) occurs in the magnetic field range where commensurability features are expected. At the position of peak (β), however, $2R_c$ is considerably smaller than the superlattice spacing.

Neither OX4531R nor OX4532′R exhibited any significant features in the Hall resistivity $\rho_{xy}(B_z)$ at the position of the peaks in $\rho_{xx}(B_z)$. The effective Hall coefficient near $B_z = 0$ differed between the modulated and the unmodulated Hall bars. As shown in Table 7.3, the changes were qualitatively consistent with the difference in the electron concentrations determined from the period of the Shubnikov–de Haas-oscillations. Generally, the calculation from the Hall coefficient led to a higher estimate of the electron density; this effect was more prominent for the modulated DHETs and in the case of sample OX4532′.

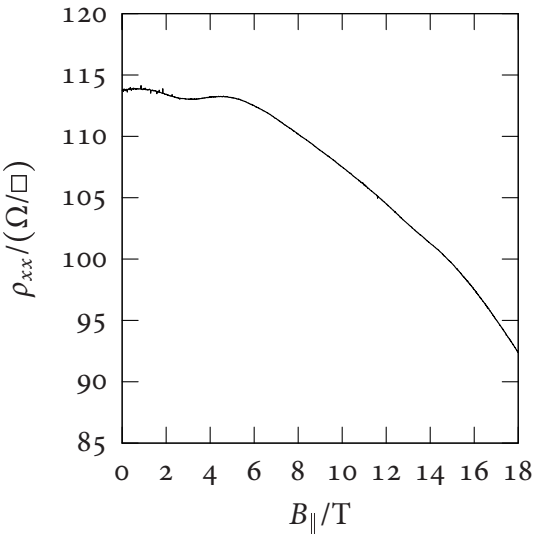
Parallel Field Dependence

As the presence of a magnetic field in the plane of the DHET affects the formation of the minigap resulting from the mixing of the electron and hole states (*cf.* Sec. 5.5.3) and can influence the commensurability peaks seen in some modulated samples (*cf.* Sec. 7.3.2), I measured the B_z-magnetoresistivity of OX4531 and OX4532′ in various parallel magnetic fields $B_\parallel$ up to 15 T as well as the $B_\parallel$-magnetoresistivity at $B_z \approx 0$. In these experiments, $B_\parallel$ was parallel to the direction of the excitation current and the rotating setup described in Sec. 4.3.1 was used.

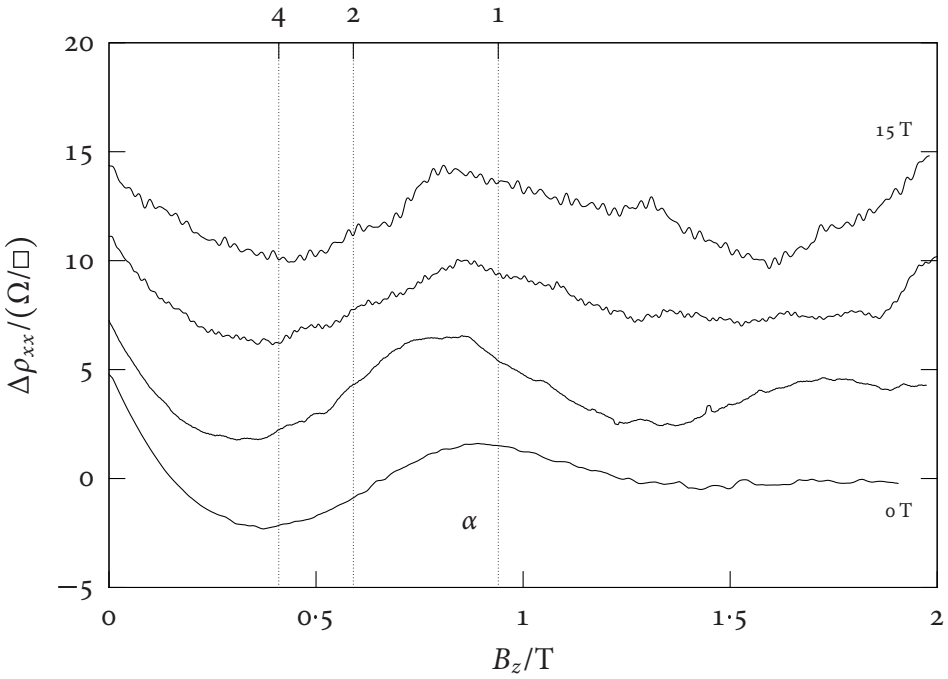
The data for sample OX4531R appears in Fig. 7.10. The overall shape of the magnetores-



(a) Resistivity as a function of B_z



(b) Resistivity at $B_z = 0$ as a function of $B_\parallel$



(c) Resistivity as a function of B_z after background subtraction

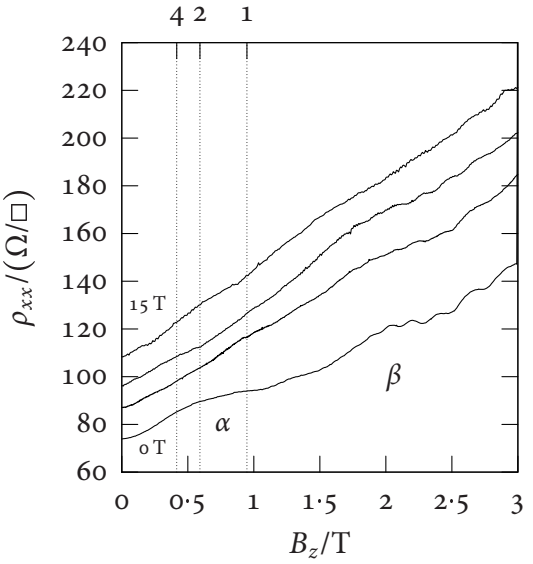
Figure 7.10: Magnetoresistivity of sample 0x4531R in a parallel field. The longitudinal resistivity measured at 4·2 K while rotating the sample towards $B_\parallel = 0$ is shown for (from bottom to top) $B_\parallel = 0$, 5, 10, and 15 T; the individual plots are shifted by 15 Ω/□ in (a) and 4 Ω□ in (c) for clarity.

istivity, which is seen in Fig. 7.10(a) remains similar. The position and relative magnitude of the commensurability peak (α), best seen after background subtraction in Fig. 7.10(c), did not change significantly in the parallel field. The shape of the peak appeared different, but this may be at least partly attributable to the additional noise appearing in the parallel field measurement as a result of the mechanical movement of the sample. Fig. 7.10(b) shows the resistivity of the modulated Hall bar as a function of the parallel field. There was a significant negative magnetoresistance, with $\rho_{xx}(B_\parallel)$ dropping to 81 % of its zero field value at 18 T; at approximately 5 T it exhibited a local maximum. It is worth noticing that the resistivity did not plateau at the largest observed value of $B_\parallel$ but continued to decrease.

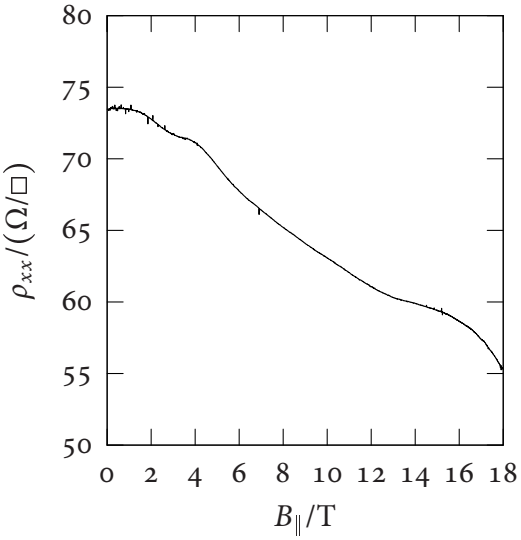
The situation was more complicated in the case of OX4532′R, as can be seen in Fig. 7.11. The well-defined low field peak (α) did not persist in the presence of an in-plane field; at 5 T and 10 T, a small peak at $B_z \approx 1$ T was seen instead, while a broad peak similar to the one seen at $B_\parallel = 0$ T reappeared at 15 T. However, these features were not discernible in all traces. In contrast, the higher field peak (β) maintained a similar size for all values of the parallel field but moved to continuously lower B_z as $B_\parallel$ increased. The small feature (γ) around $B_z \approx 1{\cdot}3$ T persisted with reduced visibility. While its magnitude only barely exceeded the detection threshold given by the noise level on the one hand and the scale of repeatable fluctuations seen in all DHETs on the other hand, it was observable in most low temperature traces. The $B_\parallel$-magnetoresistivity shown in Fig. 7.11(b) resembled that of OX4531R, decreasing continuously with increasing $B_\parallel$. At $B_\parallel = 18$ T, it dropped to approximately 75 % of its zero field value, but did not reach a plateau.

7.5.3 Behaviour in a High Perpendicular Field

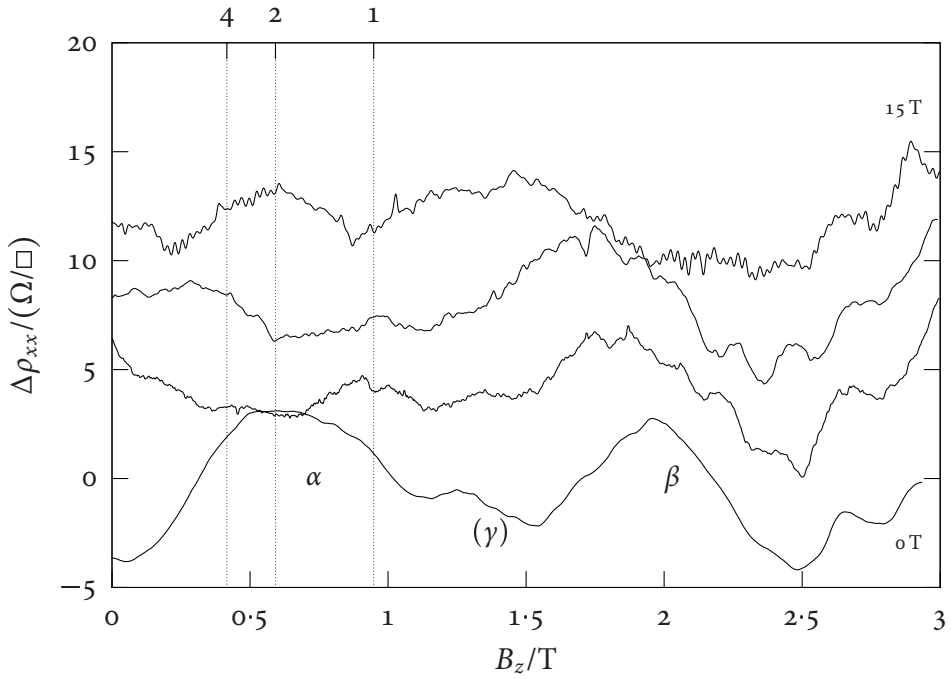
Figs. 7.12 and 7.13 show $\rho_{xx}(B_z)$ and $\rho_{xy}(B_z)$ up to $B_z = 18$ T for both the patterned and unpatterned areas of OX4531R and OX4532′R. In both samples, the Shubnikov–de Haas-oscillations in $\rho_{xx}(B_z)$ and the steps in $\rho_{xy}(B_z)$, which mark the onset of Hall plateaux, were less sharply resolved in the modulated regions, while their period in $1/B_z$ was increased.



(a) Resistivity as a function of B_z

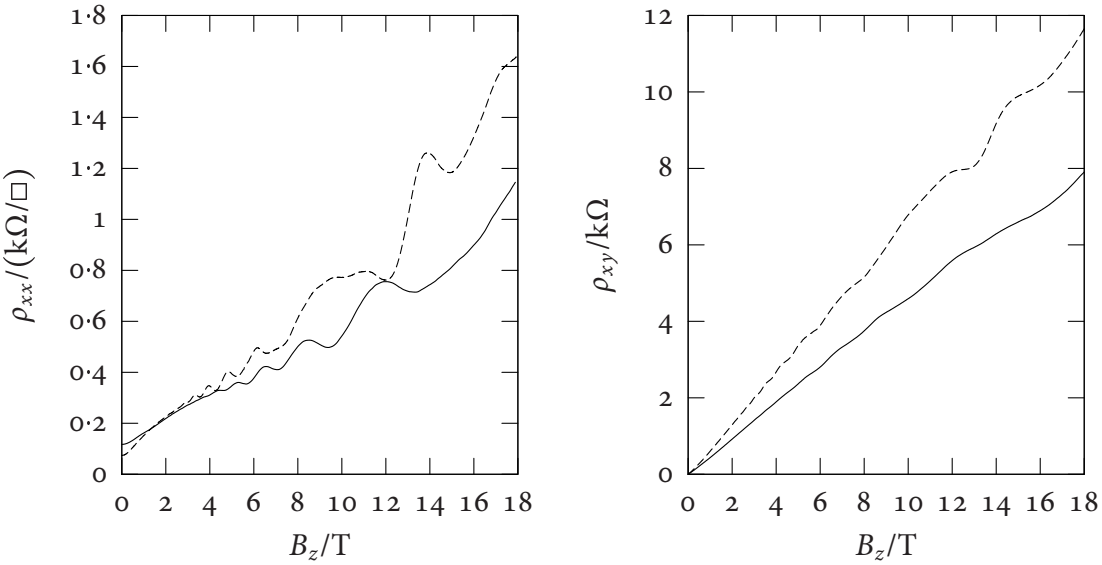


(b) Resistivity at $B_z = 0$ as a function of $B_\parallel$



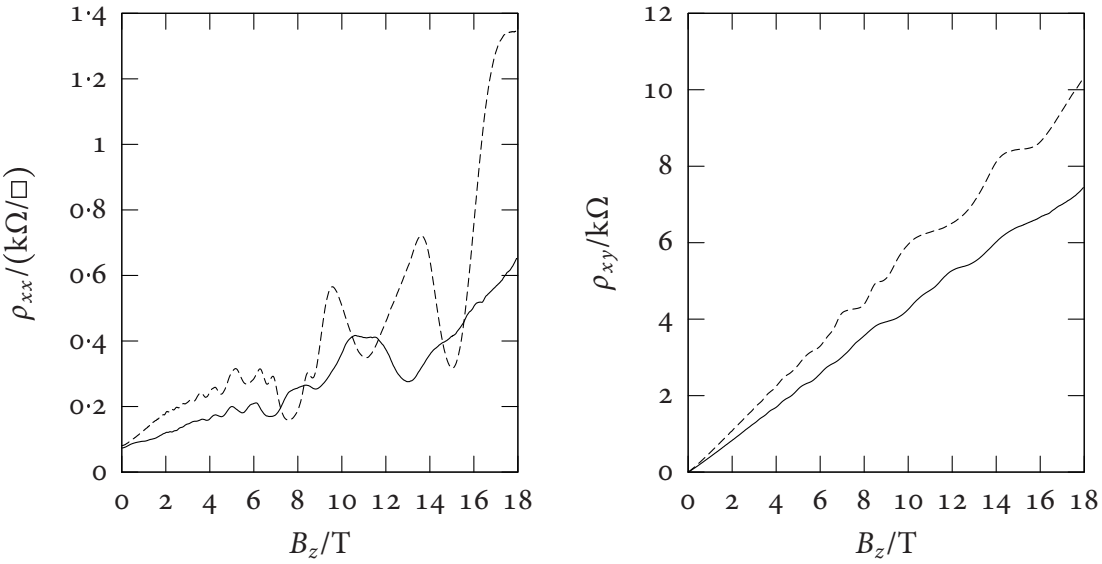
(c) Resistivity as a function of B_z after background subtraction

Figure 7.11: Magnetoresistivity of 0x4532′R in a parallel field. The longitudinal resistivity measured at 4·2 K while rotating the sample towards $B_\parallel = 0$ is shown for (from bottom to top) $B_\parallel = 0$, 5, 10, and 15 T; the individual plots are shifted by 15 Ω/□ in (a) and 4 Ω□ in (c) for clarity.



