Design of Digital Instrumentation for Scanning Probe Microscopy

A thesis presented in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

in

Physics

at Massey University, Palmerston North,
New Zealand

Henning Albrecht Michael Klank
1999

Abstract

A scanning tunneling microscope with a focus on digital instrumentation has been built. The aim of this project was to allow a digital signal processor full control over all essential microscope variables, especially simultaneous control of the vertical and horizontal tip position.

Due to the fact that its operation is controlled by software, this system offers convenient operation and considerable flexibility, allowing different modes of operation, such as topographical and spectroscopic scans. Presently this microscope is the only one in New Zealand that allows the operator full software control over the tip position and bias voltage, thereby allowing it to become a powerful research tool.

Atomic scale images on graphite were successfully recorded. The spatial resolution of the microscope was estimated to be 5 pm vertically and 40 pm horizontally. Two different imaging methods were demonstrated on a gold sputtered TEM grating with a scan area that was larger than $4\,\mu\text{m} \times 4\,\mu\text{m}$. One method has variable horizontal scan speed, while the other method can possibly be used for nanolithography. Both show the flexibility of this system.

Although digital electronics is often perceived as being slower and noisier than analog electronics, in this instrument it did not decrease the data acquisition speed nor did it reduce the signal-to-noise ratio. The bandwidth of the closed-loop controlled microscope is currently about 1 kHz, limited by the bandwidth of the current-to-voltage converter, an analog component. The resolution is limited by the large gain of the high-voltage amplifiers used to drive the actuators. With a faster current-to-voltage converter and a reduced high-voltage amplifier gain, a bandwidth of 8 kHz should be possible with a vertical resolution of less than 2 pm and a horizontal resolution of 10 pm.

Acknowledgements

Many people contributed in one way or another to the completion of this thesis.

Firstly I would like to mention the three supervisors of this project. Dr. Blair Hall initiated this project and made me come to New Zealand in the first place. I thank him for this and his good sense of humor. Dr. Craig Eccles looked after me for the last two years. I had the pleasure to work with an incredibly patient man. Associate Professor Robert O'Driscoll was also the course controller of two electronics papers where I had a teaching role. I have learnt a great deal from him regarding electronics and how handle students by telling stories.

I would like to express my gratitude for the people dwelling across the corridor in the rooms of the electronics services. Udo von Mulert explained to me the world of scientific instruments and how to repair all sorts of things. Peter Lewis built almost all of the second generation electronics and was always there to help and listen. Some day we will go and see the blue duck. Robin Dykstra was always the person to ask in complicated electronics design matters. I thank him especially for the word nano-robot and for giving me the right hint for fixing the ADC module. Keith Whitehead should also be mentioned as a source of incredible knowledge. Gerard Harrigan, who used to work at electronics services, showed me a few tricks for laying out printed circuit boards and supplied me with thyme and chives for the garden.

Our student, Mark Hunter, was the physicist in the Matlab cockpit, creating several graphs. We discussed physics and other things together at all sorts of times during day and night.

I would also like to thank the mechanics, Steve Denby, Noel Foot and Barry Evans for building various parts of the microscope. Kit Clark from Kaycee Technology built and modified several parts of our probe head assembly.

On a more personal level I would like to acknowledge the support of many good friends.

Lise cared for me in many ways. We went together through thick and thin. I thank her for her wits and her smile.

Jörg would supply us with *Wein* in the crucial moments. Martin, our long-term flatmate, discussed aspects of thesis writing with me in the long dark winter months. Luckily there is a summer after winter. I also have to thank our earlier flatmates, especially Andrew, with whom I discussed many a problem from home-made philosophy to waste-water ponds.

Last but not least I would like to thank my family, who were far away in space but not in spirit. My brother Thilo presented me with a few selected paintings, my father Eckart was always able to entertain us all and my mother Johanna was the soul of the family.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Scanning Probe Microscopy

1.1.1 Near-Field Microscopes

One can divide all microscopes into two categories, far-field microscopes and near-field microscopes. Most microscopes known before the invention of the scanning tunneling microscope [28] were of the far-field variety. Because these instruments use a wave phenomenon for image formation their resolution is diffraction limited. In other words their spatial resolution is limited by the wavelength of the radiation used.

In scanning probe microscopy one generates an image of a sample by looking at short range interactions, or the near-field of long-range effects. Because of this property a scanning probe microscope cannot image a complete sample at once. In fact, it can look only at one point at a time, due to the very short range of the physical interactions. The advantage of this technique over more traditional imaging methods is the considerable improvement in spatial resolution that can be obtained.

1.1.2 Short-Range Interaction

Scanning probe microscopes exploit many different short range interactions. Although the instrumentation described in this thesis is designed for a generic scanning probe microscope, thus far it has only been used with the scanning tunneling microscope, which uses the quantum mechanical tunneling effect. In what follows it will be assumed that the scanning tunneling microscope actually implemented is referred to, unless stated otherwise.

The tunneling microscope probes a conducting sample surface with a very sharp piece of metal wire, referred to as the tip. The tunneling effect occurs when the tip and the sample are placed in very close proximity – usually a distance of less than 1 nm. The tip must be very sharp to limit the spatial extent of the interaction, thereby gaining high lateral resolution.

Figure 1.1 shows the tip and the sample. If tip and sample come very close there will be a certain probability for electrons to tunnel through the potential barrier which separates them. When a certain bias voltage is applied between the tip and the sample, a net tunneling current results. This electron current

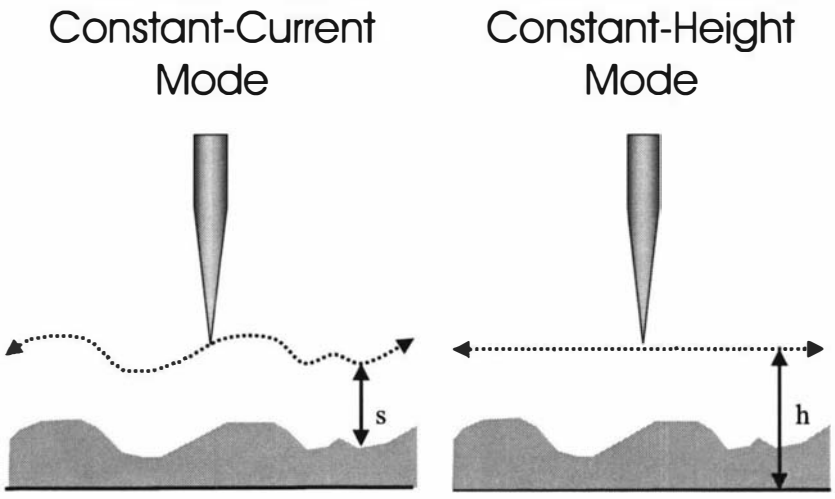


Figure 1.1: **Image Acquisition Modes:** Two different modes of scanning the sample surface with the tip are shown. The mode on the left side keeps the tunneling current constant while moving across the surface. Hence the tip must keep a constant distance (s) between it and the sample. In constant-height mode, on the right, the tip does not follow the sample contours, but keeps the same height h relative to the sample holder. In both modes, constant bias voltage is assumed. The constant-current mode requires closed-loop control and is the slower method. However, since it keeps a constant distance between tip and sample it is also the safer mode with regards to tip-sample collision.

constitutes the signal which is measured by the tunneling microscope. Typically the tunneling current is very small, usually less than 1 nA.

The need for close tip-sample proximity is due to the exponential relationship between tip-sample distance and the tunneling current. As a rule of thumb, the current decreases by a factor of 10 as the tip-sample distance is increased by 1 Å (0.1 nm) [44].

1.1.3 Image Generation

In scanning tunneling microscopy there are two modes commonly used to get an image of a sample, constant-current mode and constant-height mode. A graphical depiction of the modes is provided in figure 1.1.

In constant-current mode the tip follows the contours of the sample surface. The tip keeps a certain distance s above the surface by maintaining a set current level. Consequently the signal is constant but the tip height h changes and this change is recorded.

In constant-height mode, the height h of the tip relative to the sample holder is kept constant. The separation s between the tip and sample is changing, and therefore the signal from the tip is changing. The changing signal from the tip is recorded.

Moving in a line over the surface in either mode and recording tunneling current or height data is called a line-scan. An image is formed by doing several line-scans in succession and displaying the signal strength at each point by some means, for example as the brightness of a pixel.