(a) Longitudinal resistivity (b) Hall resistivity

Figure 7.12: High field magnetoresistivity of sample OX4531R. The dashed curves show the data for the unmodified DHET.



(a) Longitudinal resistivity (b) Hall resistivity

Figure 7.13: High field magnetoresistivity of sample OX4532′R at 0·5 K. The dashed curves show the data for the unpatterned control area.

The Hall resistivity, with the exception of these quantization features, was nearly linear over the entire field range; its slope was reduced as an effect of the antidot lattice. Apart from these observation, the high field measurements taken on the patterned Hall bars did not show qualitatively different behaviour. The Hall resistivity showed a monotonic step-like behaviour without any indication of local minima, as expected from an electron-rich DHET.[10]

7.6 Discussion

7.6.1 General Remarks

In the LAO-patterned samples, if taken collectively, there is no systematic change in the magnetotransport properties that can unambiguously be attributed to the imposed lateral superlattice. Since the DHETs used for these experiments are similar or identical to ones which do show such alterations if modified by a different method, and the inferred mean free paths are considerably longer than the attempted modulation periods, it must be concluded that the effective potential at the carrier sheets in such samples is too weak to bring about a measurable effect. In contrast, the samples prepared by masking and etching show peaks in the magnetoresistance at the approximate field values where the size of the electron cyclotron orbit is commensurate with a dimension of the superlattice. The samples OX4531R and OX4532′R only show a single commensurability peak, and in most experiments the agreement between observed and predicted peak position is not exact.

Several factors contribute to this discrepancy. First of all, since the background that has been subtracted to compensate for the strong positive magnetoresistance is not known *a priori* there is an inherent uncertainty in the measured peak position. The prediction to which the position is compared depends on the chosen analytical model and the measured electron concentration.

[10] See also the discussion in Sec. 6.4.1.

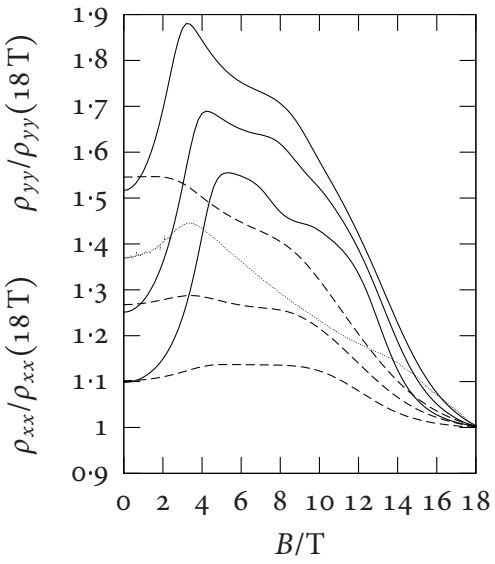
The model itself makes some simplifications, and although the commensurability conditions deduced from the pinned orbit picture agree well with a large range of experimental and simulated magnetoresistance traces, more realistic calculations do not always show an exact agreement. In general, the positions of the peaks relative to the field at which $2R_c = a$ change by a small amount with d/a; calculations based on simulated trajectories suggest a change of approximately 3 % between $d/a = 0{\cdot}2$ and $d/a = 0{\cdot}4$. The position will also depend on the precise shapes of the modulation potential and the electron dispersion.

The electron concentration is determined from the period of the low-field Shubnikov–de Haas-oscillations arising from the same magnetic breakdown orbits as the commensurability features; the calculation is quite reliable as it does not depend on the knowledge of other sample parameters. However, the field range that can be used is limited to the region between the onset of detectable oscillations at 0·4 to 1 T (depending on the electron mobility and the temperature) and the end of the validity of the description in terms of Shubnikov–de Haas-oscillations corresponding to electron orbits at 1 to 4 T (depending on the electron–hole ratio) [34]. In practice, the electron concentrations determined using Eq. (5.32) have an uncertainty between 2 to 8 % depending on the amplitude of the oscillations

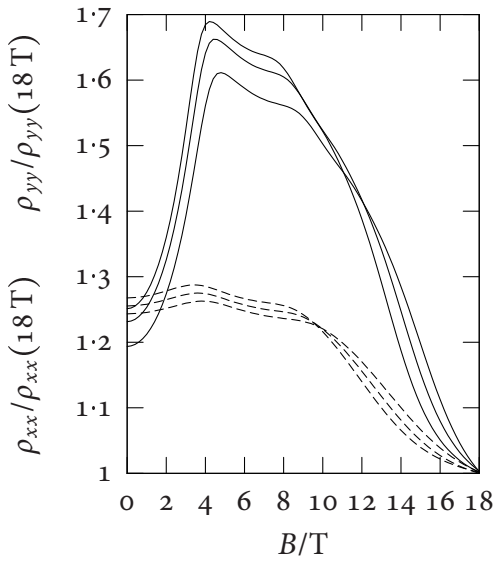
More importantly, the calculation of the commensurability condition assumes a spatially uniform electron density, which may not accurately reflect the situation between the antidots in the patterned samples. Following the considerations put forward in Sec. 7.6.3 below, an upper bound on the variation of the electron densitiy in the RIE-etched samples is given by the magnitude of the total increase of 12 to 33 % caused by the patterning.

The parallel field magnetoresistance of OX4531R and OX4532′R, shown in Figs. 7.10(b) and 7.11(b), respectively, differs considerably from thicker cap DHETS such as those discussed in Sec. 7.3, which show a behaviour close to the simulated trace of Fig. 5.8 in Chapter 5. While there is still a significant negative magnetoresistance, indicating the presence of a mixing gap close to the Fermi energy,[11] the resistance drops by a smaller fraction in the region up

[11] For finite temperatures and interaction strengths, there is still an effect even if $E_F - E_0$ exceeds the band offset, so the negative magnetoresistance does not necessarily prove the presence of mobile holes.



(a) $\Delta = 7\,m\mathrm{eV}$ (bottom), 11 meV, 15 meV (top) for $\delta z = 18\,\mathrm{nm}$



(b) $\delta z = 16\,\mathrm{nm}$ (bottom), 17 nm, 18 nm (top) for $\Delta = 11\,m\mathrm{eV}$

Figure 7.14: Magnetoresistance of thin-cap samples in a parallel magnetic field, calculated at $T = 4{\cdot}2\,\mathrm{K}$ for a DHET with $E_\mathrm{g} = 80\,m\mathrm{eV}$ and $E_\mathrm{F} - E_0 = 78\,m\mathrm{eV}$. The resistivity ρ_{xx} parallel to the in-plane field is shown by solid lines, whereas ρ_{yy} is shown with dashed lines. The single dotted line is the experimental ρ_{xx} of the control area of OX4532′R.

to $B_\parallel = 18\,\mathrm{T}$ and continues to decrease instead of saturating. The electron concentration of $9{\cdot}8 \times 10^{11}\,\mathrm{cm}^{-2}$ in the unpatterned region of sample OX4531R indicates a Fermi level 78 meV above the bottom of the electron band. To match a hole concentration of $1{\cdot}6 \times 10^{11}\,\mathrm{cm}^{-2}$ at the same time, the band offset has to be 80 meV. Using these values, I have calculated several theoretical magnetoresistance traces in Fig. 7.14, varying the electron-hole coupling Δ and the electron-hole separation δz. The larger band offset causes the effect of the minigap to extend over a larger field region compared to Fig. 5.8; for larger interaction strengths and smaller electron–hole separations the field at which the resistivity saturates moves to even higher values. The calculations predict a strong initial increase in the resistivity ρ_{xx} parallel to the in-plane field as the hole Fermi contour approaches the electron Fermi contour from the inside. This increase is small or absent in the experimental curves, which resemble more closely the calculations for ρ_{yy}. The model may not accurately reflect the situation in the

DHET, or the peak may be smeared out by misalignment between current and magnetic field or by local changes in the electron density.

The 2DEG forming in the InAs layer at the surface of the single heterostructure OX4338 has an electron mobility μ_e much lower than that seen in a comparable DHET: as it is located directly at the surface, the electron sheet is very susceptible to scattering from imperfections at the interface. Using the measured values for μ_e and the sheet density n_e, one calculates a mean free path ℓ_f of 350 nm, which is only slightly larger than the modulation period of 200 nm. Under these circumstances, the absence of commensurability features should not come as a surprise. While there is every reason to assume that an antidot potential was created in the 2DEG, the electrons will on average be scattered out of a cyclotron orbit before they can complete it, rendering the pinning mechanism ineffective. It became quickly clear that progress could only be made by procuring higher mobility heterostructures. As similar antidot lattices in pure 2DEGs had been studied in considerable detail and LAO of InAs films had already been demonstrated, I did not consider pursuing this idea a high priority.

7.6.2 Holes

Because of the presence of mobile holes and the peculiar band arrangement, the Fermi contour in InAs–GaSb DHETs has a complex shape resulting from the mixing of the electron and hole dispersion relations and a variety of orbits become possible.[12] So far the discussion has focused on nearly-free electron-like orbits, as they can successfully explain the low-field Shubnikov–de Haas-oscillations and the commensurability peaks; they are recovered in the electron–hole system if the Fermi level intersects the dispersion relation away from the minigap or via magnetic breakdown (*cf.* Sec. 5.5.4) if the cyclotron energy is comparable to the minigap. There is little direct evidence of cyclotron motion corresponding to other sections of the Fermi contour [34].

[12] See Sec. 5.5.3 for a discussion of the details.

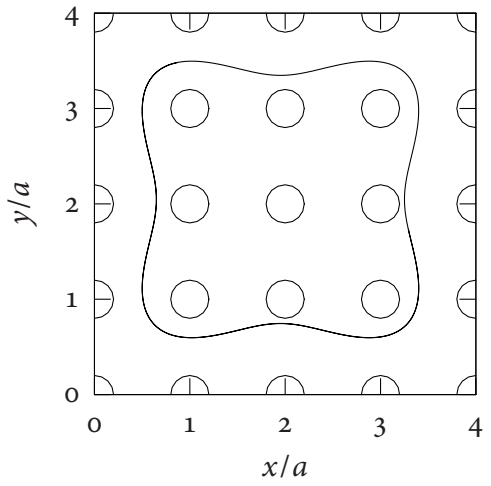
Purely hole-like trajectories can be considered directly in the numerical framework developed for electron cyclotron orbits by replacing the parabolic electron dispersion by an anisotropic relation reflecting the hole energy, such as the semi-empirical model introduced in Sec. 5.5.3.[13] As shown in Fig. 7.15, such calculations predict a peak in the hole component of the resistivity at the magnetic field where the orbit size approximately equals the superlattice spacing, similar to the situation in a 2DEG. The higher order peaks are strongly affected by the anisotropy. For the alignment assumed in Fig. 7.15, which to within a few degrees reflects that in the actual samples, the peak corresponding to orbits encircling 2 antidots is completely absent, whereas peaks corresponding to several larger orbits are clearly resolved.

The rough estimate of the hole concentration of $1{\cdot}6 \times 10^{11}\,\mathrm{cm}^{-2}$, obtained from the non-interacting Drude model for both 0X4531R and 0X4532′R, corresponds to $B_0 \approx 0{\cdot}33\,\mathrm{T}$. No feature is seen in the magnetoresistance traces around this field strength. However, the hole mean free path in DHETs of the type investigated here is sufficiently lower than the electron mean free path that the low B_z conductivity is dominated by electronic transport and there is little evidence of Shubnikov–de Haas-oscillations corresponding to cyclotron orbits other than the nearly-free-electron one. It is therefore not surprising that the effects of the commensurability of hole-like orbits with the antidot superlattice on the overall conductivity are not strong enough be measurable.

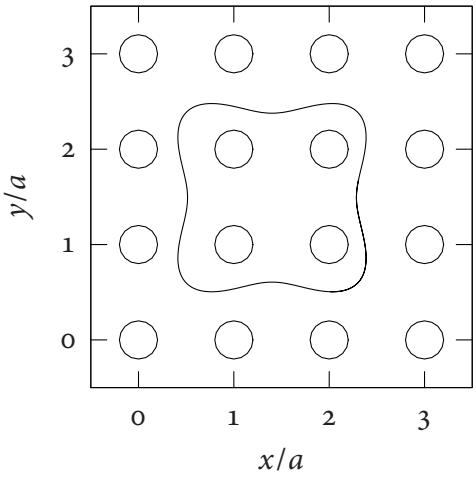
7.6.3 Surface States, Potential Strength, and Electron-Hole Interaction

In the case of the intermediate sample 0X4532R, it is safe to assume that the antidot etch depth lay between the lower bond of 30 nm established by AFM measurements and an upper bond of 68 nm given by the cap thickness, removing more than 45 % but less than 100 % of the GaSb cap. The sample as originally prepared did not exhibit any characteristic antidot effects, but further etching led to the appearance of such features. The modified sample 0X4532′R has a similar surface structure in the AFM measurement but a significantly increased electron

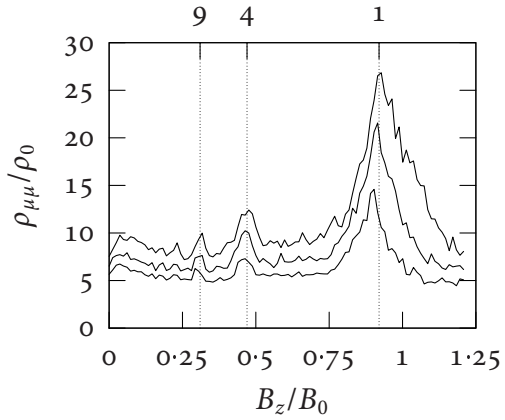
[13] See also Appendix C for the details of the calculation.



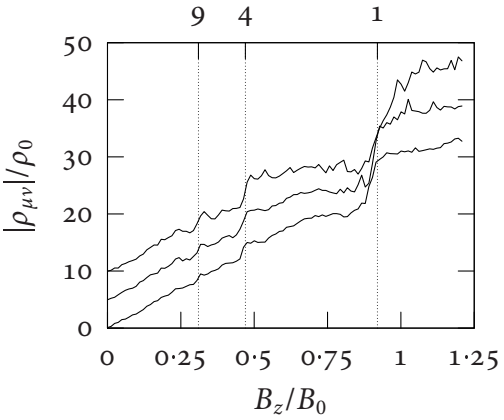
(a) A hole trajectory encircling nine antidots.



(b) A hole trajectory encircling four antidots.



(c) Diagonal components of the resistivity tensor as a function of B_z for $d/a = 0{\cdot}2$ (bottom), $0{\cdot}3$, and $0{\cdot}4$.



(d) Off-diagonal (Hall) components of the resistivity tensor as a function of B_z for $d/a = 0{\cdot}2$ (bottom), $0{\cdot}3$, and $0{\cdot}4$. The curves are offset by $5\rho_0$ for clarity.

Figure 7.15: Anisotropic hole gas resistivity from semiclassical simulations. It is assumed that the hole gas is parallel to the $\{100\}$ planes, while the superlattice is aligned to the $\langle 100\rangle$ family of crystallographic directions. B_0 is the field for which the cyclotron orbit area is πa^2 and $\rho_0 = (\pi\hbar^2)/(e^2\tau E_\mathrm{F})$.

density deduced from low temperature magnetotransport measurements, in agreement with the expected removal of a similar amount of material from *both* the original surface and the bottom of the etch pits. Based on these observations, the holes now remove at least 80 % of the reduced cap, but may extend through the InAs well. Sample OX4531R was etched under the same conditions; it is therefore plausible, but not inexorable, that a similar upper limit to the hole etch depth applies. In this case, at least 60 % of the cap have been removed, but the antidots may extend into the InAs layer.