Both modes scan the sample topographically, i.e. they map a relief of the surface. For a topographical scan the bias voltage between tip and sample re-

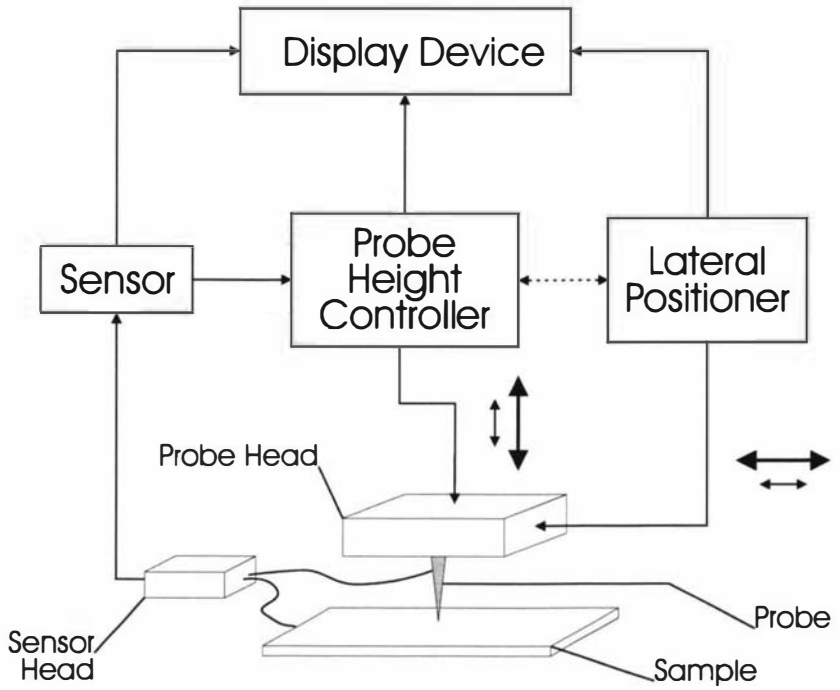


Figure 1.2: **Scanning Probe Microscope:** There are three major subsystems that make up a scanning probe microscope, a sensor, a probe height controller and a lateral positioner. The probe has to be approached from millimeter distances to within nanometers of the sample. The large and small arrows indicate the approach mechanisms which act at different length scales. In a similar fashion the probe is usually movable over large and small distances. The tip-sample interaction is recorded by the sensor. Either the sensor signal is shown directly on the display device or the height control signal is.

mains constant. Instead of mapping the contours of the sample topographically, the microscope can also be used to perform a spectroscopic investigation of the underlying sample. Spectroscopy differs from topographical imaging in that the scanning tip remains at a fixed horizontal position over the sample, while the bias voltage V_{bias} or the tip height z is varied. Either the bias voltage is kept constant, while the tip height is changed, or vice versa.

1.2 Computer Control for SPM

The microscope described in this thesis differs from an analog microscope in that it employs a digital signal processor (DSP) as the central component of the microscope control electronics. The history of computer-controlled scanning probe microscopes is explored in this section.

A scanning probe microscope is an electronic system that incorporates three different subsystems, a high-precision sensor, a sophisticated probe height controller and a lateral probe positioner. The three subsystems of a scanning probe microscope are shown in figure 1.2. The probe height controller advances the probe from distances of several millimeters down to within atomic distances of the sample, at which point the probe to sample distance is controlled by a closed-loop feedback system. The lateral positioner makes examination of a

certain scan area possible, while the system sensor measures the strength of the interaction between probe and sample. The output of the system is a display device.

As described by Aguilar *et al.* [2], early scanning probe microscopes were mainly operated manually using an analog x-y plotter as a display device. Even these early microscopes would not have been possible without electronic feedback control for the probe-sample distance and lateral probe positioning on a sub-micrometer scale. With the advancement of technology there has been a tendency to create a layer between the microscope operator and machine that simplifies use of the microscope, while increasing its flexibility. This intermediate layer consists of computer-controlled analog and digital electronics. More and more of the purely mechanical functions of the microscope have been replaced by electro-mechanical ones and within the field of electronics there has been a trend from purely analog circuits to those which are more digitally based. Since processors have become faster, smaller and cheaper, one or more processors can now be incorporated into the different subunits of the microscope. Ultimately, depending on the emphasis of the research group, the computer-controlled SPM can either be turned into an intelligent sensor system, as discussed in the book by Brignell and White [34], or it can become a nano-robot.

Automation began at the operator side of the microscope, the display device. Two of the first research groups reporting computerized microscopes were Aguilar *et al.* [2] and Schroer and Becker [169]. Both groups replaced the analog x-y plotter with a digital image display and provided computer assisted image processing. The aim was to unite the analog microscope controller with data acquisition and real-time display.

The next instrument that was replaced by a computer was the lateral probe positioner, i.e. the scan generator. While Aguilar *et al.* still used a function generator to provide the saw-tooth waveform needed for the fast scan direction, Schroer and Becker were already using their computer to replace the function generator, albeit with a rather coarse resolution of 5 mV. Cutkosky [52] presented an especially versatile digital scan generator.

Among the coarse probe height controllers, that is, those which advance the probe into the domain of the electronic distance control, several mechanisms were developed very early on to allow complete electronic control, although mechanical ones remained popular for a long period [84]. Most of the electronic methods involve a kind of piezoelectric walker [65, 130, 185] or motor, such as the Besocke design [25], the 'louse' of Binnig and Rohrer [27] or the 'Inchworm' of Burleigh Instruments [38].

The atomic-scale probe height control was addressed last in the development towards increased automation. This is mainly due to the issue of speed, which can be important for scanning probe microscopes. Early microscopes were limited by the speed of the analog x-y plotter [2] and were in general rather slow. As Pohl [156] has pointed out, the highest possible speed for a scanning probe microscope will most likely be limited to the lowest mechanical resonance of the probe-carrying actuator assembly. Pohl expects that this resonance frequency can reasonably be raised to values between 30 and 100 kHz and that the bandwidth of the microscope can then be between 10 and 30 kHz. If such a high bandwidth is desired then the electronics has to be sufficiently fast. Analog height controllers, based on operational amplifiers, are usually faster than digital ones. Supervisory control of an analog controller allows this speed advantage

to be kept and was reported by Cutkosky [52], Altman *et al.* [5], Rong *et al.* [164] and more recently by Göhde *et al.* [76] and Nakakura *et al.* [141]. The supervisory control can be as simple as enabling or disabling the analog controller. The digital controller of Göhde *et al.* is more advanced in that it can tune the analog controller.

The substitution of an analog controller with a digital one is termed direct digital control. On a level between supervisory and direct digital control is the implementation of a hardwired digital inner loop supervised by a computer. Hardwired digital electronics is more complicated to develop and build than running a programmed microprocessor, but it also results in a faster controller. The groups of Robinson *et al.* [163] and Kim *et al.* [114] used such a hard-wired digital controller.

With direct digital control, the controller is implemented entirely in the software of the processor. Despite the speed disadvantage, a digital controller that interacts directly with the microscope hardware offers greater flexibility, especially if the height control is to be interrupted to perform spectroscopic examination of the sample. The first group to publish the implementation of direct digital control were Piner and Reifenberger [154] using general purpose processors, a 68020 from Motorola together with a 68881 coprocessor. Other groups also used general purpose processors, for instance Barchesi *et al.* [18], Müller *et al.* [138], Munoz *et al.* [139] and Wilms *et al.* [191].

To solve the speed problem, faster processors such as digital signal processors (DSP) can be used in direct digital control. Operating scanning probe microscopes under DSP control have been demonstrated by the research groups of Wong and Welland [194], Clark *et al.* [46], Heuell *et al.* [93], Behler *et al.* [23], Paillard *et al.* [149], Scheuring *et al.* [167] and Wilms *et al.* [192] in chronological order. The advantages of direct digital control are discussed in more detail below.

A future option for microscope control, which will in all likelihood be explored, is to support a fast processor with other processors and assign dedicated tasks to them. This has already been suggested by Clark *et al.* [46].

1.3 DSP-Controlled Microscope

The digital signal processor at the heart of the microscope built during this project has complete control over the position of the tip. This microscope can serve as a research tool that allows the operator to position the tip at any position with respect to the sample with a precision of less than 0.1 nm, as shown in section 7.5.2. At the present time it is the only scanning probe microscope in New Zealand with this capability.

The electronic hardware provides the input and output signal channels for this software so that it can interact with the mechanical hardware. Like all instruments that are software controlled, the microscope is intrinsically flexible and versatile in its operation. The interface provides an additional advantage; convenient microscope handling for the operator. The diagram in figure 1.3 shows the two channels of data exchange between the DSP and probe head and between DSP and PC. The data exchange channel to the right includes the software based control loop. All major variables are available to the software and can be changed independently. The microscope adjustable variables are the

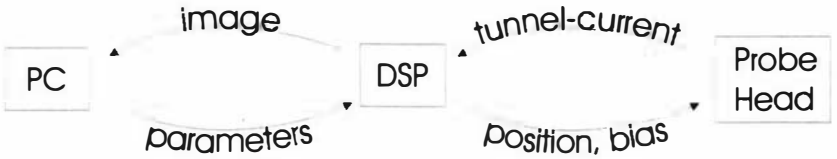


Figure 1.3: **Interaction between PC, DSP and Probe Head:** The DSP assumes a central position between PC and probe head. Not only does the DSP control the the tip-sample distance, but it also exchanges image and control data with the PC. The interaction labels indicate the main purpose of the link. Note that the DSP also controls the horizontal position and the coarse movement of the probe head.

bias voltage, horizontal tip position and vertical tip position.

The tip can be used in the conventional way to investigate the underlying sample with a raster type scan. However, this is not the only feature of this microscope. Since the operation of the microscope depends on software, the microscope is intrinsically flexible, i.e. it is adaptable to changing circumstances, and also versatile, i.e. it may find more than one use.

1.3.1 Height Control

The DSP adds functionality to the height control. Software control of the tip-sample distance allows more advanced control algorithms to be implemented, thereby promising more accurate control or a higher bandwidth. The controller does not have to be linear. Safety checks, testing actuator saturation and non-linear functions such as a logarithmic conversion are easily implemented. Using a fast processor like a DSP ensures that there is only a small penalty in reaction speed. To provide an opportunity for more accurate imaging of the surface it is also possible to account and correct for piezoelectric actuator non-linearities.