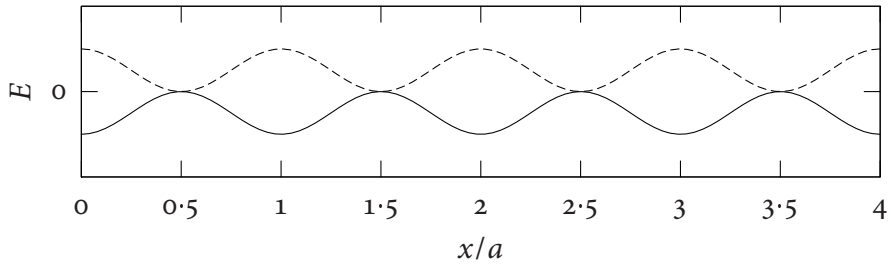
These results indicate that locally moving the GaSb surface towards the InAs well by as much as half the original distance in itself is insufficient to elicit a measurable effect on the electronic properties of the embedded DHET. This is true even if the bottom surfaces of the pits lie much closer to the carrier sheets than the lattice constant of the imposed pattern, and wholesale erosion of a comparable amount of the sample surface by the same chemical procedure has a demonstrable influence on the carrier densities. This agrees with the fact that the direct oxidation of the AFM, which, as a surface modification technique, produces similar or shallower features, fails to create a sufficiently strong modulation of the carrier sheets to induce a significant change. On the other hand, it at first seems to be at odds with the behaviour of shallow wet etched samples reported in 7.3.1, despite the fact that these were created from substrates with thicker caps and the potential modulation set up by the surface states was further away from the InAs well.

While the larger dot diameter of the samples created by the nearly isotropic wet chemical etch as well as the larger superlattice constant may play a rôle, the probable explanation of the contradiction is a different one. Whereas the influence of the surface states on the carrier gases is well understood in general and there is considerable evidence supporting the assumption that the Fermi level at oxide passivated GaSb surfaces is pinned above its bulk value [34–36], the precise level of the pinning is expected to be sensitive to the kind of oxide and the presence of contaminations, which in turn can and will vary depending on the fabrication history. This can for example be seen in the data for OX3730 presented in Table 7.1: After removal of the oxide and subsequent reoxidation the electron density is

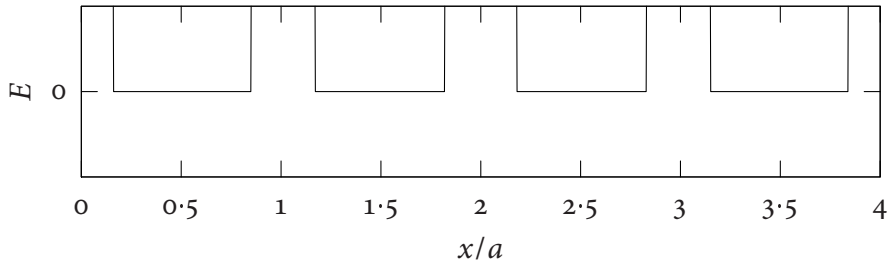
decreased, although the procedure *reduces* the cap thickness by a small amount.

Both 0X4531R and 0X4532′R show a significant increase of 23 % and 12 %, respectively, in the two-dimensional electron density in the patterned versus the control Hall bars. In contrast, the complete depletion of antidots with $d = a/3$ would lead to a *decrease* of the carrier concentration of 9 %. Since the processing is otherwise identical, the uncertainty associated with the incomplete knowledge of the exact surface states can be largely avoided, so that the antidot pattern must be directly responsible for the effect. The likely mechanism by which it is caused is the Fermi level pinning at the exposed surface of the etch pits. This may be the bottom surface if the pits do not extend into the InAs, but also the side walls. In either case, a significant smooth electric potential modulation is present, either exclusively or in addition to the steep antidot potential, leading to a more complicated effective potential landscape. Fig. 7.16 illustrates schematically the effect of different modulations on the effective potential experienced by both electrons and holes.

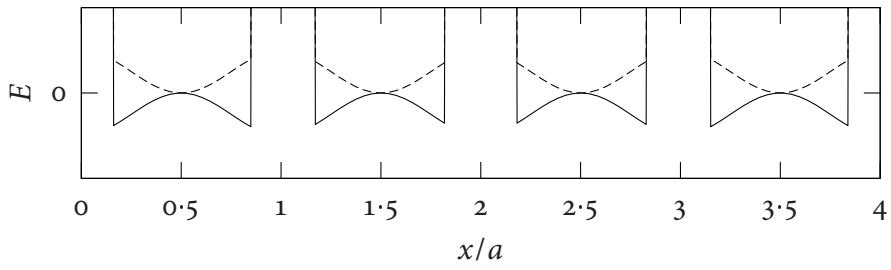
As noted in Sec. 7.3.2, the commensurability peaks in shallow wet etched samples disappear in the presence of a magnetic field parallel to the surface. A similar observation can be made for the fundamental peak in 0X4532′R, although the situation at intermediate fields is more difficult to interpret and there is evidence that the peak re-emerges at the highest fields investigated. This behaviour was explained by different potentials for electrons and holes, corresponding to the situation of Fig. 7.16(a). The potential experienced by the electrons would by itself be insufficient to produce strong magnetoresistance peaks, which would be prompted by the local depletion of the interacting hole gas; if the interaction was removed by the parallel field (*cf.* Sec. 5.5.3), the peaks would disappear. Both in these samples and in 0X4532′R, the features disappear in a considerably smaller field than required to fully decouple the electrons and holes; moreover, the exact mechanism by which the electron-hole interaction causes strong commensurability features absent from the non-interacting 2DEG has not been explained satisfactorily. In the light of these problems as well as the small hole concentration and the peak structure in the $B_{\parallel} = 15$ T trace, one cannot necessarily conclude that the parallel field behaviour of 0X4532′R implies a shallow etch depth and a strong contri-



(a) Electrostatic potential



(b) Antidots



(c) Mixed potential

Figure 7.16: Effective potentials for electrons (solid lines) and holes (dashed lines) along a two-dimensional cross-section through five antidots at $x = 0, a, \ldots, 4a$. The electrostatic potential is assumed to be of sinusoidal form and drop to zero between antidots.

bution of the electron–hole interaction to the observed structure in the magnetoresistivity. A different mechanism by which the parallel magnetic field may affect the size of the commensurability maxima is the size of the gaps in k-space through which carriers must tunnel to complete electron-like magnetic breakdown orbits. The cyclotron energy $\hbar\omega_c$ becomes similar to the minigap $\Delta \approx 7\,\mathrm{m}e\mathrm{V}$ around $B_z = 1\,\mathrm{T}$ and the feasibility of full electron orbits giving rise to the magnetoresistivity features can be expected to be sensitive to the gap size in this field region.

7.6.4 Peak Structure above the Highest Electron Commensurability Field

An interesting and unexpected feature of the magnetoresistance of the sample OX4532′R is the appearance at low temperatures of additional structure above the magnetic field value for which the diameter of the electron cyclotron orbit equals the superlattice spacing. As illustrated in Fig. 7.17, the additional peak is likely a result of the potential modulation; while the control sample shows a small dip on the high field side of the peak position which may also contribute to the measured magnetoresistivity of the patterned Hall bar, the latter exhibits structure that is clearly distinguishable before background subtraction and has no counterpart in the control measurement.

In the pinned orbit model (*cf.* Fig. 7.2), the proportion of pinned orbits decreases in this region until they become impossible where R_c falls below the antidot diameter d. In the region of the putative resistivity peak, all possible orbits are much smaller than a, and an explanation based on commensurability effects is problematic. In contrast, earlier experiments on antidot arrays in 2DEGs have revealed peaks in this field range in the special cases of large diameter [21] or asymmetric [22] antidots. The temperature dependence of these features follows that of the fundamental peak at $2R_c \approx a$.

The observations have been explained by the contribution of different classes of skipping orbits to the conductivity. Eroms *et al.* [21] have published semiclassical simulations following the *Ansatz* of Sec. 7.2.3 that predict a local magnetoresistance maximum around $2{\cdot}5B_0$

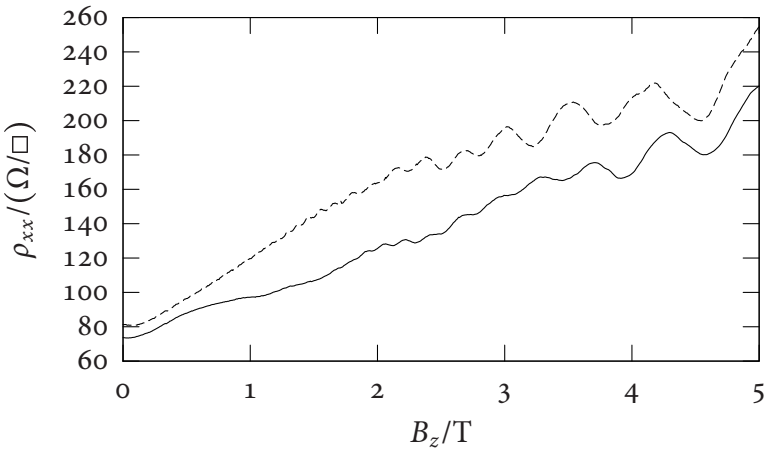
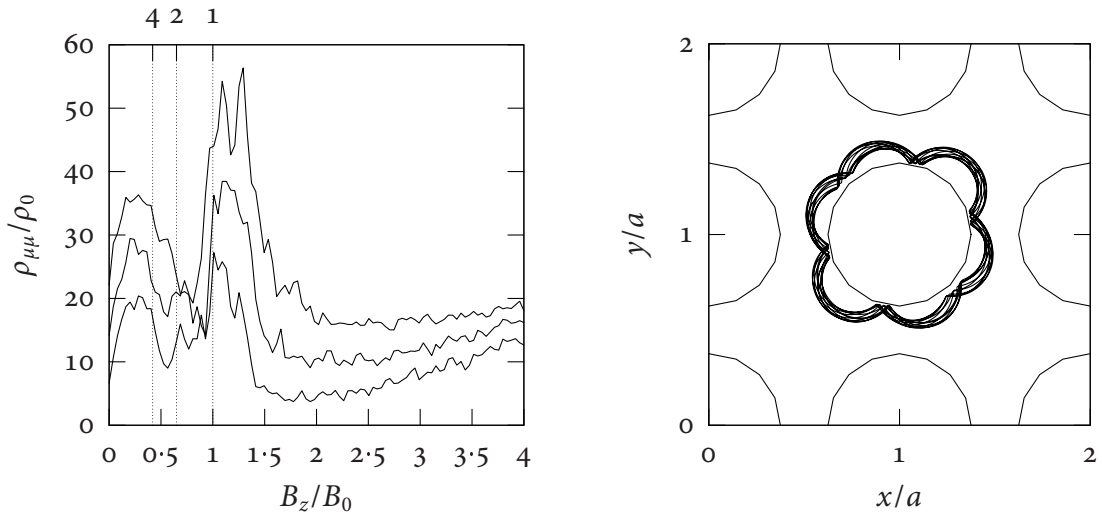


Figure 7.17: Comparison of the 1·5 K magnetoresistivity in sample OX4532′R. The dashed curves show the data for the unpatterned control area.

for large diameter antidots with $d/a \approx 0{\cdot}66$. It has been attributed to quasi-periodic rosette-shaped orbits trapped at single antidots as seen in Fig. 7.18(b). As the simulated resistivities in Fig. 7.18(a) show, I have not been able to reproduce this high field peak, although preliminary calculations have yielded orbits and phase space cross sections similar to those reported. At $B \gg B_0$, most orbits are periodic or quasi-periodic with a slowly decaying autocorrelation function, and the calculated conductivity is sensitive to the accuracy of the simulation, often exhibiting oscillations in B which depend on the time step and cut-off. In all simulations, a further peak develops just above the fundamental peak at $B_z = B_0$, which can be identified with orbits trapped in between antidots; this effect is already indicated in the smaller diameter data of Fig. 7.4(a). There is no corresponding feature in OX4532′R, but the shoulder observed on the commensurability peak in the magnetoresistivity of several larger period antidot lattices on InAs–GaSb DHETs may be attributed to trapped orbits of this kind [30].

In OX4532′R, the high field peak (β) disappears at a relatively low temperature and changes its position in a parallel magnetic field, which is not explained by a semiclassical theory of rosette-shaped orbits. Moreover, the physical etch pit diameter of $0{\cdot}25a$ is much smaller than the antidot size at which the effect is predicted.

Alternatively, a more complicated effective potential similar to that of Fig. 7.16(c) may lie at the origin of the additional higher field structure. Because of the significant effect of the



(a) Diagonal components of the resistivity tensor as a function of B_z for $d/a = 0{\cdot}5$ (bottom), 0·6, and 0·7 (top). The curves are offset by $5\rho_0$.

(b) Rosette shaped orbit at $2{\cdot}5B_0$.

Figure 7.18: Calculations for large diameter antidot lattices with $d/a = 0{\cdot}66$. $B_0 = 2m^* v_\mathrm{F}/ea$ is the field for which $2R_\mathrm{c} = a$ and $\rho_0 = (2\pi\hbar^2)/(m^* e^2 \tau v_\mathrm{F}^2)$ is the Drude resistivity.

surface states, it is probable that the potential between antidots is not completely flat, and in this case smaller scale variations are present that could lead to resonances at higher magnetic fields corresponding to smaller cyclotron radii. Halving the distance along the line between two antidots corresponds to a cyclotron radius at twice the magnetic field value B_0 of the fundamental peak, whereas halving the distance along the diagonal of a square formed by four antidots matches a field of $\sqrt{2}B_0$. While the estimate is overly simplistic and one would not necessarily expect commensurability peaks at exactly these magnetic fields, this covers the range of the observed features.

7.7 Conclusions

If modulated by a sufficiently strong two-dimensional superlattice potential with a lattice constant a in the range $\lambda_\mathrm{F} < a < \ell_\mathrm{f}$, high mobility InAs–GaSb DHETs, which contain sheets

of mobile electrons *and* holes, typically show commensurability peaks in the perpendicular field magnetoresistance very similar to those observed in pure 2DEGs. These features persist for temperatures in excess of 20 K for thick cap samples and disappear above 10 K in lower mobility thin cap samples. The peak positions are compatible with electron-like orbits that occur as a result of magnetic breakdown and also manifest themselves in Shubnikov–de Haas quantum oscillations.

Despite the strong interaction between the electrons and holes, the effect of the latter is not apparent. It has been probed by applying an additional magnetic field component parallel to the plane of the carrier sheets, which is known to shift the electron and hole dispersion relations in momentum space. Depending on the sample, the peaks in the magnetoresistance are either unaffected or disappear at a parallel magnetic field lower than that required to completely decouple electrons and holes. For the samples in the latter category, the etch pits cannot be assumed to extend through the InAs layer, so that only the holes see an antidot potential. It has therefore been proposed that the electron-hole coupling is essential for the appearance of commensurability features in these samples; also, the probability of magnetic breakdown may be affected by the changing size of the gaps in momentum space as the in-plane field is varied. A theoretical model that can explain all observations has, however, not been achieved.

A single device, which after an additional etch step has the thinnest GaSb cap of the samples showing commensurability effects, shows further structure of unknown origin above the fundamental peak. The temperature dependence differs, with features disappearing between 2·5 and 4·5 K, and the strongest peak in this region moves to lower perpendicular fields as the parallel field component is increased.

Although LAO can be used to modify GaSb surfaces with high resolution, the method cannot be used to impose an antidot potential on the mobile electrons and holes in typical InAs–GaSb DHETs. In order to maintain an electron mean free path $\ell_f > a$ and avoid full depletion of the hole gas, a minimum GaSb cap depth exceeding the maximum anodization depth is required. While the sensitivity of the carrier gases to the Fermi level pinning at the

free surface implies that a partial removal of the cap layer can be sufficient to create a significant modulation potential, which has been demonstrated in prior experiments on large chemically etched antidots, AFM lithography cannot capitalize on this phenomenon. The probable explanation is that the surface states created by direct anodization or by careful removal of the anodic oxide and reoxidation in air (which preserves the lithographic pattern) are significantly different from those created by the wet etch used earlier and do not set up a sufficiently strong potential at the active layer. This idea is corroborated by the fact that pits etched using RIE and allowed to form a native oxide in air also can remove a significant percentage of the cap thickness without generating an effect on the magnetotransport properties.

Bibliography

[1] R. W. Winkler and J. P. Kotthaus. Landau-band conductivity in a two-dimensional electron system modulated by an artificial one-dimensional superlattice potential. *Physical Review Letters*, 62(10):1177–1180, March 1989.

[2] P. D. Ye, D. Weiss, R. R. Gerhardts, M. Seeger, K. von Klitzing, K. Eberl, and H. Nickel. Electrons in a periodic magnetic field induced by a regular array of micromagnets. *Physical Review Letters*, 74(15):3013–3016, April 1995.

[3] D. Weiss, K. von Klitzing, K. Ploog, and G. Weimann. Magnetoresistance oscillations in a two-dimensional electron gas induced by a submicrometer periodic potential. *Europhysics Letters*, 8(2):179–184, January 1989.

[4] D. Weiss, C. Zhang, R. R. Gerhardts, K. von Klitzing, and G. Weimann. Density of states in a two-dimensional electron gas in the presence of a one-dimensional superlattice potential. *Physical Review B*, 39(17):13020–13023, June 1989.

[5] R. R. Gerhardts, D. Weiss, and K. v. Klitzing. Novel magnetoresistance oscillations in a periodically modulated two-dimensional electron gas. *Physical Review Letters*, 62(10):1173–1176, March 1989.

[6] R. R. Gerhardts, D. Weiss, and U. Wolf. Magnetoresistance oscillations in a grid potential: Indication of a Hofstadter-type energy spectrum. *Physical Review B*, 43(6):5192–5195, February 1991.

[7] R. Kubo. Statistical-mechanical theory of irreversible processes. I. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957.

[8] C. Zhang and R. R. Gerhardts. Theory of magnetotransport in two-dimensional electron systems with unidirectional periodic modulation. *Physical Review B*, 41(18):12850–12861, June 1990.

[9] D. Pfannkuche and R. R. Gerhardts. Theory of magnetotransport in two-dimensional electron systems subjected to weak two-dimensional lattice potentials. *Physical Review B*, 46(19):12606–12626, November 1992.

[10] P. Vasilopoulos and F. M. Peeters. Periodic magnetic modulation of a two-dimensional electron gas: Transport properties. *Superlattices and Microstructures*, 7(4):393–395, 1990.

[11] F. M. Peeters and P. Vasilopoulos. Quantum transport of a two-dimensional electron gas in a spatially modulated magnetic field. *Physical Review B*, 47(3):1466–1473, January 1993.

[12] D. P. Xue and G. Xiao. Magnetotransport properties of two-dimensional electron gases under a periodic magnetic field. *Physical Review B*, 45(11):5986–5990, March 1992.

[13] C. W. J. Beenakker. Guiding-center-drift resonance in a periodically modulated two-dimensional electron-gas. *Physical Review Letters*, 62(17):2020–2793, April 1989.

[14] R. R. Gerhardts. Quasiclassical calculation of magnetoresistance oscillations of a two-dimensional electron gas in an anharmonic lateral superlattice potential. *Physical Review B*, 45(7):3449–3454, February 1992.

[15] R. R. Gerhardts. Quasiclassical calculation of magnetoresistance oscillations of a two-dimensional electron gas in spatially periodic magnetic and electrostatic fields. *Physical Review B*, 53(16):11075, April 1996.

[16] R. Menne and R. R. Gerhardts. Magnetoresistance of a two-dimensional electron gas with spatially periodic lateral modulations: Exact consequences of Boltzmann's equation. *Physical Review B*, 57(3):1707–1722, January 1998.

[17] R. R. Gerhardts and S. D. M. Zwerschke. Guiding-center picture of magnetoresistance oscillations in rectangular superlattices. *Physical Review B*, 64:115322, 2001.

[18] D. Weiss, M. L. Roukes, A. Menschig, P. Grambow, K. von Klitzing, and G. Weimann. Electron pinball and commensurate orbits in a periodic array of scatterers. *Physical Review Letters*, 66(21):2790–2793, May 1991.

[19] M. Wendel, B. Irmer, J. Cortes, R. Kaiser, H. Lorenz, J. P. Kotthaus, A. Lorke, and E. Williams. Nanolithography with an atomic force microscope. *Superlattices and Microstructures*, 20(3):349–356, 1996.

[20] A. Dorn, M. Sigrist, A. Fuhrer, T. Ihn, T. Heinzel, K. Ensslin, W. Wegscheider, and M. Bichler. Electric properties of antidot lattices fabricated by atomic force lithography. *Applied Physics Letters*, 80(2):252–254, January 2002.

[21] J. Eroms, M. Zitzlsperger, and D. Weiss. Skipping orbits and enhanced resistivity in large-diameter InAs/GaSb antidot lattices. *Physical Review B*, 59(12):7829–7832, March 1999.

[22] A. Lorke, S. Wimmer, B. Jager, J. P. Kotthaus, W. Wegscheider, and M. Bichler. Far-infrared and transport properties of antidot arrays with broken symmetry. *Physica B*, 249–251:312–316, 1998.

[23] S. de Haan, A. Lorke, R. Hennig, M. Suhrke, W. Wegscheider, and M. Bichler. Magneto-transport properties of arrays of cross-shaped antidots. *Physical Review B*, 60(12):8845–8848, September 1999.

[24] B. G. L. Jager, S. Wimmer, A. Lorke, J. P. Kotthaus, W. Wegscheider, and M. Bichler. Edge and bulk effects in terahertz photoconductivity of an antidot superlattice. *Physical Review B*, 63, 2001.

[25] J. Rychen, T. Vančura, and T. Heinzel. Commensurability oscillations of rectangular antidot arrays: A classical diffusion model. *Physical Review B*, 58(7):3458–3571, August 1998.

[26] R. Fleischmann, T. Geisel, and R. Ketzmerick. Magnetoresistance due to chaos and nonlinear resonances in lateral surface superlattices. *Physical Review Letters*, 68(9):1367–1370, March 1992.

[27] S. Ishizaka and T. Ando. Detailed analysis of the commensurability peak in antidot arrays with various periods. *Physical Review B*, 55(24):16331–16338, 1997. June.

[28] R. Hennig and M. Suhrke. Absence of commensurability in magnetic antidot arrays with classical chaos. *Physical Review B*, 60(16):11535–11539, October 1999.

[29] R. Onderka, M. Suhrke, and U. Rössler. Anisotropic magnetotransport in a rectangular antidot superlattice: Classical and semiclassical aspects. *Physical Review B*, 62(16):10918–10922, October 2000.

[30] Y. J. Rehman. *Processing and Magneto-transport Studies of InAs/GaSb Low Dimensional Structures*. DPhil thesis, The Queen's College, University of Oxford, 1999.

[31] T. O. Stadelmann, B. Kardynal, R. J. Nicholas, K. Takashina, and N. J. Mason. Magnetotransport studies of antidot superlattices in coupled two-dimensional electron-hole gases. *Physica E*, 12(1–4):293–295, January 2002.

[32] S. Sasa, A. Nakashima, S. Yodogawa, T. Kita, and M. Inoue. Magnetotransport properties of InAs nanostructure devices produced by AFM oxidation. *Physica B*, 314:95–98, 2002.

[33] S. Sasa, A. Nakashima, Y. Nakajima, and M. Inoue. Aharonov–Bohm oscillations observed in nanoscale open-dot structures fabricated in an InAs surface inversion layer. *Physica E*, 20:224–227, 2004.

[34] K. Takashina. *Magneto-transport Studies of GaSb/InAs/GaSb Double-heterostructures*. DPhil thesis, University College, University of Oxford, 2002.

[35] M. Altarelli, J. C. Maan, L. L. Chang, and L. Esaki. Electronic states and quantum Hall effect in GaSb–InAs–GaSb quantum wells. *Physical Review B*, 35(18):9867–9870, June 1987.

[36] C. Petchsingh. *Cyclotron Resonance Studies on InAs/GaSb Heterostructures*. DPhil thesis, Wolfson College, University of Oxford, 2002.

8 Summary

> Causas rerum naturalium non plures admitti debere, quam quae et verae sint et earum phaenomenis explicandis sufficiant.
>
> *(Sir Isaac Newton, Pricipia Mathematica)*

8.1 Introduction

Real life research projects frequently follow a less than straightforward evolution as planning is adjusted to reflect new discoveries and practical limitations.[1] This can result in a thesis that presents itself as a smörgåsbord of different results and insights with less clear structure than one might hope for. The purpose of this summary is to collect the salient points and put them into perspective.

8.2 Technology

I have demonstrated that local anodic oxidation (LAO) with an atomic force microscope (AFM) can be performed on a variety of GaSb and InAs surfaces, including those grown by metal-organic vapour phase epitaxy (MOVPE), with a performance comparable to that reported for other III-V semiconductors [1, 2]. The technique does meet the expectations in terms of versatility and attainable resolution, with a minimum lateral feature size of less than 30 nm, but is marred by two significant limitations: On the one hand, despite considerable

[1] In fact, one of the lessons I have learnt in the course of the present research is that it might not be wise to resist this process too much.

efforts to improve reproducibility, the results are not very reliable. With sufficient care it is possible to produce highly uniform patterns several square micrometres in size, but the success rate remains disappointingly low. On the other hand, the anodization is limited to the surface of the sample, extending less than 20 nm into the substrate, and modifications with a high aspect ratio are unachievable.

It is possible to avoid this restriction by using LAO to pattern an etch mask instead of directly modifying the semiconductor sample [3], and I have shown how this approach can be implemented on GaSb using an aluminium mask together with a silicon dioxide spacer layer to avoid detrimental reactions of the developing etch with the substrate. The resulting mask can be used directly as a stop layer for a plasma etch process. Then, however, much of the simplicity of the AFM lithography process is lost and it has to compete directly with more reliable conventional pattern transfer techniques such as electron beam lithography (EBL).

8.3 Solid State Physics

InAs–GaSb double heterostructures (DHETS) containing quasi-two-dimensional sheets of both holes and electrons that have been patterned with a regular array of antidots with a period smaller than the electron mean free path exhibit commensurability features in the perpendicular field magnetoresistance similar to those observed in unipolar systems [4]. These features can be attributed unambiguously to the commensurability of the electron cyclotron orbit with the antidot lattice: even though the hybridization between the electron and hole levels breaks up the non-interacting Fermi contours and leads to more complex orbits, electron-like orbits remain possible via magnetic breakdown. The effect of the commensurability on the resistivity can be demonstrated from a semiclassical simulation of electron orbits by means of KUBO's formula. A more intuitive explanation is afforded by the 'pinning' of suitably sized orbits around antidots.

One idea underlying the project discussed in this thesis has been to use LAO to create such potential modulations on InAs–GaSb DHETs. The Fermi level at the InAs layer of DHETs with a thin GaSb cap is known to be highly sensitive to surface states, and preliminary experiments with wet etched antidot patterns have suggested that a complete removal of the cap layer is not needed for creating a strong modulation potential. After investigating a considerable number of shallow LAO antidot samples as originally envisaged, it is now clear that a shallow surface modification with the AFM cannot produce a strong enough potential modulation to cause a measurable effect, even if a substantial percentage of the cap layer is affected.

As the behaviour of these samples is dominated by electron dynamics because of the higher electron mobility and the possibility of magnetic breakdown, the effect of the of mobile holes which set GaSb–InAs structures apart from more familiar systems such as two-dimensional electron gases (2DEGS) based on GaAs–$Al_xGa_{1-x}As$ is elusive. Samples with different potential profiles show different behaviour in a parallel magnetic field which is expected to decouple electrons and holes, with commensurability features disappearing for shallow dots that are expected to present an attractive potential to negatively charged carriers. The presence of commensurability peaks in the absence of a parallel field is then attributed to electron–hole interaction. A single electron-rich thin cap sample exhibits at least one novel peak at a magnetic field higher than that corresponding to the commensurability of the electron cyclotron orbit with the period of the potential. Different explanations have been investigated, but it is improbable that the feature is connected directly to the hole gas.

Supplementary magnetotransport studies on close to intrinsic double well structures have shown that the two wells behave essentially as two independent DHETs conducting in parallel for well distances above 100 angstrom. For smaller barrier widths, there are circumstantial signs for a change in the hole levels in the barrier. In no case an asymmetry between the conduction band states in the two wells has been demonstrated.

8.4 Concluding Remarks

Scanning probe microscopes provide an unprecedented ability to manipulate matter at scales ranging down to atomic dimensions for imaging and fabrication purposes. This capability makes them indispensable for many areas of contemporary research, and we will undoubtedly continue to hear of ever smaller devices constructed with the help of SPMs. It is less obvious which rôle LAO and similar techniques have in this development. Despite considerable effort by many researchers over the past years, the reliability of the method is limited, and while the attainable resolution is currently competitive for many applications, there is not much room for improvement. As pointed out in Chapter 2, the range of the interaction between probe and sample is often more critical for the resolution than the size of the probe itself. In the case of LAO, the underlying interaction is a chemical anodization process, and even in the limit of an infinitely small tip the thickness of the electrolyte layer and the scale of ionic diffusion processes put a lower limit on the possible feature size.

Bibliography

[1] R. Nemutudi, N. J. Curson, N. J. Appleyard, D. A. Ritchie, and G. A. C. Jones. Modification of a shallow 2DEG by AFM lithography. *Microelectronic Engineering*, 57–58:967–973, 2001.

[2] N. J. Curson, R. Nemutudi, N. J. Appleyard, M. Pepper, D. A. Ritchie, and G. A. C. Jones. Ballistic transport in a GaAs/$Al_xGa_{1-x}As$ one-dimensional channel fabricated using an atomic force microscope. *Applied Physics Letters*, 78(22):3466–3468, May 2001.

[3] A. Boisen, K. Birkelund, O. Hansen, and F. Grey. Fabrication of submicron suspended structures by laser and atomic force microscopy lithography on aluminum combined with reactive ion etching. *Journal of Vacuum Science and Technology B*, 16(6):2977–2981, November/December 1998.

[4] D. Weiss, M. L. Roukes, A. Menschig, P. Grambow, K. von Klitzing, and G. Weimann. Electron pinball and commensurate orbits in a periodic array of scatterers. *Physical Review Letters*, 66(21):2790–2793, May 1991.

A Scanner Design for Microscopy Applications

> He heaved his monumentally vile body round in his ill-fitting, slimy seat and stared at the monitor screen on which the starship *Heart of Gold* was being systematically scanned.
>
> *(Douglas Adams, The Restaurant at the End of the Universe)*

A.1 Piezoelectric Scanners

A.1.1 The Piezoelectric Effect

In crystalline dielectrics that lack inversion symmetry, such as quartz or tourmaline, mechanical stress along certain crystal axes causes the unit cell to become polarized as the charged atoms forming its basis move relative to each other. The direct piezoelectric[1] effect was discovered by Pierre and Jacques Curie in 1880. The polarization of the material is a linear function of the stress and in the absence of an external electrical field is given by

$$P_i = d_{ijk}\sigma_{jk}, \tag{A.1}$$

where P_i is the polarization vector, σ_{jk} is the mechanical stress, d_{ijk} is called the piezoelectric strain tensor, and the Einstein summation convention has been used [1]. As a third-rank tensor, d_{ijk} has 27 components, but since the stress tensor σ_{jk} is symmetric no more than 18 of them are independent.[2]

[1] from Greek πιέζειν, 'to squeeze'

[2] In engineering texts, j and k are sometimes contracted into a single index that runs over the 6 independent components of the strain tensor. This is known as Voigt's notation.

Conversely, an external electric field changes the equilibrium configuration of the basis atoms and leads to a deformation of the crystal, which is described by another linear relation:

$$\varepsilon_{jk} = d_{ijk} E_i, \tag{A.2}$$

where ε_{jk} is the strain tensor, E_i is the electric field vector, and the material is assumed to be relaxed. The converse piezoelectric effect constitutes the underlying principle of piezoelectric actuators [1].