1.3.2 Spectroscopy

The microscope described here is used in air. Spectroscopy with a clean tip and sample is therefore not an option, since molecules from the surrounding air will form adsorbates on both the tip and the sample. Additionally, the tip material, tungsten, used so far in this microscope, oxidizes. However, the same circumstances open up different opportunities for this microscope. For example, phenomena can be studied that occur only in an air environment, such as the effects of adsorption or desorption of molecules, since this results in $1/f$ noise [129]. An issue of broader interest is surface manipulation in air, which is discussed below.

The DSP-controlled microscope is flexible in regard to the type of sample examination that can be applied. The continuous control of the tip-sample distance can be temporarily suspended, which means that sample investigation at one point over the sample, such as spectroscopy, is easily achieved. All essential variables of the tip-sample system; the bias voltage and the horizontal and vertical tip position, can be accessed independently by the DSP-controlled microscope. A topographical scan can easily be combined with spectroscopic investigations, as is done in current imaging tunneling spectroscopy (CITS) [86], a method mentioned in section 2.1.2. Other combinations are also possible, such as measuring the apparent barrier height at selected places. With an analog

microscope this would involve a more elaborate scheme, since the analog tip height controller would have to be interrupted in hardware [164].

1.3.3 Convenient Service Tool

The DSP-controlled microscope can be handled conveniently with software instructions, by having a PC connected to the DSP. The recorded image is visualized on the computer screen and the operator can adjust control variables such as scan frequency, image area and orientation with computer software. The operator does not have to go through the process of selecting a stored image to be digitized and digitize it for further image processing as all images are already digitized by default.

The DSP-controlled microscope can become a convenient service tool for other researchers. The software on the DSP provides an extra layer between researcher and microscope, which allows this microscope to become an intelligent sensor system. The researcher need only define the desired scan path, operating conditions and information to be collected, while the DSP handles all the details. For instance, a certain location within a sample may be of interest and could be imaged in detail by simply selecting a region around this point. Sometimes the first or second derivative of the measured current might hold valuable information [138]. Certain samples are very delicate and so may need to be imaged in a nonstandard way to prevent damage.

The DSP can also provide quality control of the tip-height feedback mechanism or data collection process by passing figures of merit on to the operator, such as the standard deviation of the collected data. Remaining at one sample point, until the data set has a minimum statistical quality, is a well known method for ensuring quality data are collected.

Finally one should be able to exchange the currently used probe head, so that the DSP-controlled microscopy system can be reconfigured to work as another type of scanning probe microscope, such as an AFM. This could be done with comparative ease, since the closed-loop control can be established with any other physical effect that establishes an interaction between tip and sample surface.

1.3.4 Position Control

Wong and Welland [194] discuss the two functions of tip height control and scan generation. They point out that in early systems these functions were provided by analog electronics and were kept separate, while later instruments had digital scan generators, although only some designs were realized implementing both functions digitally. This is the case in their microscope as well as in the one described here, giving rise to additional features.

Since both vertical and horizontal tip movement are subject to software control, they may interact. For example the imaging path of the tip need not be a conventional raster scan with fast x-scan and slow y-scan directions with independent tip height (z) control. Line scans with variable scan speeds have already been described by Piner and Reifenberger [154], Wong and Welland [194], Heuell [94] and Robinson *et al.* [163]. This is but one way of using the power of positioning control. For instance the microscope may be programmed to follow paths on the surface which have constant height, thereby producing a contour map. Such a scheme was proposed by Pohl and Möller [155], who used an analog

microscope. Direct digital control makes it easier to implement new scan modes and adapt them to the sample under investigation.

1.3.5 Novel Imaging Modes and Nano-Manipulation

Apart from imaging, the microscope can be used to manipulate the sample on a nanometer scale. This microscope is designed to be able to supply comparatively high bias voltages to the sample, which will result in non-imaging tunneling currents. The tunneling electrons are accelerated in the electric field and impact with high velocity on the sample, where they can cause a change in its physical or chemical make-up. Full digital control allows the operator to place the tip at any spot within the actuator range, manipulate the sample there and subsequently monitor the change. The microscope can change the sample locally at specific points to create patterns or structures on the nanometer scale. This microscope is especially suited to be a nanomanipulator, since apparently the changes in the sample, such as hillocks or moulds, depend on the presence of atmospheric components, presumably water vapor and oxygen [126, 159].

Obviously the processor needs to be able to control the x, y and z motion of the tip to do that. Very few previously presented microscopes have control over all axes by a single processor. The ones that have, Piner and Reifenberger [154], Wong and Welland [194], Heuell [94], Robinson *et al.* [163], Müller *et al.* [138] and Baselt *et al.* [20], have not yet used the full potential that complete position control offers. Some groups, such as Morgan and Stupian [134], Paillard *et al.* [149], Scheuring *et al.* [167], Wilms *et al.* [192], employed a DSP only to replace the conventional analog tip height controller. Wilms *et al.* [191] even went a step back and reverted from using a DSP for height control to using an analog controller. This is not surprising, since DSPs are more costly and more difficult to use than a small amount of analog circuitry. DSPs will show their real power when they have control over the complete tip movement as well as the bias voltage. Paillard *et al.* specifically mention that a design is on its way to integrate their height controlling DSP with horizontal tip control.

Since this microscope has been built fairly recently, our research group has only just begun to scratch the surface of all possibilities. What is presented here is the description of a potentially powerful research tool.

Chapter 2

Probe Microscopy

2.1 Theoretical Background

2.1.1 One Dimensional Tunneling

The tunneling effect is usually explained with the aid of a simple one-dimensional model. However the theory also has to explain how a two-dimensional image is obtained. The standard model of scanning tunneling theory will be introduced to explain the lateral dependence of the tunneling effect.

The tunneling current depends on the electronic properties of both the sample and the microscope tip. Consider two metals, metal 1 and 2, which are not yet in close contact as shown in figure 2.1 to the left. Each metal has a specific Fermi energy (E_F) and work function (Φ) that depend on the type of metal. In the following discussion both the Fermi energy and work function are arbitrarily chosen for metal 1 and 2. In practice the sample is often chosen such that the relative work functions leave the sample negative with respect to the tip. In this case metal 1 corresponds to the sample and metal 2 to the tip. When the metals come into contact, for example by a conducting wire, then the Fermi energies level out, $E_{F1} = E_{F2}$, and as a result a contact potential

$$\Delta\Phi = \Phi_1 - \Phi_2$$

exists between both metals [13]. The Fermi energies become equal because charges are transferred from one metal to another. The charge exchange can also occur when the separation distance between the metals becomes very small. The one-dimensional model of the two metals separated by a distance s is shown in figure 2.2, where s is the separation between the closest nuclei of the tip and sample. In this case electrons 'tunnel' through the potential barrier

$$\Phi \approx \frac{\Phi_1 + \Phi_2}{2} \tag{2.1}$$

between the two materials, which is indicated in figure 2.2. The one-dimensional stationary Schrödinger equation for an electron in the potential $U(z)$ is

$$(-\frac{\hbar^2}{2m}\frac{\mathrm{d}^2}{\mathrm{d}z^2} + U(z))|\Psi\rangle = E|\Psi\rangle.$$

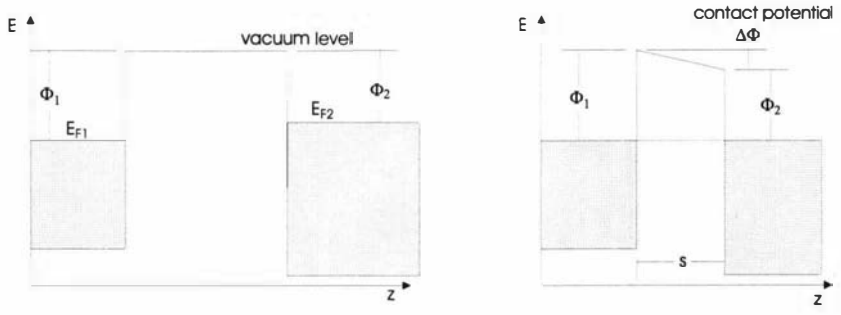


Figure 2.1: **Energy Levels of Two Metals:** Both diagrams show a one-dimensional model of two metals with their respective Fermi energies E_{F1} and E_{F2} and work functions Φ_1 and Φ_2. The left diagram shows the energy levels at a large distance. The diagram to the right shows the two metals in close contact with a barrier distance s, which is defined as the distance between the nearest nuclei of the tip and sample. In scanning tunneling microscopy the distance s is usually smaller than 1 nm. The slope of the barrier is due to the contact potential $\Delta\Phi = \Phi_1 - \Phi_2$.