A.1.2 Ferroelectric Ceramics

Actuators are most commonly assembled from ferroelectric ceramics such as lead zirconate titanate (PZT, $Pb(Zr_xTi_{1-x})O_3$). Ferroelectric materials are piezoelectric materials that spontaneously polarize below the Curie temperature T_C because of symmetry breaking and can produce a strong piezoelectric response even as a polycrystal since the orientation of the polarization can be changed externally. The individual pieces of ceramic are produced by pressing a precursor powder in the desired shape and sintering the workpiece. The resulting actuator is polycrystalline with randomly oriented grains and polarization domains and has initially almost no ability to expand or contract. In order to polarize the ceramic, the pieces are heated above T_C and cooled down in the presence of a strong electric field. After this process, the polarization direction in the grains mostly coincides with the equivalent crystal axis that is best aligned to the external field [1].

A.1.3 Deviations from Linearity

Although the piezoelectric effect, Eq. (A.2), is linear in nature, practical actuators exhibit a certain degree of nonlinearity [2–5]: For ferroelectric ceramics, the piezoelectric strain tensor depends on the average remnant polarization, which can change with time or in response to an external electric field. The effect is small for low driving voltages and small extensions and is consequently of limited importance to atomic scale imaging, which was

the initial focus of SPM. Metrology and lithography applications, however, require accurate calibration and repeatability over a large scanning area and hence are critically affected by nonlinearity and drift [2, 3].

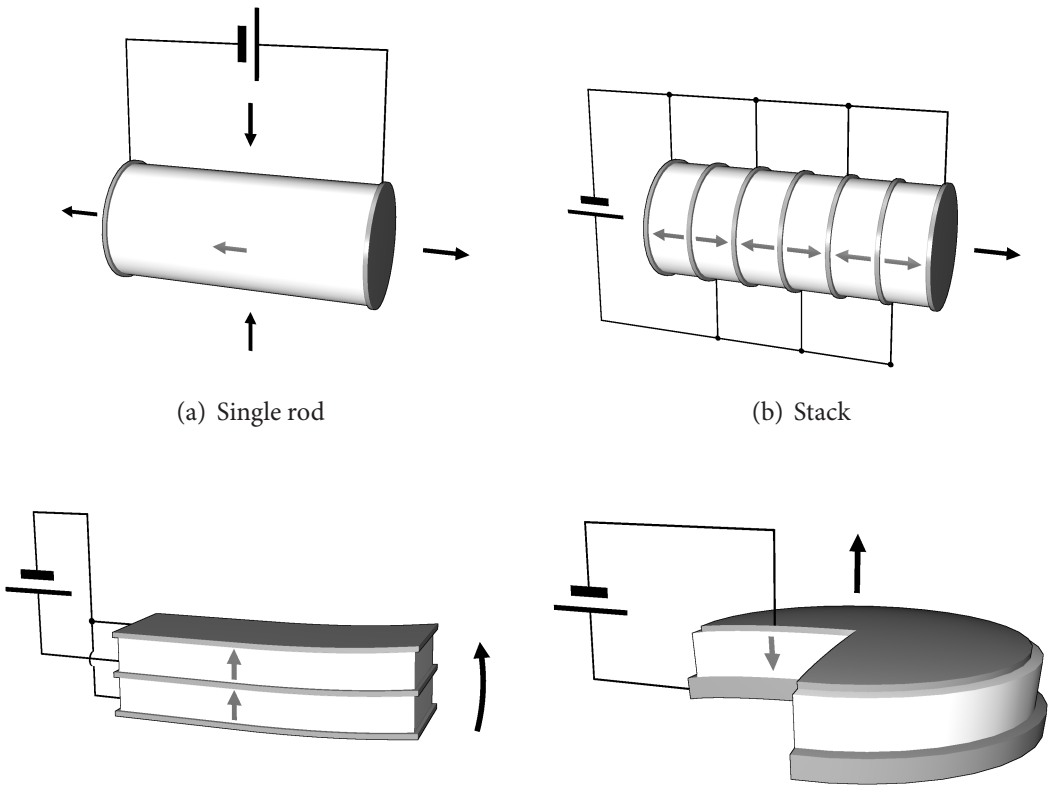
As the strength of the piezoelectric response is affected by the external electric field, the strain depends on the speed and direction of the change in the driving voltage. Accordingly, the relation between input voltage and displacement exhibits hysteresis. In typical ferroelectric actuators the error caused by hysteresis is 5 % to 10 % of the extension [4].

Even in a constant electric field the extension of the ferroelectric material initially continues to change slowly, an effect which is known as 'creep' or 'drift'. It is caused by the movement of the walls between individual polarization domains, which results in a change of the average polarization.

Over the lifetime of an actuator the alignment created by the poling process may degrade. This is especially true if the piezoceramic is heated to temperatures close to T_C, exposed to high electric fields in the direction opposite to its polarization, or used only rarely. Conversely, regular use of the actuator prevents degradation and can actually improve the alignment.

A.1.4 Basic Actuators

In the presence of an external electric field parallel to its average polarization, a rod of ferroelectric material will expand along the polarization axis and contract in the direction normal to it, as shown in Fig. A.1(a). If the electric field is reversed, the rod will contract along the poling axis and expand in the perpendicular direction. A rod of this kind can be used as a simple one-dimensional actuator, the range of which is proportional to its length and the range of the electric field. The field strength in the poling direction is limited by dielectric breakdown, while the allowable field strength in the inverse direction is typically much lower and given by the onset of depoling. The field required to achieve the theoretical maximal extension of a rod of commercially available PZT is several kV/mm, which makes this design



(a) Single rod (b) Stack

(c) Bimorph (d) Unimorph disk

Figure A.1: Basic piezoelectric actuators made from ferroelectric ceramic. Grey arrows indicate average polarization. Black arrows show deformation or movement.

impractical for many applications. Instead, the arrangement shown in Fig. A.1(b) is commonly used. Here the rod is replaced by a stack of thin piezoelectric disks with alternating poling connected in parallel. The voltage required to reach full extension is divided by the number of disks.

The deformation normal to the direction of the electric field can also be exploited in piezoelectric actuators. In Fig. A.1(c) two layers of ferroelectric material are combined in an arrangement similar to a bimetallic strip. Application of an external voltage causes one of the layers to expand while the other contracts so that the device bends upwards. If one end is clamped, the other will move in an arc by a distance much larger than the deformation of the

ceramic. In general, a stacked structure comprising an arbitrary number of piezoelectric and elastic layers is known as a multimorph, which makes the structure in Fig. A.1(c) a bimorph. A familiar application of this idea—frequently used as a transducer or loudspeaker—is the unimorph[3] disk shown in Fig. A.1(d) consisting of just one piezoelectric layer attached to an elastic metal base, which doubles as one of the contacts. If a voltage is applied, the disk buckles axially.

A.1.5 Scanner Design

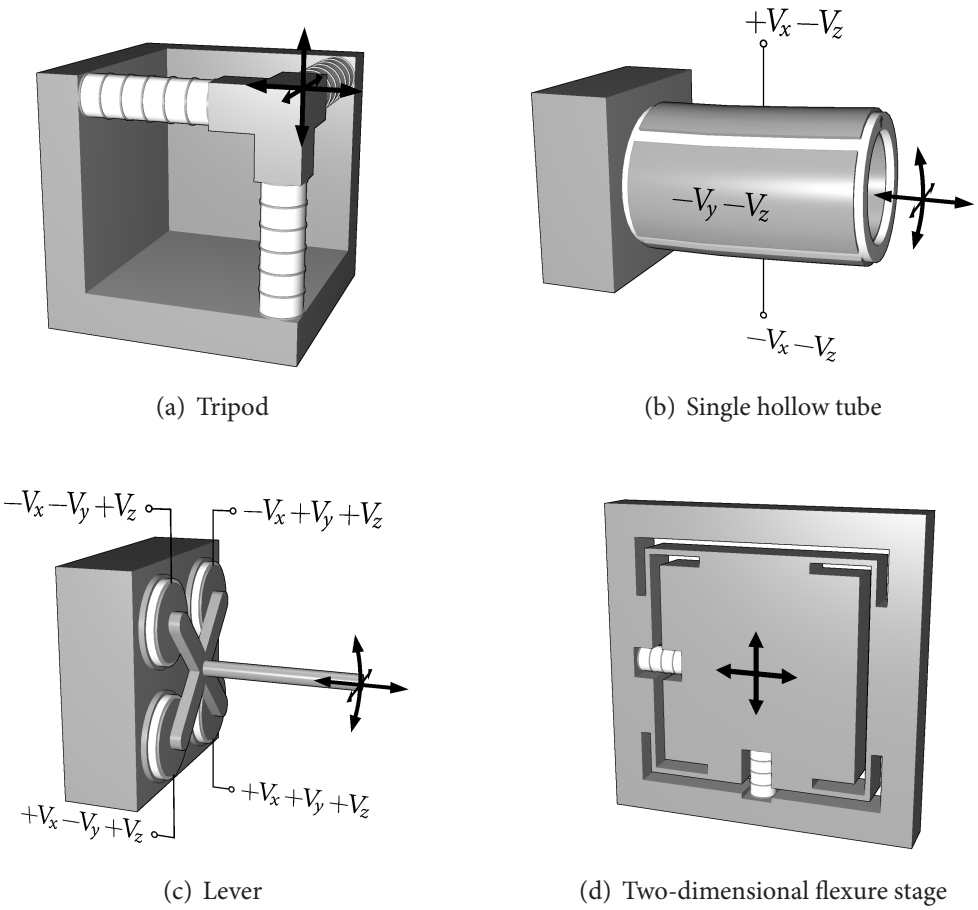
Tripod

Fig. A.2(a) shows the most straightforward three-dimensional scanner geometry, which was used by Binnig *et al.* [6] in the earliest SPM experiments. A separate one-dimensional piezoelectric actuator—typically a standard PZT stack—is used for each axis of movement. Since the position on the individual axes can be controlled independently, the feedback and data acquisition modules can be kept simple. Even so, because of the geometrical coupling of the actuators cross talk of the order 0·1 % to 1 % between the axes is present and may need to be corrected for [7]. The disadvantage of this design is an increased mechanical complexity and a limited range of movement in the plane of the sample.

Single Hollow Tube

In 1986 Binnig *et al.* introduced a sophisticated scanner, which requires only one piece of piezoelectric material [8,9]. The design makes use of the transverse component of the piezoelectric response and is now used in the majority of new SPMs. As illustrated in Fig. A.2(b), the piezoceramic is formed as a hollow tube with a single contact on the inside as well as four contacts on the outer surface, where each covers a quarter of the circumference. By applying equal and opposite voltages V_x (V_y) to opposing quadrants of the tube, the sides

[3] Rather confusingly, the similar term 'monomorph' is sometimes used to describe a different actuator lacking the elastic layer. Buckling is then induced by an inhomogeneous field.



(a) Tripod

(b) Single hollow tube

(c) Lever

(d) Two-dimensional flexure stage

Figure A.2: Piezoelectric SPM scanners. Arrows indicate scanner movement.

are induced to contract and expand respectively, and the tube bends. At the same time, a bias voltage V_z applied to all four quadrants with respect to the central contact changes the length of the entire tube. If one end of the tube is clamped to the support of the instrument, the other end, which may hold either the sample or the probe, can scan a three-dimensional volume [8,10]. The range of sideways movement that can be achieved in this manner is much larger than the longitudinal expansion of the sides of the hollow tube. The drawback is that the movement in the (x, y)-plane and along the z-axis is no longer independent. The interference between the three axes is much larger than with the tripod geometry: The free end of the tube maps out a curved surface if V_z is kept constant. Moreover, a complex controller module is required to translate (x, y, z)-values into voltages.

Lever Design

Fig. A.2(c) shows a design proposed by Mariani *et al.* [11] that is conceptually similar to the tube scanner but does not require a specialized piece of piezoceramic. Instead, four standard unimorph disk actuators are laid out in the corners of a square and connected by a rigid cross supporting a lever. If the actuators on one side of the square retract while the ones on the other side move upwards the lever swings to that side. Movement along the lever axis—with a limited range given by the travel of the disks—can be achieved by driving all four actuators in parallel. The design suffers from the same cross talk problems as the tube scanner and a lessened stiffness, but can be implemented very cheaply and allows for potentially very large backlash-free amplification of the movement in the (x, y)-plane (at the expense of reduced stiffness) [11]. A further simplification uses a single disk divided into four quadrants instead of four separate actuators.

Flexure Stages

Conventional three-dimensional scanners are not only afflicted by cross talk, but take up space directly above or below the point where the probe comes in contact with the sample, making it difficult to integrate the SPM with other microscopy techniques. This is particularly problematic for near field optical microscopy. Several schemes have been proposed to move the scanner hardware away from the optical axis and contain it in a flat package [12]. A design that uses a set of leaf springs and levers to transmits the movement of an off-axis actuator to the sample stage as illustrated in Fig A.2(d) has proven particularly useful. The entire device is machined from a single piece of metal, and hinges are provided by the flexure of thin metal bridges, avoiding backlash and the need for lubrication [5, 13]. Geometrical amplification of the movement range is possible by a suitable arrangement of levers as in the original design by Scire [13], but reduces the stiffness of the translator [7]. It is possible to combine several independent stages, although the combination of x and y scanners in a single frame as shown in Fig. A.2(d) helps minimizing Abbe errors [13].

A.2 Electromagnetic Scanners

A.2.1 Electrodynamic Actuators

In an electrodynamic or inductive actuator a solenoid carrying the driving current moves in the radial field between the poles of a permanent magnet. The solenoid experiences a force, which is balanced by a diaphragm or spring, so that the displacement of the actuator is

$$d = \frac{F}{k} = \frac{nI\Phi}{k}, \tag{A.3}$$

where F is the force, k the spring constant, n the number of windings in the solenoid, I the current, and Φ the magnetic flux in the gap.

This arrangement, often know as a voice coil, is familiar from acoustic loudspeakers and used to see widespread use in the laboratory for micropositioning applications. Despite the fact that electrodynamic actuators are virtually free from hysteresis and drift and do not require high voltage electronics, they have been superseded by piezoelectric elements for fine positioning [14]. In conventional voice coil designs a moderate current causes a large displacement and the device has a low resonant frequency. Nanometre resolution would require a sensitive current control that is impossible to achieve in practice and the actuator position is highly susceptible to acoustic and electromagnetic noise on this length scale.

Binnig *et al.* [15] have shown that these disadvantages may be overcome by increasing the spring constant that the solenoid pushes against to such an extent that the travel for conveniently controllable driving currents is reduced to the desired scanner range. The deviations caused by instabilities in the driving current are demagnified proportionally so that the need for sensitive electronic control is obviated. At the same time, the effective stiffness of the scanner is increased, leading to higher mechanical resonance frequencies and reducing the interference from the scanning frequency and low frequency acoustic noise.

A.2.2 Scanner Design

Electrodynamic actuators can be used to construct a three-dimensional scanner in much the same way as piezoelectric actuators. While there is no analogy to the the hollow tube design of Fig. A.2(b), the other scanners shown in Fig. A.2, which use individual one-dimensional positioners, can be implemented readily. Most commercial offerings use a form of flexure stage, while Mariani has realized a variant of his inexpensive lever scheme that employs standard voice coils [14]. Binnig *et al.* [15] have used a central lever that doubles as the elastic load for both x and y actuators.

Unlike piezoelectric scanners, electromagnetic devices can deliver large travel ranges even with small actuator sizes. This makes them the preferred solution for applications where miniaturization is important, for example where a large number of probes is to be operated independently within a limited area [16].

Bibliography

[1] L. E. Cross. Ferroelectric materials for electromechanical transducer applications. *Materials Chemistry and Physics*, 43:108–115, 1996.

[2] L. Libioulle, A. Ronda, M. Taborelli, and J. M. Gilles. Deformations and nonlinearity in scanning tunneling microscope images. *Journal of Vacuum Science and Technology B*, 9(2):655–658, March/April 1991.

[3] K. R. Koops, P. M. L. O. Scholte, and W. L. de Koning. Observation of zero creep in piezoelectric actuators. *Applied Physics A*, 68:691–697, 1999.

[4] K. Dirscherl, J. Garnæs, and L. Nielsen. Modeling the hysteresis of a scanning probe microscope. *Journal of Vacuum Science and Technology B*, 18(2), 621–625 2000.

[5] H.-C. Yeh, W.-T. Ni, and S.-S. Pan. Digital closed-loop nanopositioning using rectilinear flexure stage and laser interferometry. *Control Engineering Practice*, 13:559–566, 2005.

[6] G. Binnig, H. Rohrer, C. Gerber, and E. Weibel. Surface studies by scanning tunneling microscopy. *Physical Review Letters*, 49(1):57–61, July 1982.

[7] H.-C. Zhang, A. Sasaki, J. Fukaya, and H. Aoyama. Surface roughness observation by scanning tunneling microscopy using a monolithic parallel spring. *Journal of Vacuum Science and Technology B*, 12(3):1669–1672, May/June 1994.

[8] G. Binnig and D. P. E. Smith. Single-tube three-dimensional scanner for scanning tunneling microscopy. *Review of Scientific Instruments*, 57(8):1688–1689, March 1986.

[9] A. Franks. Nanotechnology. *Journal of Physical Engineering: Scientific Instruments*, 20:1442–1451, 1987.

[10] E. T. Yu. Nanoscale characterization of semiconductor materials and devices using scanning probe techniques. *Materials Science and Engineering*, R17:147–206, 1996.

[11] T. Mariani, C. Frediani, and C. Ascoli. Non-conventional, inexpensive 3-D scanners for probe microscopy. *Applied Physics A*, 66:S35–S40, 1998.

[12] A. Lewis and K. Lieberman. Flat scanning stage for scanned probe microscopy. US Patent 5,705,878, January 1998.

[13] F. E. Scire. Planar biaxial micropositioning stage. US Patent 4,506,154, March 1985.

[14] T. Mariani, C. Frediani, and C. Ascoli. A three-dimensional scanner for probe microscopy on the millimetre scale. *Applied Physics A*, 66:S861–S866, 1998.

[15] G. K. Binnig, W. Haeberle, H. Rohrer, and D. P. E. Smith. Fine positioning apparatus with atomic resolution. US Patent 5,808,302, September 1998.

[16] H. Rothuizen, U. Drechsler, G. Genolet, W. Häberle, M. Lutwyche, R. Stutz, R. Widmer, and P. Vettiger. Fabrication of a micromachined magnetic $X/Y/Z$ scanner for parallel scanning probe applications. *Microelectronic Engineering*, 53:509–512, 2000.

B Calculating Conductivity

> The mathematics is not there till we put it there.
>
> *(Sir Arthur Stanley Eddington, The Philosophy of Physical Science)*

B.1 The Boltzmann Equation

The Boltzmann Equation describes, in the presence of scattering, the evolution of the distribution function $f(\mathbf{r}, \mathbf{k}, t)$ of an ensemble of particles obeying the semiclassical equations of motion

$$\dot{\mathbf{r}} = \mathbf{v}(\mathbf{k}) \qquad \text{and} \tag{B.1}$$

$$\hbar\dot{\mathbf{k}} = \mathbf{F}(\mathbf{r}, \mathbf{k}), \tag{B.2}$$

where $\mathbf{r}(t)$ and $\mathbf{k}(t)$ denote position and crystal momentum, and $\mathbf{v}$ and $\mathbf{F}$ have their usual meanings of velocity and force, respectively. It can then be shown that [1]

$$\frac{df(\mathbf{r}, \mathbf{k}, t)}{dt} = \left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla_{\mathbf{r}} + \mathbf{F} \cdot \frac{1}{\hbar} \nabla_{\mathbf{k}} \right) f(\mathbf{r}, \mathbf{k}, t) = \left[\frac{\partial f(\mathbf{r}, \mathbf{k}, t)}{\partial t} \right]_{\text{coll}}, \tag{B.3}$$

where the right hand side describes the change in $f(\mathbf{r}, \mathbf{k}, t)$ due to collisions.

Once $f(\mathbf{r}, \mathbf{k}, t)$ is known, the carrier and current densities can be calculated by integrating over momentum space, yielding

$$n(\mathbf{r}, t) = (2\pi)^{-d} \int g(\mathbf{k}) f(\mathbf{r}, \mathbf{k}, t)\, d\mathbf{k} \qquad \text{and} \tag{B.4}$$

$$\mathbf{j}(\mathbf{r}, t) = (2\pi)^{-d} \int g(\mathbf{k}) f(\mathbf{r}, \mathbf{k}, t) q\mathbf{v}(\mathbf{k})\, d\mathbf{k}, \tag{B.5}$$

where d is the number of dimensions and $g(\mathbf{k})$ is the density of states.

To make the Boltzmann equation (B.3) tractable, the relaxation time approximation is often used. It amounts to the assumption that the scattering probability for a carrier with momentum $\mathbf{k}$ at position $\mathbf{r}$ is of the form $dt/\tau(\mathbf{r},\mathbf{k})$, and that the distribution of carriers emerging from collisions is the local equilibrium distribution $f_0(\mathbf{r},\mathbf{k},t)$, *i.e.*, the Fermi-Dirac distribution $f_{\mathrm{FD}}(E) \stackrel{\mathrm{def}}{=} \{\exp[(E-\mu)/k_{\mathrm{B}}T]+1\}^{-1}$ evaluated at $E = E[\mathbf{k}(t)]$ and allowing for a local temperature $T = T[\mathbf{r}(t)]$ and chemical potential $\mu = \mu[\mathbf{r}(t)]$. The collision term then has the simple form

$$\left[\frac{\partial f(\mathbf{r},\mathbf{k},t)}{\partial t}\right]_{\mathrm{coll}} = -\frac{f(\mathbf{r},\mathbf{k},t) - f_0(\mathbf{r},\mathbf{k},t)}{\tau(\mathbf{r},\mathbf{k})}, \tag{B.6}$$

and the Boltzmann equation (B.3) becomes

$$\frac{df(\mathbf{r},\mathbf{k},t)}{dt} = -\frac{f(\mathbf{r},\mathbf{k},t) - f_0(\mathbf{r},\mathbf{k},t)}{\tau(\mathbf{r},\mathbf{k})}. \tag{B.7}$$

Eq. (B.7) can be now be solved using standard techniques, leading to

$$f(\mathbf{r},\mathbf{k},t) = e^{-\int \tau(\mathbf{r},\mathbf{k})^{-1}\,dt}\left(\int_{-\infty}^{t} \frac{f_0(\mathbf{r},\mathbf{k},t)e^{\int \tau(\mathbf{r},\mathbf{k})^{-1}\,dt}}{\tau(\mathbf{r},\mathbf{k})}\,dt + \text{const.}\right). \tag{B.8}$$

The constant (with respect to t) depends on the initial conditions and the corresponding term decays exponentially for physical $\tau(\mathbf{r},\mathbf{k})$. The first term, which describes the intrinsic behaviour of the system, may also be written as [1]

$$f(\mathbf{r},\mathbf{k},t) = f_0(\mathbf{r},\mathbf{k},t) - \int_{-\infty}^{t} P(t,t')\frac{d}{dt'}f_0(\mathbf{r},\mathbf{k},t)\,dt', \tag{B.9}$$

where $P(t,t') \stackrel{\mathrm{def}}{=} \exp\left\{-\int_{t'}^{t} 1/\tau[\mathbf{r}(t''),\mathbf{k}(t'')]\,dt''\right\}$ is the probability for a particle with momentum $\mathbf{k}(t)$ at position $\mathbf{r}(t)$ *not* to have undergone collision in the time interval $[t',t]$.

If $\mathbf{v}(\mathbf{k}) = (1/\hbar)\,\nabla_{\mathbf{k}}E(\mathbf{k})$, the force is $\mathbf{F}(\mathbf{r},\mathbf{k}) = q[\mathbf{E}(\mathbf{r},t)+\mathbf{v}(\mathbf{k})\times\mathbf{B}(\mathbf{r},t)]$, and the chemical potential μ as well as the temperature T are independent of position, then

$$\frac{df_0(\mathbf{r},\mathbf{k},t)}{dt} = \frac{\partial f_0(\mathbf{r},\mathbf{k},t)}{\partial E}\,\hbar\mathbf{v}(\mathbf{k})\cdot\frac{q\mathbf{E}(\mathbf{r},t)}{\hbar} = q\frac{\partial f_{\mathrm{FD}}(E)}{\partial E}\,\mathbf{v}(\mathbf{k})\cdot\mathbf{E}(\mathbf{r},t); \tag{B.10}$$

if one also makes the simplifying assumption of a constant relaxation time τ,[1] $P(t,t') = \exp[-(t-t')/\tau]$. Under these conditions, Eq. (B.9) becomes

$$f(\mathbf{r},\mathbf{k},t) = f_0(\mathbf{r},\mathbf{k},t) - q\int_{-\infty}^{t} e^{-(t-t')/\tau}\frac{\partial f_{\mathrm{FD}}(E)}{\partial E}\mathbf{v}(\mathbf{k})\cdot\mathbf{E}(\mathbf{r},t)\,dt', \tag{B.11}$$

[1] If τ only depends on energy, $\partial\tau/\partial t$ is still approximately zero in metals and the same arguments can be used [1].

which further simplifies to

$$f(\mathbf{k}) = f_0(\mathbf{k}) - \Re\left[\frac{q}{\tau^{-1} - i\omega}\frac{\partial f_{\mathrm{FD}}(E)}{\partial E}\mathbf{v}(\mathbf{k})\cdot\mathbf{E}(\omega)\right] \tag{B.12}$$

if $\mathbf{B}(\mathbf{r},t) = \mathbf{o}$ and $\mathbf{E}(\mathbf{r},t) = \Re[\mathbf{E}(\omega)\exp(-i\omega t)]$. Using Eq. (B.12) in Eq. (B.5), the current density $\mathbf{j}(\omega)$ is calculated as

$$\mathbf{j}(\omega) = \frac{q^2}{(2\pi)^d}\int \frac{g(E)}{\tau^{-1} - i\omega}\frac{\partial f_{\mathrm{FD}}(E)}{\partial E}\mathbf{v}(\mathbf{k})\cdot\mathbf{E}(\omega)\,\mathbf{v}(\mathbf{k})\,d\mathbf{k}. \tag{B.13}$$

Since $\mathbf{j}(\omega) = \boldsymbol{\sigma}(\omega)\mathbf{E}(\omega)$, the conductivity tensor is accordingly

$$\sigma_{\mu\nu}(\omega) = \frac{q^2}{(2\pi)^d}\int \frac{g(E)}{\tau^{-1} - i\omega}\left[-\frac{\partial f_{\mathrm{FD}}(E)}{\partial E}\right]v_\mu(\mathbf{k})v_\nu(\mathbf{k})\,d\mathbf{k}. \tag{B.14}$$

The derivation is independently valid for every energy band in which $\partial f_{\mathrm{FD}}(E)/\partial E$ is nonzero (*i.e.*, which intersects the Fermi level), and the total conductivity is obtained by summing the contributions of all such bands multiplied by their individual degeneracies.

B.2 Kubo's Linear Response Theory

B.2.1 Fundamentals

In his 1957 paper [2], Ryogo Kubo discusses how the linear response to a small perturbation of a system in equilibrium can be expressed in terms of the fluctuations of dynamical variables of the unperturbed system. His linear response theory, which extended and unified prior work on the statistical description of transport properties, has since become an important tool in fields such as condensed matter physics and fluid dynamics.

If $B(t)$ is a physical quantity corresponding to the expectation value of an observable $\hat{B}$ at time t, the linear change $\Delta B(t)$ in $B(t)$ due to a spatially homogeneous[2] field $\xi_A(t)$ can be written as

$$\Delta B(t) = \int_{-\infty}^{\infty}\alpha_{BA}(t-t')\xi(t')\,dt', \tag{B.15}$$

[2] The generalization to inhomogeneous fields is complicated, but conceptually straightforward.

where $\alpha_{BA}(t)$ is the response at time t caused by a unit impulse at $t = 0$, and causality implies $\alpha_{BA}(t) = 0$ for $t < 0$. For a periodic field $\xi_A(t) = \Re\big[\xi_A^{(0)} \exp(i\omega t)\big]$, the response is expressed as

$$\Delta B_\omega(t) = \Re\left[\chi_{BA}(\omega)\xi_A^{(0)} e^{i\omega t}\right], \tag{B.16}$$

where the complex susceptibility $\chi_{BA}(\omega)$ is the Fourier transform of $\alpha_{BA}(t)$.[3]

To calculate $\alpha_{BA}(t)$, Kubo assumes that the system is described by the Hamiltonian $\hat{H}(t) = \hat{H}_0 - \xi_A(t)\hat{A}$, where $\hat{H}_0$ governs the natural motion of the unperturbed system and $\hat{A}$ is the generalized force conjugate to $\xi_A(t)$. The density matrix $\rho(t)$ then obeys the equation

$$\frac{\partial\rho(t)}{\partial t} = \frac{1}{i\hbar}\big[\hat{H}(t), \rho(t)\big] = \frac{1}{i\hbar}\Big(\big[\hat{H}_0, \rho(t)\big] - \xi_A(t)\big[\hat{A}, \rho(t)\big]\Big). \tag{B.17}$$

With the *Ansatz* $\rho(t) = \rho_0 + \Delta\rho(t)$, where $\Delta\rho(t) \to 0$ as $t \to -\infty$ and ρ_0 is taken to be the equilibrium distribution, *i.e.*, $[\rho_0, \hat{H}_0] = 0$, he simplifies Eq. (B.17) to

$$\frac{\partial\Delta\rho(t)}{\partial t} \approx \frac{1}{i\hbar}\Big(\big[\hat{H}_0, \Delta\rho(t)\big] - \xi_A(t)\big[\hat{A}, \rho_0\big]\Big), \tag{B.18}$$

which can be solved for $\Delta\rho(t)$ [2,3], giving

$$\Delta\rho(t) = -\frac{1}{i\hbar}\int_{-\infty}^{t}\big[\hat{A}(t'-t), \rho_0\big]\xi_A(t')\,dt', \tag{B.19}$$

where $\hat{A}(t) \stackrel{\text{def}}{=} \hat{U}_0^\dagger(t)\hat{A}\hat{U}_0(t)$ is a Heisenberg operator constructed using the time evolution operator $\hat{U}_0(t) \stackrel{\text{def}}{=} \exp(-i\hat{H}_0 t/\hbar)$. From Eq. (B.19) $\Delta B(t)$ follows as

$$\Delta B(t) = \operatorname{Tr}\Delta\rho(t)\hat{B} = -\frac{1}{i\hbar}\int_{-\infty}^{t}\operatorname{Tr}\big[\hat{A}(t'-t), \rho_0\big]\hat{B}\xi_A(t')\,dt'. \tag{B.20}$$

Comparing Eqs. (B.15) and (B.20), one finds

$$\alpha_{BA}(t) = \theta(t)\varphi_{BA}(t), \tag{B.21}$$

where $\theta(t)$ is the Heaviside function and

$$\varphi_{BA}(t) = -\frac{1}{i\hbar}\operatorname{Tr}\big[\hat{A}(-t), \rho_0\big]\hat{B}. \tag{B.22}$$

[3] In the following I shall define the Fourier transform of $F(t)$ as $\tilde{F}(\omega) \stackrel{\text{def}}{=} \int_{-\infty}^{\infty} F(t)\exp(-i\omega t)\,dt$ so that $F(t) = (1/2\pi)\int_{-\infty}^{\infty}\tilde{F}(\omega)\exp(i\omega t)\,d\omega$.