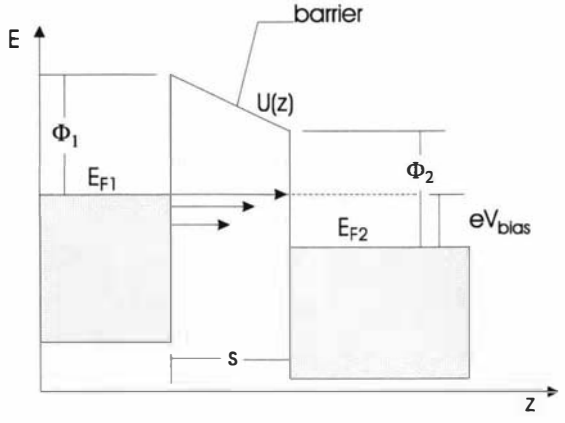


Figure 2.2: **Energy Levels:** The diagram shows the same model as in figure 2.1, now with an additional bias voltage applied between the metals. The electrons at higher energy levels can tunnel through the barrier into empty levels. More than one electron state can and will be involved while tunneling, which is indicated by arrows. The electrons can also lose energy within the tunneling barrier, to adsorbed molecules for instance, and arrive at a lower energy level.

If the barrier height is constant, then the solution of the Schrödinger equation within the barrier is

$$\Psi(z) = \Psi(0)e^{-\kappa z}$$

where κ is defined by

$$\kappa = \frac{\sqrt{2m(U-E)}}{\hbar}.$$

The wavefunction $\Psi(z)$ depends on the z direction defined in figure 2.2. The barrier Φ is known as the effective barrier height between the metals and the barrier shown in figure 2.2 is only an approximation. When the Fermi levels are equal, the metals are in equilibrium and there is no net charge transfer. However, if an external voltage V_{bias} is applied between the metals, then a certain electron current will tunnel from one metal to the other. Consider an electron in the negative metal (metal 1) described by the state $|\Psi_\mu\rangle$ with an energy E_μ. If E_μ lies in the energy window between the Fermi energy (E_{F1}) and the smaller energy $E_{\text{F1}} - eV_{\text{bias}}$, then the electron can tunnel to the positive metal (metal 2) with a probability that is proportional to

$$\langle\Psi_\mu|\Psi_\mu\rangle = |\Psi_\mu(0)|^2 e^{-2\kappa s}.$$

The decay constant κ is given by

$$\kappa = \frac{\sqrt{2m\Phi}}{\hbar}, \tag{2.2}$$

where the effective barrier Φ from equation (2.1) is used [44]. The barrier height is around $\Phi = 4\,\text{eV}$ for the metals used in scanning tunneling microscopy, which results in a typical decay constant of $\kappa = 1.0 \times 10^{10}\,1/\text{m}$. The tunneling current will be proportional to the number of states in metal 1 that have an energy within the energy window defined above. The proportionality is expressed as [44]

$$I_t(s) \propto \sum_{E_\mu = E_{\text{F1}} - eV_{\text{bias}}}^{E_{\text{F1}}} |\Psi_\mu(0)|^2 e^{-2\kappa s}. \tag{2.3}$$

Under the conditions that the states $|\Psi_\mu\rangle$ do not vary considerably within the energy window, equation (2.3) can be expressed by substituting for the local density of states (ρ_1) of metal 1, which is defined as [44]

$$\rho_1(z, E) = \frac{1}{\epsilon} \sum_{E_\mu = E-\epsilon}^{E} |\Psi_\mu(z)|^2, \tag{2.4}$$

for a small energy interval ϵ. The local density of states describes the number of states per unit energy and unit volume at energy E and at location z. The local density of states can of course also be defined for three dimensions. The condition is usually true for small bias voltages. Hamers [85] defines small voltages by

$$eV_{\text{bias}} < \Phi$$

corresponding to $V_{\text{bias}} = 10\,\text{mV}$ to $4\,\text{V}$.

Substituting equation (2.4) into equation (2.3), using $\epsilon = eV_{\text{bias}}$, shows that the tunneling current (I_t) is proportional to the local density of states, i.e.

$$I_t \propto V_{\text{bias}}\, \rho_1(0, E_{\text{F1}})\, e^{-2\kappa s}. \tag{2.5}$$

The following conclusions can be drawn from equation (2.5):

- The tunneling current falls exponentially with increasing tip-sample distance (s).
- The tunneling current is proportional to the local density of states at the Fermi level (E_{F1}).
- The tunneling current is proportional to the bias voltage V_{bias}. This is only true for small bias voltages, since for higher voltages field emission will begin. This is discussed below.

In the following chapters the tunneling current from equation (2.5) is written as

$$I_t = V_{bias} \Lambda e^{-\frac{s}{s_0}}, \tag{2.6}$$

where Λ is a constant with the dimensions of a conductance, while s_0 is clearly related to the decay constant κ by

$$s_0 = \frac{1}{2\kappa}. \tag{2.7}$$

Using the value for κ given above we find that:

$$s_0 \approx \frac{1}{2 \times 1.0 \times 10^{10}\ {}^1/_{\mathrm{m}}} = 0.05\,\mathrm{nm}.$$

Apparent Barrier Height

The proportionality between tunneling current and bias voltage can be used to define a tunneling conductance (G_t) as [144]

$$G_t = \frac{I_t}{V_{bias}}. \tag{2.8}$$

Olesen *et al.* [144] use the tunneling conductance and equation (2.5) to define an apparent barrier height (Φ_{ap}) as

$$\Phi_{ap} = \left(\frac{1}{2} \frac{\hbar}{\sqrt{2m}} \frac{\partial \ln G_t}{\partial s} \right)^2, \tag{2.9}$$

where the decay constant κ from equation (2.2) has been used. The apparent barrier height can be determined experimentally if the tunneling current is recorded as a function of the tip-sample distance s.

2.1.2 Standard Model

The s-wave model of Tersoff and Hamann [181] is the standard model for scanning tunneling microscopy [84]. It is a three-dimensional model that gives the tunneling current as a function of the tip geometry and its distance from the sample, as well as the electronic properties of tip and sample. Tersoff and Hamann fall back on a more general expression for the tunneling current, which gives the tunneling current in terms of the electronic states of the tip and sample. The s-wave model is explained below, after the derivation of the more general expression, the approach of Bardeen.

Bardeen Approach

Chen [44] describes the approach of Bardeen to tunneling theory, which is based on time-dependent perturbation theory. The main idea is to calculate the current (I_z) in the z-direction from the current density (j_z)

$$I_z = \int_A j_z \mathrm{d}A = \int_A n e v_z \mathrm{d}A\,,$$

where n is the number of electrons, with charge e, per unit volume that pass the surface S with a speed v_z. The electron density n is related to the density of states and the Fermi levels on each side of the tunnel barrier, while the speed v_z depends on the barrier width and electron transfer rate. Bardeen considers a system consisting of two metals in close vicinity as shown in figure 2.2. He separates the system into two sub-systems of one metal only, of which the solution to the Schrödinger equation is often known. The probability of electron transfer between the metals is determined by the overlap of the wavefunctions of the two sub-systems. Bardeen showed that the electron transfer is related to a tunneling matrix element ($M_{\mu\nu}$), which incorporates the wavefunction of both metals (Ψ_μ and Ψ_ν)

$$M_{\mu\nu} = -\frac{\hbar^2}{2m} \int_A (\Psi_\mu^* \nabla \Psi_\nu - \Psi_\nu \nabla \Psi_\mu^*) \mathrm{d}S\,, \tag{2.10}$$

where the integration is over any surface A between the two metals. In this section the index μ stands for the sample (metal 1), while index ν stands for the tip[1] (metal 2). The rate of electron transfer is calculated using Fermi's Golden Rule, which gives the probability (w) that an electron in state $|\Psi_\mu\rangle$ with energy E_μ will tunnel into a state $|\Psi_\nu\rangle$ with energy E_ν as

$$w = \frac{2\pi}{\hbar} |M|^2 \delta(E_\mu - E_\nu), \tag{2.11}$$

where M is given by equation (2.10). The tunneling current is finally given by

$$I_t = e \sum_{\mu,\nu} w f_1(E_\mu)[1 - f_2(E_\nu - eV_{\mathrm{bias}})]\,, \tag{2.12}$$

where $f(E)$ is the Fermi-Dirac distribution and the product

$$f_1(E_\mu)[1 - f_2(E_\nu - eV_{\mathrm{bias}})]$$

describes the probability of the state $|\Psi_\mu\rangle$ being occupied and at the same time the state $|\Psi_\nu\rangle$ being empty.

Standard Model

The standard model for scanning tunneling microscopy goes back to work published by Tersoff and Hamann [181]. They use Bardeen's result, given in equation (2.12), and use the approximations of small bias voltages, $eV_{\mathrm{bias}} \ll \Phi$, and

[1] Note that Tersoff and Hamann [181] chose the opposite notation. In their paper electrons tunnel from tip to sample, while in this description, as in many publications (e.g. [83, 193]), it is the other way round. The theoretical result is unaffected.

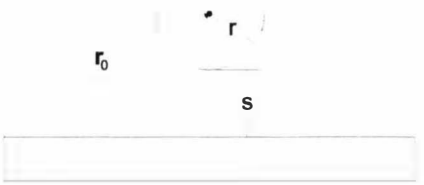


Figure 2.3: **Standard Model:** The tip has a radius of curvature r and is at a distance s from the sample. The center of tip curvature is at $\mathbf{r_0}$.