Because of the cyclic property of the trace and the definition of the Heisenberg operators, Eq. (B.22) can be rewritten as

$$\begin{aligned}\varphi_{BA}(t) &= \frac{1}{i\hbar}\operatorname{Tr}\big[\rho_0, \hat{A}(-t)\big]\hat{B} \\ &= \frac{1}{i\hbar}\operatorname{Tr}\rho_0\big[\hat{A}(-t), \hat{B}\big] \\ &= \frac{1}{i\hbar}\operatorname{Tr}\rho_0\big[\hat{A}(0), \hat{B}(t)\big] \qquad \text{(B.23)} \\ &= \frac{1}{i\hbar}\left\langle\big[\hat{A}(0), \hat{B}(t)\big]\right\rangle_0. \qquad \text{(B.24)}\end{aligned}$$

An analogous expression is found in the classical case; the derivation proceeds along exactly the same lines if ρ is taken as the classical phase space distribution function, the commutator $(1/i\hbar)[\hat{A}, \hat{B}]$ is replaced by the classical Poisson bracket $(A, B) \stackrel{\text{def}}{=} \sum_i \big[(\partial A/\partial q_i)(\partial B/\partial p_i) - (\partial A/\partial p_i)(\partial B/\partial q_i)\big]$ and the trace by integration over the full set $\{p_i, q_i\}$ of phase space coordinates.

In terms of matrix elements $A_{mn} \stackrel{\text{def}}{=} \langle m|\hat{A}|n\rangle$ in the basis of the eigenstates $|n\rangle$ of $\hat{H}_0$, the response function Eq. (B.24) is given explicitly as

$$\varphi_{BA}(t) = \frac{1}{i\hbar}\sum_{n,m} p_n \left(A_{nm}B_{mn}e^{-i\omega_{nm}t} - B_{nm}A_{mn}e^{i\omega_{nm}t}\right), \qquad \text{(B.25)}$$

where p_n are the weights of the density matrix $\rho_0 = \sum_n |n\rangle\, p_n\, \langle m|$, I define $\omega_{nm} \stackrel{\text{def}}{=} (E_n - E_m)/\hbar$, and E_n are the eigenvalues of $\hat{H}_0$. In the same basis, the Fourier transform of $\varphi_{AB}(t)$ is written as

$$\begin{aligned}\tilde{\varphi}_{BA}(\omega) &= \frac{2\pi}{i\hbar}\sum_{n,m} p_n\big[A_{nm}B_{mn}\delta(\omega+\omega_{nm}) - B_{nm}A_{mn}\delta(\omega-\omega_{nm})\big] \qquad \text{(B.26)} \\ &= \frac{2\pi}{i\hbar}\sum_{n,m}(p_n - p_m)A_{nm}B_{mn}\delta(\omega+\omega_{nm}). \qquad \text{(B.27)}\end{aligned}$$

B.2.2 Correlation Functions

To relate them to physically meaningful quantities, it is convenient to express $\tilde{\varphi}_{BA}(\omega)$—and hence $\varphi_{BA}(t)$, $\alpha_{BA}(t)$, and $\chi_{BA}(\omega)$—in terms of the symmetrized correlation function

$$\Psi_{BA}(t) \stackrel{\text{def}}{=} \left\langle\left\{\hat{A}(0), \hat{B}(t)\right\}\right\rangle_0 \stackrel{\text{def}}{=} \frac{1}{2}\left\langle\hat{A}(0)\hat{B}(t) + \hat{B}(t)\hat{A}(0)\right\rangle_0, \qquad \text{(B.28)}$$

which is one of the possible quantum mechanical equivalents of the classical correlation function $\langle A(0)B(t)\rangle$. In the basis of the energy eigenstates, the symmetrized correlation function becomes

$$\Psi_{BA}(t) = \frac{1}{2}\sum_{n,m} p_n \left(A_{nm}B_{mn}e^{-i\omega_{nm}t} + B_{nm}A_{mn}e^{i\omega_{nm}t}\right) \qquad \text{and} \tag{B.29}$$

$$\tilde{\Psi}_{BA}(\omega) = \pi\sum_{n,m} p_n\big[A_{nm}B_{mn}\delta(\omega+\omega_{nm}) + B_{nm}A_{mn}\delta(\omega-\omega_{nm})\big] \tag{B.30}$$

$$= \pi\sum_{n,m}(p_n+p_m)A_{nm}B_{mn}\delta(\omega+\omega_{nm}). \tag{B.31}$$

If $\rho_0 = \rho_0^{(\beta)}$ is a canonical distribution at the temperature $T = 1/k_B\beta$ [2, 4], the weights are $p_n = \exp(-\beta E_n)/\sum_m \exp(-\beta E_m)$ and $(p_n - p_m)/(p_n + p_m) = \tanh[-\frac{1}{2}\beta(E_n - E_m)]$. Eq. (B.27) then becomes

$$\begin{aligned}\tilde{\varphi}_{BA}^{(\beta)}(\omega) &= \frac{2\pi}{i\hbar}\sum_{n,m}\tanh\left(\frac{1}{2}\beta\hbar\omega_{nm}\right)(p_n+p_m)A_{nm}B_{mn}\delta(\omega+\omega_{nm})\\ &= \frac{2}{i\hbar}\tanh\left(\frac{1}{2}\beta\hbar\omega\right)\tilde{\Psi}_{BA}^{(\beta)}(\omega) = \frac{\omega}{iE_\beta(\omega)}\tilde{\Psi}_{BA}^{(\beta)}(\omega),\end{aligned} \tag{B.32}$$

where $E_\beta(\omega) = \hbar\omega/2 + \hbar\omega/[\exp(\beta\hbar\omega) - 1]$ is the average energy of an harmonic oscillator with frequency ω at $T = 1/k_B\beta$. On the other hand, if $\rho_0 = \rho_0^{(\mathrm{mp})}$ describes a system of many non-interacting particles [4], we can write the $\tilde{\varphi}_{BA}^{(\mathrm{mp})}(\omega)$ and $\tilde{\Psi}_{BA}^{(\mathrm{mp})}(\omega)$ in terms of the single particle microcanonical spectra $\tilde{\varphi}_{BA}^{(E)}(\omega)$ and $\tilde{\Psi}_{BA}^{(E)}(\omega)$ as

$$\tilde{\varphi}_{BA}^{(\mathrm{mp})}(\omega) = \sum_n p_n\tilde{\varphi}_{BA}^{(E_n)}(\omega) \qquad \text{and} \qquad \tilde{\Psi}_{BA}^{(\mathrm{mp})}(\omega) = \sum_n p_n\tilde{\Psi}_{BA}^{(E_n)}(\omega), \tag{B.33}$$

where $\tilde{\varphi}_{BA}^{(E_n)}(\omega)$ and $\tilde{\Psi}_{BA}^{(E_n)}(\omega)$ can be identified by comparing Eq. (B.33) with Eqs. (B.26) and (B.30). By exchanging summation indices in Eq. (B.26), adding the permutated version to the original, and dividing by 2, we can write

$$\begin{aligned}\tilde{\varphi}_{BA}^{(\mathrm{mp})}(\omega) &= \frac{\pi}{i\hbar}\sum_{n,m}(p_n-p_m)\big[A_{nm}B_{mn}\delta(\omega+\omega_{nm}) - B_{nm}A_{mn}\delta(\omega-\omega_{nm})\big]\\ &= -\omega\frac{\pi}{i}\sum_{n,m}\frac{p_n-p_m}{E_n-E_m}\big[A_{nm}B_{mn}\delta(\omega+\omega_{nm}) + B_{nm}A_{mn}\delta(\omega-\omega_{nm})\big].\end{aligned} \tag{B.34}$$

If $f(E_n) \stackrel{\text{def}}{=} p_n$ and the level spacing $\hbar\omega_{nm}$ is small, so that $\big[f(E_n) - f(E_m)\big] \,/\, (E_n - E_m) \approx \partial f_{\text{FD}}(E_n)/\,\partial E_n$ independent of m, Eq. (B.34) becomes

$$\tilde{\varphi}^{(\text{mp})}_{BA}(\omega) = -\frac{\omega}{i} \sum_n \frac{\partial f_{\text{FD}}(E_n)}{\partial E_n} \tilde{\Psi}^{(E_n)}_{BA}(\omega). \tag{B.35}$$

In the same limit of small $\hbar\omega_{nm}$ (and specifically if the density of states $g(E)$ is well defined) the sum over n can be replaced by integration over energy, leading to

$$\tilde{\varphi}^{(\text{mp})}_{BA}(\omega) = \int_0^\infty L^d g(E) f(E) \tilde{\varphi}^{(E)}_{BA}(\omega)\, dE = -\frac{\omega}{i} \int_0^\infty L^d g(E) \frac{\partial f_{\text{FD}}(E)}{\partial E} \tilde{\Psi}^{(E)}_{BA}(\omega)\, dE, \tag{B.36}$$

where L^d is the d-dimensional real space volume per particle, so that $L^d g(E)$ is the energy density of single-particle states.

B.2.3 Electrical Conduction

Kubo's linear response theory is directly applicable to the problem of electrical conduction. The complex conductivity tensor $\sigma_{\mu\nu}(\omega)$ relates the current density $J_\mu(t)$ in the μ-direction to the electric field $E_\nu(t) = \Re\big[E^{(0)}_\nu \exp(i\omega t)\big]$ in the ν-direction and obeys the relation

$$J_\mu(t) = \Re\left[\sigma_{\mu\nu}(\omega) E^{(0)}_\nu e^{i\omega t}\right]. \tag{B.37}$$

If $\hat{H} = \hat{H}_0 - \sum_i q_i \sum_\nu E_\nu(t) \hat{r}^{(i)}_\nu$ is the Hamiltonian of a system consisting of many particles with charges q_i and position operators $\hat{\mathbf{r}}^{(i)}$, while $\hat{A}$ is the polarization $\hat{\Pi}_\nu \stackrel{\text{def}}{=} \sum_i q_i \hat{r}^{(i)}_\nu$, Eqs. (B.16), (B.21) and (B.24) imply

$$\sigma_{\mu\nu}(\omega) = \frac{1}{i\hbar} \int_0^\infty \left\langle \big[\hat{\Pi}_\nu(0), \hat{J}_\mu(t)\big] \right\rangle_0 e^{-i\omega t}\, dt \tag{B.38}$$

$$= \frac{V}{\hbar\omega} \int_0^\infty \left\langle \big[\hat{J}_\nu(0), \hat{J}_\mu(t)\big] \right\rangle_0 e^{-i\omega t}\, dt, \tag{B.39}$$

where V is the total volume of the system. The last step follows from $\hat{J} = \dot{\hat{\Pi}}/V$, where $\dot{\hat{\Pi}} \stackrel{\text{def}}{=} (1/i\hbar)\big[\hat{\Pi}(t), \hat{H}_0\big]$, and the fact that $\tilde{\varphi}_{AB}(\omega) = (i/\omega)\tilde{\varphi}_{CB}(\omega)$ if $\hat{C}(t) = \dot{\hat{A}}(t)$.[4] If the system is in thermal equilibrium with a heat reservoir, by Eq. (B.32)

$$\sigma_{\mu\nu}(\omega) = \frac{V}{E_\beta(\omega)} \int_0^\infty \left\langle \big\{\hat{J}_\nu(0), \hat{J}_\mu(t)\big\} \right\rangle^{(\beta)}_0 e^{-i\omega t}\, dt, \tag{B.40}$$

[4] This can be verified by expressing both sides in terms of matrix elements.

which is one of the results commonly referred to as the *'Kubo formula'* of conductivity.[5] For non-interacting electrons, Eq. (B.39) becomes

$$\sigma_{\mu\nu}(\omega) = \frac{e^2}{\hbar\omega L^d} \int_0^\infty \left\langle \left[\hat{v}_\nu(0), \hat{v}_\mu(t)\right]\right\rangle_0^{(\mathrm{F})} e^{-i\omega t}\, dt \tag{B.41}$$

$$= \frac{e^2\hbar}{iL^d} \sum_{n,m} \frac{f_{\mathrm{FD}}(E_n) - f_{\mathrm{FD}}(E_m)}{E_n - E_m} \frac{\langle n|\hat{v}_\nu|m\rangle \langle m|\hat{v}_\mu|n\rangle}{\hbar\omega + (E_n - E_m) + i0^+} \tag{B.42}$$

where $f_{\mathrm{FD}}(E)$ is the Fermi-Dirac distribution. If the requirements for Eq. (B.36) are met, we may further approximate Eq. (B.41) as

$$\sigma_{\mu\nu}(\omega) = e^2 \int_0^\infty dE \int_0^\infty dt\; g(E) \left[-\frac{\partial f_{\mathrm{FD}}(E)}{\partial E}\right] \left\langle \left\{\hat{v}_\nu(0), \hat{v}_\mu(t)\right\}\right\rangle_0^{(E)} e^{-i\omega t}. \tag{B.43}$$

Starting from Eq. (B.43), one can calculate the semiclassical static conductivity $\sigma_{\mu\nu}$ of a two-dimensional electron gas at zero temperature as

$$\sigma_{\mu\nu} = \frac{m^* e^2}{\pi\hbar^2} \int_0^\infty \left\langle v_\nu(0) v_\mu(t)\right\rangle^{(E_\mathrm{F})} dt, \tag{B.44}$$

where $\left\langle \left\{\hat{v}_\nu(0), \hat{v}_\mu(t)\right\}\right\rangle_0^{(E_\mathrm{F})}$ has been replaced by the classical velocity autocorrelation function $\left\langle v_\nu(0) v_\mu(t)\right\rangle^{(E_\mathrm{F})} \stackrel{\mathrm{def}}{=} C_{\mu\nu}(t)$, which can be obtained numerically from simulated trajectories. The simplest way to take impurity scattering into account in such a calculation is to assume statistically independent scattering events that completely destroy any velocity correlation. If τ is the average time between two such scattering events, the probability that no scattering occurs in the time interval $[0, t]$ is $P(t) = \exp(-t/\tau)$, and the velocity autocorrelation function becomes $C_{\mu\nu}(t) = P(t)\check{C}_{\mu\nu}(t)$, where $\check{C}_{\mu\nu}(t)$ is calculated in the absence of impurity scattering [6].

Bibliography

[1] N. W. Ashcroft and D. N. Mermin. *Solid State Physics*, chapter 16. Harcourt College Publishers, Fort Worth, Texas, college edition, 1976.

[5] Kubo [5] defines a *canonical momentum* $\left\langle \hat{A}; \hat{B}\right\rangle \stackrel{\mathrm{def}}{=} (1/\beta)\int_0^\beta \left\langle \exp(\lambda\hat{H}_0)\hat{A}\exp(-\lambda\hat{H}_0)\hat{B}\right\rangle_0^{(\beta)} d\lambda$. Then $(1/i\hbar)\left\langle \left[\hat{A}(0), \hat{B}(t)\right]\right\rangle_0^{(\beta)} = \beta\left\langle \dot{\hat{A}}(0); \hat{B}(t)\right\rangle$ and Eq. (B.39) is written $V\beta\int_0^\infty \left\langle \hat{J}_\nu(0); \hat{J}_\mu(t)\right\rangle \exp(-i\omega t)\, dt$. In the classical limit, both $\left\langle \hat{A}(0); \hat{B}(t)\right\rangle$ and $\left\langle \left\{\hat{A}(0), \hat{B}(t)\right\}\right\rangle_0^{(\beta)}$ become the classical correlation function $\left\langle A(0)B(t)\right\rangle$.

[2] R. Kubo. Statistical-mechanical theory of irreversible processes. I. *Journal of the Physical Society of Japan*, 12(6):570–586, 1957.

[3] R. Kubo, M. Toda, and M. Hashitsume. *Statistical Physics II. Nonequilibrium Statistical Mechanics*, volume 31 of *Springer Series in Solid State Sciences*, chapter 4. Springer Verlag, Berlin, 1978.

[4] D. Cohen. Quantum pumping in closed systems, adiabatic transport, and the Kubo formula. *Physical Review B*, 2003.