Figure 2.4: **Sinusoidal Sample Model:** Chen [44] as well as Binnig and Rohrer [27] assume this sample surface to estimate the lateral resolution of the s-wave tip model of Tersoff and Hamann [181].

low temperatures, $kT \ll \Phi$, where k is Boltzmann's constant and T the absolute temperature. As a result, the Fermi-Dirac distribution can be replaced with a step function and the local density of states can be introduced as above. With these approximations and substituting equation (2.11) into equation (2.12) they get

$$I_t = \frac{2\pi}{\hbar} e^2 V_{\text{bias}} \sum_{\mu,\nu} |M_{\mu\nu}|^2 \delta(E_\mu - E_{F1}) \delta(E_\nu - E_{F2}) \,, \tag{2.13}$$

Tersoff and Hamann assume the geometric conditions depicted in figure 2.3. The tip is assumed to end spherically with a radius of curvature r at a distance s from the sample. Tersoff and Hamann continue by expanding both wavefunctions. The crucial step in the derivation of Tersoff and Hamann is the assumption that for the expansion of the tip wavefunction it is sufficient to only take spherical waveforms, i.e. s waves, into account. As a result Tersoff and Hamann [181] find

$$I_t = 32\pi^3 \frac{e^2}{\hbar} V_{\text{bias}} \Phi^2 \rho_2(E_{F2}) r^2 \frac{e^{2\kappa r}}{\kappa^4} \sum_{\mu} |\Psi_\mu(\mathbf{r}_0)|^2 \delta(E_\mu - E_{F1}) \,, \tag{2.14}$$

where ρ_2 is the density of states of the tip and $\Psi_\mu(\mathbf{r}_0)$ is a state of the sample, but at the tip position ($\mathbf{r}_0$). Under constant current conditions the position of the tip ($\mathbf{r}_0$) is adjusted such that the tunneling current (I_t) does not change. In equation (2.14) the first factors do not depend on the tip position, so that the tip moves on a equal-current area given by

$$\sum_{\mu} |\Psi_\mu(\mathbf{r}_0)|^2 \delta(E_\mu - E_{F1}) = \rho_1(\mathbf{r}_0, E_{F1}) = \text{const.} \,.$$

Therefore equation (2.14) shows that 'a topographical image of the surface' is an area of constant local density of sample states (ρ_1) at the Fermi level taken at the position of the tip ($\mathbf{r}_0$) [44, 84]. With Tersoff and Hamann's initial expansion of the surface wave function ($|\Psi_\mu\rangle$) in the beginning of their derivation, the surface wave function evaluated at the tip position ($\Psi_\mu(\mathbf{r}_0)$) is found to be

$$|\Psi_\mu(\mathbf{r}_0)|^2 \propto e^{-2\kappa(r+s)} \,.$$

The tunneling current in equation (2.14) is therefore

$$I_t \propto e^{-2\kappa s}$$

as the one-dimensional model of equation (2.5) predicted.

Lateral Resolution The model of Tersoff and Hamann offers an estimate of

$$\sqrt{2\frac{r+s}{\kappa}} \propto \frac{\sqrt{r+s}}{\Phi^{\frac{1}{4}}} \tag{2.15}$$

for the lateral resolution of a scanning tunneling microscope. The smaller the tip radius and the closer the tip is to the sample, the better the resolution of the instrument. The resolution is also related to the effective barrier Φ. Binnig and Rohrer [27] point out that the resolution, after improving with smaller tip-sample distance (s), will rapidly worsen at very small s, because the effective barrier decreases under this condition.

Chen [44] discusses the standard model of Tersoff and Hamann as well as another model of Stoll. He considers a metal with a sinusoidal surface corrugation of a certain periodicity (a_s) and amplitude (c_s) as sketched in figure 2.4. Chen states that for a large periodicity, $a_s \gg \pi/\kappa$, and large tip-sample distance, $s \gg 2/\kappa$, both models predict that the measured corrugation (c_m) is

$$c_m = c_s \exp\left(-\frac{\pi^2(r+s)}{\kappa a_s^2}\right) . \tag{2.16}$$

An important conclusion from equation (2.16) is that the measured corrugation decreases exponentially with tip-sample distance. The fact that the sum $r + s$ stands in the exponent means that, as a result of these models, it is not *just* the tip-sample distance which is the important length.

Chen [44] points out that in this s wave model the electronic states of the tip do not play a role. He continues saying that while the model provides an adequate explanation of topographical STM images of large structures, it cannot explain the experimentally observed atomic resolution found on many metal surfaces, the earliest ones being reported by Hallmark *et al.* [83] and Wintterlin *et al.* [193]. Chen suggests that the electronic states of the tip as well as forces between tip and sample have to be taken into account to explain the high lateral resolution of the STM.

Chen cites Baratoff as being the first to explain the high resolution as coming from a localized electron orbital of the tip. Such an orbital, or dangling bond, is likely to be a d_z or p_z-like state located at the tip apex. Chen says that there is evidence that tips made from tungsten, a popular tip material, have such a d_z dangling bond protruding from the apex. Chen cites several more authors that stress the importance of taking the tip electronic states into account when discussing the lateral resolution of the STM.

Tip-Sample Interactions

Chen continues with a discussion of the tip-sample interaction. Already Wintterlin *et al.* [193] suggest that forces between the tip and sample are responsible for the observed resolution in their experiments. In the case of graphite Soler

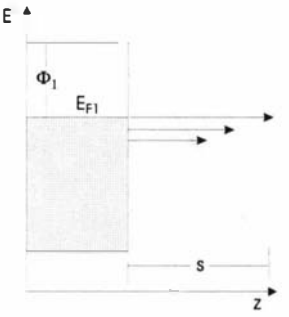


Figure 2.5: **Field Emission:** At high voltages the barrier becomes triangular and electrons can more readily tunnel through the barrier.

et al. [174] discuss the deformation of the relatively soft graphite by the hard tip, which leads to what they call giant measured corrugations.

When the tip is moved closer and closer towards the sample the tip will eventually come into contact, as observed by Gimzewski and Möller [75]. Depending on the tip surface, whether it is metallic or has an oxide layer, they find different forces acting on the tip when in close contact. With a clean tip and sample there are attractive forces, corresponding to metallic adhesion, that make the tip jump into contact with the sample at sufficiently small distances. This is also reported by Dürig *et al.* [63] and Olesen *et al.* [144]. With a contaminated tip, Gimzewski and Möller [75] find that, once in contact, there is also adhesive bonding between tip and sample, but that the bonding is very weak.

When discussing tip-sample forces, the image force also has to be mentioned, which lowers the potential barrier between tip and sample. Chen [44] summarizes the discussion of tip-sample forces as follows. There are attractive forces at large tip-sample distances, about 0.5 nm, while at very short distances, about 0.15 nm, the forces are repulsive. The equilibrium distance is about 0.20 to 0.25 nm and usually tunneling occurs at a distance of 0.3 to 0.7 nm.

Field Emission

At high enough voltages the tunneling current is not proportional to the bias voltage any longer. The potential barrier seen by the electrons becomes triangular in shape and it becomes easier for electrons in states with higher energy to tunnel through the barrier. A sketch is shown in figure 2.5. Since the barrier is thinner than in the metal-to-metal tunneling case, the electrons tunnel into vacuum or air. This phenomenon is known as field emission and begins at voltages of about 4 V [48]. Field emission was extensively studied by Fowler and Nordheim early this century. The relationship between current, voltage and tip-sample distance in the Fowler-Nordheim regime of tunneling is given by [84]

$$I \propto \left(\frac{V_{\text{bias}}}{s}\right)^2 \exp\left(-\text{const.}\,\frac{s}{V_{\text{bias}}}\right). \tag{2.17}$$

Once the electrons have tunneled through the barrier, they feel an electric field of a strength of about $E = V_{\text{bias}}/s$, which accelerates them towards the sample. This fact makes the STM useful as a tool for nano-manipulation, since the impacting stream of electrons on the sample can be used to modify the sample locally. The stream is easily controlled via the bias voltage.

In Fowler-Nordheim mode, the tip can be used at greater distances of up to 100 nm [84], which can be useful for the tip-sample approach. With the tip at

moderate distances of several nanometers, there is still a great vertical sensitivity due to the strong dependence on s in equation (2.17). Sáenz and García [166] suggest that these facts might be exploited to observe soft materials (perhaps biological), at great distances, with high vertical resolution.

Surface Spectroscopy

Examining the sample by recording the tunneling current as a function of the bias voltage is known as spectroscopy. An extensive review of spectroscopy using the scanning tunneling microscope is given by Hamers [85].

The actual relationship between tunneling current and bias voltage can be complicated, since the electronic tip and sample states are involved. For small bias voltages the window of energies between E_{F1} and $E_{F1} - eV_{bias}$ is small and only electrons with energies near the Fermi level contribute to the tunneling current. At higher voltages the energy window increases and more states contribute to the tunneling current. This contribution is weighted by the occupation probability of the states. According to the definition of the decay constant, $\kappa \propto \sqrt{U - E}$, the higher the energy E of the state, the smaller κ will be and the less rapidly the wave function decays [85]. Therefore the strongest contributors to the tunneling current will still be the states with energies closest to the Fermi level (E_{F1}).

In equation (2.5) the tunneling current is found to be proportional to the local density of states of the more negative metal, i.e.

$$I_t \propto \rho(E_{F1}).$$

With the scanning tunneling microscope either the tip or the sample can be more negative, simply by choosing the right polarity for the bias voltage. By varying the bias voltage the energy dependent local density of states of either tip or sample can be examined. While the polarity of the bias voltage determines the direction of tunneling, the zero volt level determines whether the states are usually occupied or unoccupied.

Binnig suggested a spectroscopic constant-current mode, where the height of the tip is varied, while the bias voltage is adjusted so as to keep the tunneling current constant. The result would be a characteristic z-V curve. This mode was named scanning tunneling spectroscopy (STS). The main reason for keeping the current constant was an instrument requirement, i.e. it was not possible to switch the analog feedback controller off.

As Hamers and coworkers [86] point out, it is desirable to record a complete I-V curve at constant tip-sample distance s and at an arbitrary point on the sample. The simultaneous recording of these curves and scanning the sample topographically is known as current imaging tunneling spectroscopy (CITS) [86]. The microscope of Hamers *et al.* made this mode possible by allowing the analog controller to be gated with a sample-and-hold circuit.

Usually it is only the sample that is to be investigated and not the combined tip-sample system. To exclude contributions from the tip, a clean metal tip is used for spectroscopy, where a constant density of states for the tip can be assumed. According to equation (2.5) the tunneling current is proportional to the bias voltage. Any deviation of proportionality in the recorded curve will be due to a change in the local density of states of the sample, such as a surface state or the bandgap of a semiconductor.

2.2 General Design Considerations

General design considerations for the microscope relate to the resolution, scan area and scan speed. The considerations should start with the equation of the tunneling effect, which is given in equation (2.6) and repeated here

$$I_t = \Lambda V_{bias} e^{-\frac{s}{s_0}},$$

where I_t is the tunneling current, V_{bias} the bias voltage and s the tip-sample distance. The symbols Λ and s_0 represent constants that determine the magnitude of the tunneling current and tip-sample separation. Both constants depend on the tip and the sample. The constant s_0 is approximately 0.05 nm as derived in equation (2.7). The constant Λ is not usually known. The tunneling current varies over several magnitudes, typical values are between 1 pA and 10 nA with bias voltages less than 1 V.

The short range of the tunneling effect and the necessity of controlling the microscope electronically confines the choice of possible electromechanical actuators. The chosen actuator needs to be reasonably fast and have very high resolution. A popular choice, which was also followed here, is the use of piezoelectric actuators. In this microscope the so-called 'beetle' design invented by Besocke [73] has been used. The beetle design is also used by other research groups [23, 68, 133, 191] and is commercially available from Zeiss [41]. A photo of the probe head used in this project is shown in figure 2.6, while a photo of the Zeiss probe head can be seen in figure 2.7. The choice of piezoelectric actuators and the specific probe head design influence all other design specifications, especially the bandwidth. Range and resolution also depend on the driving voltages, but the bandwidth of the system cannot be higher than that allowed by the mechanical properties of the piezoelectric actuator assembly.

An upper limit to the scanning speed is set by the lowest resonance of the piezoelectric actuator assembly carrying the scanning tip. The tip scanning rate should be reasonably high, not only for convenience, but also to avoid thermal drift influences. Another reason to have a high scanning speed is to minimize the possibility of sample surface modification such as adsorption when operating the microscope in a vacuum. Since the microscope is currently operated in air, such modifications are not yet a problem, because in air only relatively inert surfaces can be observed. Since electronic components usually react faster than mechanical ones, an important design specification was to not limit the scan speed of the microscope by the electronics.

Atomic features on sample surfaces have a typical size of 100 pm. If one wants to resolve atoms with the microscope, one would want a horizontal resolution of a tenth of that value, i.e. 10 pm. The horizontal resolution depends strongly on the vertical resolution because of the way in which surface images are recorded. If the surface features are as deep as they are wide, then an equal horizontal and vertical resolution is needed. Since the depth of surface features at an atomic level is often several times smaller than the width, a vertical resolution that is several times better than the horizontal one is needed. A vertical resolution of 1 pm could be chosen [156].

The limit of the achievable resolution is determined by the amount of noise in the system. The lower limit is the intrinsic measurement noise, which in the case of tunneling current is shot and tunneling junction noise. An estimate of

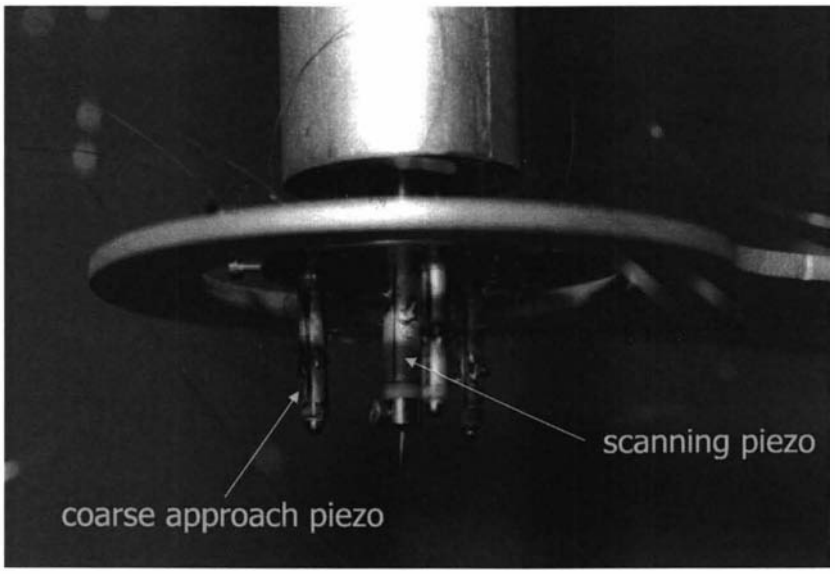


Figure 2.6: **Probe Head:** A side view of the probe head is shown together with the probe head resting ring and the tube-shaped shield for the tunneling current. The four piezoelectric actuator tubes can be seen, mounted on the probe head disk. The tip-carrying actuator is in the center and is larger than the outer actuators. In the lower middle of the photo the tip, the tip holder and the macor insulator can be seen. The three outer actuators each carry a steel ball at the bottom. There are sixteen wires, four per actuator, which connect the high-voltage positioning signals from the probe electronics to the actuator electrodes.

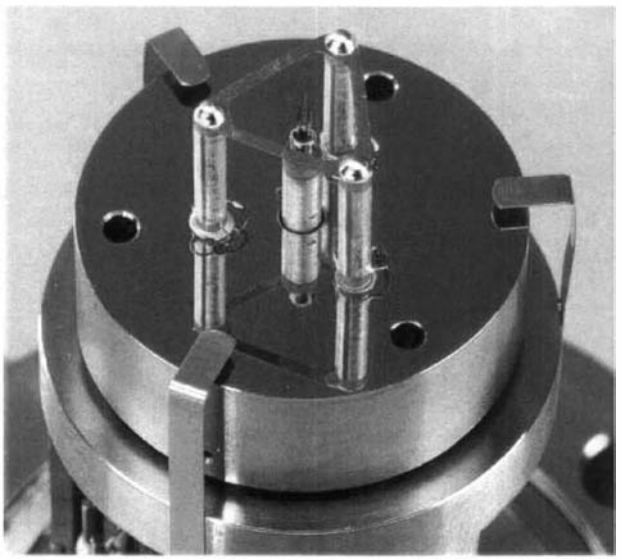


Figure 2.7: **Zeiss Probe Head:** The photo shows an upside down photo of a probe head of the Besocke design, commercially available from Zeiss. (Photo from OMICRON Vakuumphysik GmbH). Note the connections between the outer actuators, which give the Zeiss probe head extra rigidity.

the intrinsic noise level provides an upper limit for the vertical resolution of the microscope. The actual figures are discussed in section 5.1.2.

The electronic components that provide and condition the control and measurement signals also contribute to the noise. Keeping the amount of noise added by the electronic components as low as possible is the task of the electronic design (discussed in section 5.1.2). Finally there are perturbations to the tip and the sample that originate from external sources. The most important ones are vibrations, acoustic noise and thermal drift. Minimization of the external perturbations has to be addressed in the mechanical design of the microscope, which is discussed in section 3.3.

Once an acceptable noise level has been determined and the horizontal sensitivity of the piezoelectric actuator chosen, one can select the actuator control voltages. Here a compromise must be made between scan area and noise level.

One important consideration when designing the electronic instrumentation for a scanning probe microscope is the standard set by other custom-made and commercial microscopes. Introductory texts that overview microscope design are the books by Hamann and Hietschold [84] and Chen [44]. A review of microscope instrumentation was given by Kuk and Silverman [121].

An important design goal was to try to at least match the performance of these instruments. Pohl [156] discusses all major design considerations in detail. Typical specifications are 10 pm for the horizontal resolution, 1 pm for the vertical resolution, a scan area of a few micrometers squared and a bandwidth of about 10 kHz.

2.3 Essentials for Scanning Probe Microscopy

A scanning probe microscope has several essential features. These include a sample, a tip, a mechanism to move the tip horizontally and vertically and a facility to generate and record the interaction signal.

To minimize environmental influences, vibrational damping and thermal compensation are also needed. For practical reasons a tip holder is needed as well as a sample holder. Finally an approach mechanism is needed to move the tip from a distance of about 1 mm to within 1 nm of the sample.