[5] R. Kubo. The fluctuation–dissipation theorem. *Reports on Progress in Physics*, 29:255–284, 1966.

[6] R. Fleischmann, T. Geisel, and R. Ketzmerick. Magnetoresistance due to chaos and nonlinear resonances in lateral surface superlattices. *Physical Review Letters*, 68(9):1367–1370, March 1992.

C Numerical Calculation of Carrier Trajectories

> Mutationem motus proportionalem esse vi motrici impressae, et fieri secundum lineam rectam qua vis illa imprimitur.
>
> *(Sir Isaac Newton, Pricipia Mathematica)*

Electron trajectories in an arbitrary electrostatic potential $V(\mathbf{r})$ can be simulated by numerically integrating the semiclassical equations of motion [1]

$$\dot{\mathbf{r}} = \frac{1}{\hbar} \nabla_{\mathbf{k}} E(\mathbf{k}) \qquad \text{and} \tag{C.1}$$

$$\hbar\dot{\mathbf{k}} = -e\big[\mathbf{E}(\mathbf{r}) + \dot{\mathbf{r}} \times \mathbf{B}(\mathbf{r})\big], \tag{C.2}$$

where $\mathbf{r}$ and $\mathbf{k}$ denote position and crystal momentum, $E(\mathbf{k})$ is the energy dispersion, q the charge, $\mathbf{E}(\mathbf{r}) = -\nabla_{\mathbf{r}} V(\mathbf{r})$ the electric field, and $\mathbf{B}(\mathbf{r})$ the magnetic field. The equations of motion for electron holes with energy $E_{\mathrm{h}}(\mathbf{k})$ are obtained by replacing $E(\mathbf{k})$ by $-E_{\mathrm{h}}(\mathbf{k})$ in Eq. (C.1) and $-e$ by e in Eq. (C.2) [2].

Integrating the equations of motion of one or more particles is an ubiquitous problem, and a number of general algorithms have been developed for this purpose, particularly in the context of molecular dynamics simulations. For the present work I have chosen an approach based on Beeman's algorithm [3], which gives both position and momentum up to third order in the time step and provides for improved energy conservation compared to the

unmodified Verlet algorithm. In this scheme, the position and momentum at the next time step are

$$\mathbf{r}(t+\delta t) = \mathbf{r}(t) + \dot{\mathbf{r}}(t)\delta t + \frac{2}{3}\ddot{\mathbf{r}}\delta t^2 - \frac{1}{6}\ddot{\mathbf{r}}(t-\delta t)\delta t^2 \qquad \text{and} \tag{C.3}$$

$$\mathbf{k}(t+\delta t) = \mathbf{k}(t) + \frac{5}{12}\dot{\mathbf{k}}(t+\delta t)\delta t + \frac{3}{2}\dot{\mathbf{k}}(t)\delta t - \frac{1}{12}\dot{\mathbf{k}}(t-\delta t)\delta t. \tag{C.4}$$

If $\mathbf{B}(\mathbf{r}) \neq \mathbf{0}$, according to Eq. (C.2) the force and hence the momentum change $\dot{\mathbf{k}}$ depends on the velocity $\dot{\mathbf{r}}$. As $\dot{\mathbf{r}}(t+\delta t)$ is not known yet, one cannot determine $\dot{\mathbf{k}}(t+\delta t)$ in Eq. (C.4) directly. Instead, a predictor–corrector method is used, in which Eq. (C.4) is replaced by

$$\mathbf{k}_0(t+\delta t) = \mathbf{k}(t) + \frac{3}{2}\dot{\mathbf{k}}(t)\delta t - \frac{1}{2}\dot{\mathbf{k}}(t-\delta t)\delta t \qquad \text{and} \tag{C.5}$$

$$\mathbf{k}_{n+1}(t+\delta t) = \mathbf{k}(t) + \frac{1}{3}\dot{\mathbf{k}}(t+\delta t)\delta t + \frac{5}{6}\dot{\mathbf{k}}(t)\delta t - \frac{1}{6}\dot{\mathbf{k}}(t-\delta t)\delta t \tag{C.6}$$

and Eq. (C.6) is iterated until the correction $\mathbf{k}_{n+1}(t+\delta t) - \mathbf{k}_n(t+\delta t)$ falls below a predetermined threshold.

If $\mathbf{B}(\mathbf{r}) = (0, 0, B_z)$ independent of $\mathbf{r}$ and $\partial V(\mathbf{r})/\partial z = 0$, corresponding to a laterally modulated two-dimensional electron gas (2DEG) in a perpendicular magnetic field, Eqs. (C.1) and (C.2) can be written in terms of x, y, k_x, and k_y as

$$\dot{x} = \frac{1}{\hbar}\frac{\partial E(k_x, k_y)}{\partial k_x}, \qquad \dot{y} = \frac{1}{\hbar}\frac{\partial E(k_x, k_y)}{\partial k_y}, \tag{C.7}$$

$$\hbar\dot{k}_x = e\frac{\partial V(x,y)}{\partial x} - e\dot{y}B_z, \qquad \text{and} \qquad \hbar\dot{k}_y = e\frac{\partial V(x,y)}{\partial y} + e\dot{x}B_z. \tag{C.8}$$

A square lattice of antidots with period a can be represented by a phenomenological potential of the form [4]

$$V(x,y) = V_0\left[\cos\left(\frac{\pi x}{a}\right)\cos\left(\frac{\pi y}{a}\right)\right]^{2\alpha} \qquad \text{with} \quad \alpha \in \mathbb{Z}^+, \tag{C.9}$$

where $V_0 < 0$ and α is a parameter controlling the steepness. A weak potential modulation corresponds to $-eV_0 \ll E_\mathrm{F}$, where E_F is the Fermi energy, while $-eV_0 > E_\mathrm{F}$ implies the local depletion of the carrier sheet. If $-eV_0 \gg E_\mathrm{F}$, the intersection contours $-eV(x,y) = E_\mathrm{F}$

are approximately circular and one can identify the diameter of the antidots as $d \approx a\left[\frac{1}{2} - \sqrt[2\alpha]{E_\mathrm{F}/(-eV_0)}/\pi\right]$.

It is convenient to express Eqs. (C.7), (C.8) and (C.9) using dimensionless variables. The lattice constant gives the natural scale for the position, leading to $\tilde{x} \stackrel{\text{def}}{=} x/a$ and $\tilde{y} \stackrel{\text{def}}{=} y/a$, while the energy scale is set by E_F so that one can define $\tilde{U} \stackrel{\text{def}}{=} -eV/E_\mathrm{F}$. With these definitions Eq. (C.9) simplifies to

$$\tilde{U}(x,y) = \tilde{U}_0\left[\cos(\pi\tilde{x})\cos(\pi\tilde{y})\right]^{2\alpha}, \tag{C.10}$$

where $\tilde{U}_0 = -eV_0/E_\mathrm{F} \approx [\cos(\pi d/2a)]^{-2\alpha}$. The energy dispersion should be scaled accordingly as $\tilde{E} \stackrel{\text{def}}{=} E/E_\mathrm{F}$, while the momentum coordinates will be $\tilde{k}_x \stackrel{\text{def}}{=} k_x/k_0$ and $\tilde{k}_y \stackrel{\text{def}}{=} k_y/k_0$. If $k_0 = \sqrt{2m_\mathrm{c}E_\mathrm{F}/\hbar^2}$, where m_c is the cyclotron effective mass at E_F, and one further defines $\tilde{t} = t/t_0$ and $\tilde{B} = B_z/B_0$ with $t_0 = \sqrt{2m_\mathrm{c}a^2/E_\mathrm{F}}$ and $B_0 = 2\sqrt{2m_\mathrm{c}E_\mathrm{F}}/ea$, Eqs. (C.7) and (C.8) become

$$\frac{d\tilde{x}}{d\tilde{t}} = \frac{\partial\tilde{E}(\tilde{k}_x,\tilde{k}_y)}{\partial\tilde{k}_x}, \qquad \frac{d\tilde{y}}{d\tilde{t}} = \frac{\partial\tilde{E}(\tilde{k}_x,\tilde{k}_y)}{\partial\tilde{k}_y}, \tag{C.11}$$

$$\frac{d\tilde{k}_x}{d\tilde{t}} = -\frac{\partial\tilde{U}(\tilde{x},\tilde{y})}{\partial\tilde{x}} - 2\frac{d\tilde{y}}{d\tilde{t}}\tilde{B}, \qquad \text{and} \qquad \frac{d\tilde{k}_y}{d\tilde{t}} = -\frac{\partial\tilde{U}(\tilde{x},\tilde{y})}{\partial\tilde{y}} + 2\frac{d\tilde{x}}{d\tilde{t}}\tilde{B}. \tag{C.12}$$

For nearly free electrons, $E(k_x,k_y) = \hbar^2(k_x^2+k_y^2)/2m^*$, so that $m_\mathrm{c} = m^*$ and a well-defined Fermi wave number k_F and Fermi velocity v_F exist. The above definitions then imply $k_0 = k_\mathrm{F}$ and $t_0 = 2a/v_\mathrm{F}$, while $B_0 = 2m^*v_\mathrm{F}/ea$ is the field for which the cyclotron radius $R_\mathrm{c} = m^*v_\mathrm{F}/eB_z$ equals $a/2$.

For electron holes, Eqs. (C.11) and (C.12) need to be modified as described above, so that one now has

$$\frac{d\tilde{x}}{d\tilde{t}} = -\frac{\partial\tilde{E}(\tilde{k}_x,\tilde{k}_y)}{\partial\tilde{k}_x}, \qquad \frac{d\tilde{y}}{d\tilde{t}} = -\frac{\partial\tilde{E}(\tilde{k}_x,\tilde{k}_y)}{\partial\tilde{k}_y}, \tag{C.13}$$

$$\frac{d\tilde{k}_x}{d\tilde{t}} = -\frac{\partial\tilde{U}(\tilde{x},\tilde{y})}{\partial\tilde{x}} + 2\frac{d\tilde{y}}{d\tilde{t}}\tilde{B}, \qquad \text{and} \qquad \frac{d\tilde{k}_y}{d\tilde{t}} = -\frac{\partial\tilde{U}(\tilde{x},\tilde{y})}{\partial\tilde{y}} - 2\frac{d\tilde{x}}{d\tilde{t}}\tilde{B}\cdot \tag{C.14}$$

If the hole band $E_\mathrm{h}(\mathbf{k})$ is given by Eq. (5.8), the reduced energy $\tilde{E}(\tilde{k}_x,\tilde{k}_y)$ becomes

$$\tilde{E}(\tilde{k}_x,\tilde{k}_y) = \frac{E_\mathrm{g}}{E_\mathrm{F}} - 2\frac{m_\mathrm{c}}{m_\mathrm{e}}\left[|A|\left(\tilde{k}_x^2+\tilde{k}_y^2\right) - \sqrt{B^2\left(\tilde{k}_x^2+\tilde{k}_y^2\right)^2 + C^2k_x^2k_y^2}\right], \tag{C.15}$$

where—as before—I take the anisotropy parameters to be $A = -14{\cdot}3$, $B = -10{\cdot}4$, and $C = 13{\cdot}5$ [5]. With this dispersion relation, one finds $m_c \approx 0{\cdot}185 m_e$ independent of energy. The scale B_0 then identifies the magnetic field for which the unperturbed cyclotron orbit has the same area as a circle of radius $a/2$.

Bibliography

[1] N. W. Ashcroft and D. N. Mermin. *Solid State Physics*, chapter 12. Harcourt College Publishers, Fort Worth, Texas, college edition, 1976.

[2] C. Kittel. *Introduction to Solid State Physics*, chapter 7. John Wiley & Sons, New York, seventh edition, 1996.

[3] D. Beeman. Some multistep methods for use in molecular dynamics calculations. *Journal of Computational Physics*, 20(2):130–139, February 1976.

[4] R. Fleischmann, T. Geisel, and R. Ketzmerick. Magnetoresistance due to chaos and nonlinear resonances in lateral surface superlattices. *Physical Review Letters*, 68(9):1367–1370, March 1992.

[5] *Landolt-Börnstein Numerical Data and Functional Relationships in Science and Technology: New Series: Group III: Crystal and Solid State Physics*, volume 17. Springer, 1982.

D Physical Constants

Πάντες ἄνθροποι τοῦ εἰδέναι ὀρέγονται φύσει.

(Aristotle, Metaphysics)

Quantity	Symbol	Value		Unit
speed of light in vacuum	c	2·997 924 58	$\times 10^{8}$	m/s
magnetic constant $4\pi \times 10^{-7}$	μ_0	1·256 637 061 4⋯	$\times 10^{-6}$	N/A^2
electric constant $1/\mu_0 c^2$	ε_0	8·854 187 817⋯	$\times 10^{-12}$	F/m
Planck constant	h	6·626 069 3(1 1)	$\times 10^{-34}$	J s
$h/2\pi$	$\hbar$	1·054 571 68(18)	$\times 10^{-34}$	J s
elementary charge	e	1·602 176 53(14)	$\times 10^{-19}$	C
Dirac flux quantum h/e	$\Phi_0^{(\mathrm{D})}$	4·135 667 44(36)	$\times 10^{-15}$	Wb
conductance quantum $2e^2/h$	G_0	7·748 091 733(26)	$\times 10^{-5}$	S
von Klitzing constant h/e^2	R_K	2·581 280 744 9(8 6)	$\times 10^{4}$	Ω
electron mass	m_e	9·109 382 6(1 6)	$\times 10^{-31}$	kg
Bohr magneton $e\hbar/2m_\mathrm{e}$	μ_B	9·274 009 49(80)	$\times 10^{-24}$	J/T
Bohr radius $4\pi\varepsilon_0\hbar^2/m_\mathrm{e}e^2$	a_0	0·529 177 210 8(1 8)	$\times 10^{-10}$	m
Boltzmann constant	k_B	1·380 650 5(2 4)	$\times 10^{-23}$	J/K

Bibliography

[1] P. J. Mohr and B. N. Taylor. CODATA recommended values of the fundamental physical constants: 2002. *Reviews of Modern Physics*, 77:1–107, January 2005.

E Typography and Notation

> Anyone who would letterspace blackletter would shag sheep.
>
> *(Frederic Goudy)*

This thesis has been typeset by the TEX typesetting program [1] with the help of the LATEX document preparation system [2]. The software is freely available from a number of sources, including the Comprehensive TEX Archive Network (CTAN) [3].

The mathematical and physical notation follows ISO 31 [4], with the exception of literal numbers, which are typeset according to traditional British conventions. The *Système International d'Unités* (SI) [5] has been used throughout for physical units; in particular, the definition of electromagnetic constants follows this standard, and the fact that some quantities are given in terms of centimetres (cm) does not imply the use of cgs units. In addition, some length scales are quoted in ångströms (Å); while the ångström is not an SI unit, it is directly related to the metre ($1\,\text{Å} = 10^{-10}\,\text{m}$) and approved for use with the SI.

I have tried to follow the usual conventions for formula symbols in so far as they are clearly established and do not conflict with each other. However, care has been taken to define all such quantities—possibly by referring to a figure—where they are first used; this excludes the fundamental physical constants listed in Appendix D. Superscripts that do not represent exponents have been enclosed in parentheses everywhere in this thesis to avoid ambiguity.

Bibliography

[1] D. E. Knuth. *The TEXbook*. Addison-Wesley, Reading, Massachusetts, 1984.

[2] L. Lamport. *LATEX. A Document Preparation System*. Addison-Wesley, Reading, Massachusetts, second edition, 1994.

[3] The comprehensive TEX archive network. http://www.ctan.org/, Februrary 2006.

[4] International Organization for Standardization. *ISO Standards Handbook: Quantities and units*. International Organization for Standardization, Geneva, Switzerland, 1993.

[5] Organisation intergouvernementale de la Convention du Mètre. *Le Système international d'unités (SI)*. Bureau international de poids et mesures, seventh edition, 1998.

Index