2.3.1 Probe Head Overview

Bias Voltage

The first electronic stage is a current-to-voltage converter to measure the tunneling current. The current-to-voltage converter and the bias voltage are connected to the tip and the sample, as shown in figure 2.8. Either tip or sample is set to a certain electrical potential by the bias voltage, while the other electrode is set to virtual ground by the current-to-voltage converter. It is at the discretion of the operator, whether the tip or the sample is at ground potential. However, for the current-to-voltage converter it is favorable to have the tip at ground potential, since the sample holder has a bigger capacitance to ground than the tip. Depending on the sign of the bias voltage, the electrons tunnel from the tip to the sample or vice versa. In practice most researchers prefer negative sample bias, thus making the tip positive with respect to the sample. One reason for

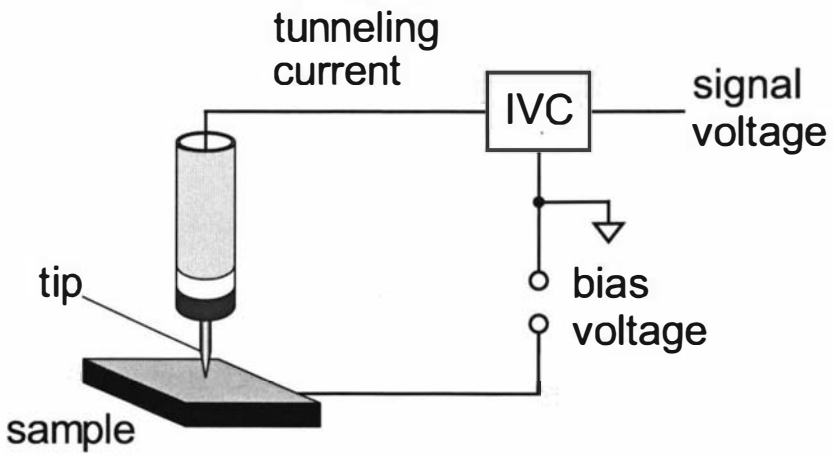


Figure 2.8: **Tip and Sample:** The tip is mounted at the bottom of the center actuator and held at a certain distance from the sample below. Between tip and sample a bias voltage can be applied. A current-to-voltage converter (IVC) holds the tip at virtual ground and converts the tunneling current to a signal voltage. Note that it is also possible to connect the IVC to the sample instead of the tip.

manual	probe head	center actuator
5 to 0.4 mm	500 to 0.01 μm	1200 to 0.01 nm

Table 2.1: **Tip-Sample Approach:** The different control distances over which the tip is approached towards the sample are listed. Bringing the tip to within 0.4 mm of the sample is done manually. After that the probe head and center actuator approach the tip.

this preference is that the metal ions of the tip are then attracted to the sample. If the ions can move, for instance due to a high electric field, then they will wander to the tip apex thereby sharpening the tip.

Tip Positioning

A mechanism to move the tip in all directions of space is another essential prerequisite of scanning probe microscopy. In addition, an approach mechanism to move the tip towards the sample is also needed. The approach of the tip has to work over several scales; millimeter, micrometer and nanometer. Table 2.1 shows a summary. The tip-sample approach is discussed in sections 6.4 and 7.3. Moving the tip is done with the help of piezoelectric actuators. A diagram of an actuator used in this microscope is shown in figure 2.9. The actuator is tube-shaped and is made of a piezoelectric ceramic, which contracts or elongates when a voltage is applied across it. There are four tube actuators in the probe head, each one having four electrodes. Putting a voltage on all electrodes makes the tube contract or elongate. This produces vertical movement. Putting a voltage on one pair of electrodes ($\pm x$ or $\pm y$) makes the tube bend in one direction, because one side shrinks and the opposite side grows. This produces horizontal movement.

A diagram of the probe head is shown in figure 2.10. This probe head design is of the so-called 'beetle' type [73]. A photo of the probe head showing

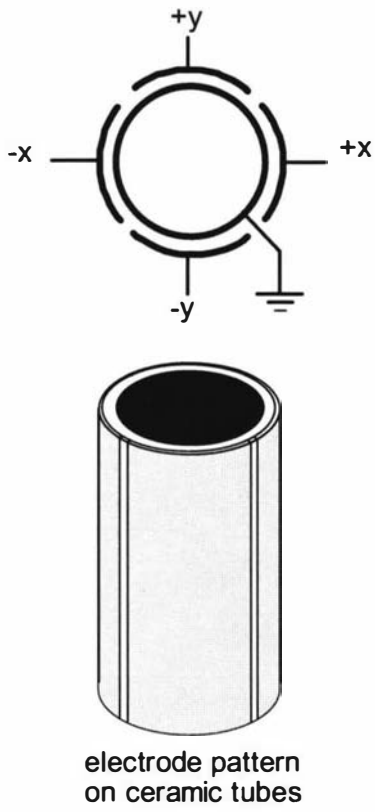


Figure 2.9: **Piezoelectric Actuator Tube:** The inside and outside electrodes of the actuator are shown along with the connection pattern. The inside electrode is grounded, while the outside electrodes are driven in pairs, labeled $\pm x$ and $\pm y$.

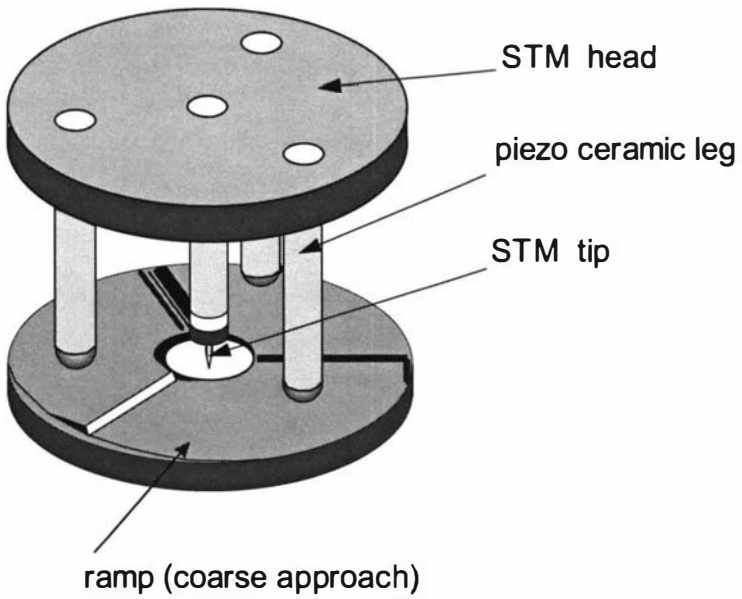


Figure 2.10: **Beetle-type Probe Head:** The probe head followed the design of Besocke [73]. The head consists of a disk and four piezoelectric actuators. The disk is supported by the three outer actuators, while the center actuator holds the tip, which is the probe for the microscope. The probe head sits on three ramps. The ramps are arranged such that the lowest point of each ramp neighbors the highest point of another ramp. A clock-wise circular movement of the probe head causes the complete assembly to move towards the sample. A photo of the probe head is shown in figure 2.6.

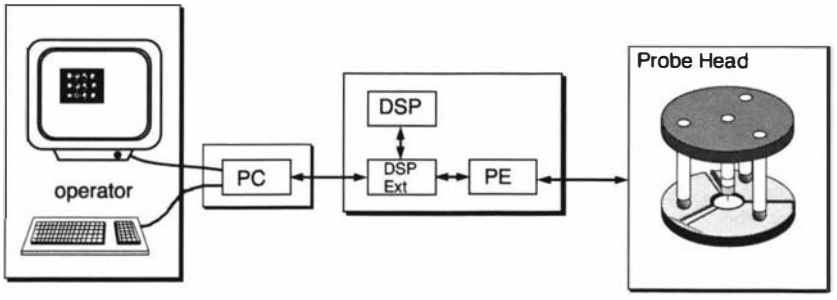


Figure 2.11: **Overview of the Scanning Tunneling Microscope:** This diagram shows the whole microscopy system. To the right is the mechanical part of the microscope, the probe head. In the center is the electronics, consisting of the probe electronics (PE), the digital signal processor (DSP) and the interfaces for the DSP (DSP Ext). To the left is the PC. The way the microscope is operated is symbolized by a keyboard and a screen. The electronics between the PC and the microscope controls the microscope and also uploads image data to the PC. The PC provides a user interface for the operator of the microscope. Most of the electronics was designed and built in-house.

the center and outer actuators is presented in figure 2.6. The center actuator carries the tip and scans it across the surface. There are three more actuators, which support the probe head. These outer actuators approach the probe head towards the sample. The three legs stand on a ramp, which is mounted on top of the sample holder. Figure 2.12 shows a photo of the sample holder ring. The top of each ramp is 0.4 mm higher than the bottom of the ramp. The outer actuators can rotate the probe head around the center of the ramp. The tip is thereby moved very slowly towards the sample through a maximum height of 0.4 mm. The legs 'walk' down the ramps in many small steps. In engineering this walking is generally known as slip-stick movement, which is based on the difference between static and dynamic friction. Each step consists of a slip and a stick phase. In the slip phase the legs bend rapidly sideways so as to overcome static friction. During this process the legs glide over the surface of the ramp, while the probe head remains fixed owing to its greater inertia. During the stick phase the legs relax slowly back to their original length. In this phase static friction allows the legs to grip the ramp and the probe head moves with the relaxing legs.

There are practical reasons in favor of the beetle design, connected to vibrational damping and thermal compensation. This is discussed in sections 3.3 and 3.4.

2.3.2 Microscope Overview

The scanning probe microscopy system consists of three major units, shown in figure 2.11. These parts are the probe head to the right, the digital signal processor (DSP) and assorted electronics in the center and the PC on the left. In figure 2.11 the mechanical part of the microscopy system is symbolically represented by the probe head standing on the sample holder ramps. Apart from the approach mechanism, the mechanical part of the microscope also provides acoustic and vibrational damping for the probe head as well as mechanical support. A photo of the scanning probe microscopy system is shown in figure 2.13. Instead of the transparent cover shown in the photo, the mechanical parts of

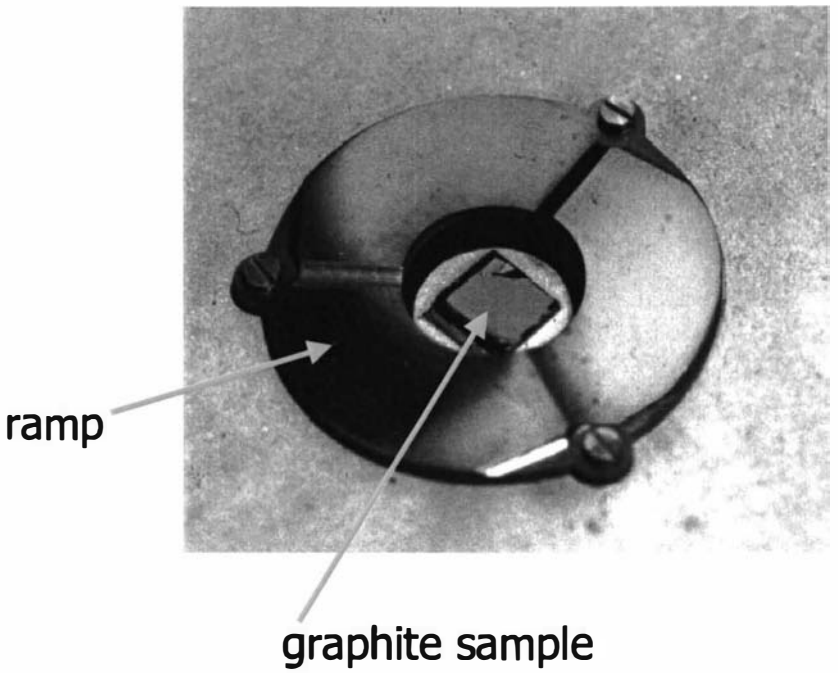


Figure 2.12: **Coarse Approach Ramp:** This photo shows the sample and the three ramps the coarse approach actuators stand on. The surface of the ramp has to be as flat as possible to allow a smooth approach. The ramp is made from tungsten carbide to prevent scratching. The sample is highly orientated pyrolytic graphite. The outer diameter of the sample holder ring ramp is 25 mm, the inner diameter is 10 mm. The ramp height is 0.4 mm.

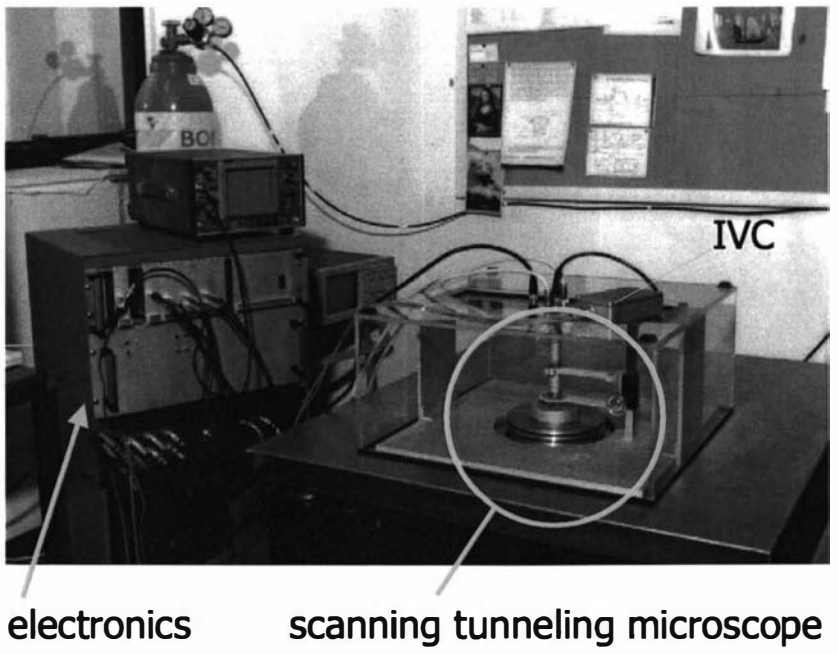


Figure 2.13: **SPM System:** The probe head is indicated by a circle on the right, sitting on a stack of brass plates. Between the plates are viton O-rings. The probe head and the stack sit on top of a pneumatically damped table. This air table and the stack of small plates together form the vibration damping system for the microscope. The transparent cover has meanwhile been replaced by a cover with improved acoustical damping. The electronics are mounted in the rack to the left.

the microscope are now accommodated in a housing furnished with acoustical damping. The scanning actuator as well as the actuators for the finer approach mechanism are connected to the electronics unit. The electronics also provides the bias voltage between tip and sample and acquires the tunneling current. A part of the electronic hardware, the DSP, is commercial equipment. The remainder has been custom-made.

The microscope is operated from the PC, which interfaces via custom-made electronics to the DSP. The software in the PC receives commands from the operator and then these commands are passed on to the DSP. In exchange, the data gleaned from the underlying sample are uploaded from the DSP to the PC and presented to the operator on the PC screen. The DSP software continuously controls the microscope with the help of the electronics. A diagram of the interaction between PC, DSP and probe head is shown in figure 1.3 in section 1.3, where the role of the DSP is discussed in detail.

2.3.3 Operation of the Microscope

The first step in operating the microscope is to approach the tip to the sample. After the tip has been brought close enough to the sample, a suitable scan area and resolution can be chosen and then a topographical scan can be made.

Tip-Sample Approach

The tip approaches the sample surface by three different mechanisms, each working at ever decreasing length scales. After approaching as close as possible with one mechanism, the one with a smaller range takes over and brings the tip even closer to the sample surface. One can distinguish between manual, coarse and fine approach mechanisms. The tip is attached to the probe head and is brought close to the sample surface by hand. At this stage the tip has to be within 0.4 mm of the sample surface. By contrast the coarse and fine mechanisms of approach are electronically controlled. During the coarse approach the whole probe head is moved towards the sample in a series of small steps. Each of the small steps moves the microscope vertically a maximum distance of 120 nm. Although each step size is comparatively small, the number of steps is unlimited. A finer movement is possible by elongation of the actuator the tip is mounted on, while the microscope body itself remains at rest. The resolution depends mainly on the electronic noise and is a fraction of a nanometer. The elongation can carry the tip over a maximum distance of ± 650 nm.

While the tip is approaching the sample there is always a possibility of destroying the tip through accidental impact. The approach is therefore under electronic feedback control. During the fine approach phase the tunneling current is monitored. As soon as a tunneling current value reaches a certain threshold, electronic control is initiated to regulate the distance between tip and sample. If no tunneling current is detected within the range of the fine approach mechanism, the tip will be removed from the domain of the fine approach and the coarse approach brings the microscope head closer to the sample. The process of alternate searching and approaching is then repeated until tunneling current is detected. Note that it is necessary to continuously control the tip-sample distance electronically because of the presence of thermal drift and

vibrations. The approach algorithm is discussed in detail in section 6.4, while the step size used during the approach is discussed in section 7.3.

Scan

After the tip has come within tunneling distance of the sample, the surface can be examined by performing a topographical scan in either constant-height or constant-current mode. The different modes have been previously described in section 1.1.3. The first step in a topographical scan is to adjust the bias voltage to the desired value, thereafter keeping it constant. The scan itself consists of moving the tip to a series of new x-y positions and reading the tunneling current. In the constant-current mode this is followed by an adjustment of the tip-sample distance.

The sample can also be examined by recording characteristic curves, which have been described in section 1.3.2. Prior to recording, the tip is first moved to the desired x-y position. Following this, the tunneling current is measured as a function of bias voltage or tip height.

Region of Interest

To find another region of interest the tip has to be retracted away from the sample surface. The microscope head is then moved horizontally towards the new area of interest. Afterwards the tip is brought into tunneling distance again. Some means are necessary to move the microscope head horizontally. The same mechanism that provides the coarse approach towards the sample can also move the head horizontally. This requires a redistribution of the applied positioning signals to the outer piezoelectric actuator legs. The amount and distribution of these positioning signals is discussed in the following paragraph.

High-Voltage Positioning Signals

A total of seven high-voltage positioning signals are connected to the piezoelectric actuators of the microscope. Since electronic control should be operational at all times, it must be possible to move the tip in all three dimensions. This means that a minimum of three independent positioning signals are needed. However the situation is complicated by the need to provide three basic modes of operation; coarse and fine approach and horizontal probe head movement.

For convenience the fine approach mechanism is run from four high voltage signals, one signal for each of the four electrodes of the center actuator.

For the coarse approach one requires circular movement which in principle requires only one independent signal, however the system is faster and more flexible when three independent signals are used.

Since vertical and horizontal motion of the beetle never occur at the same time, the two modes share some of the high voltage lines; the horizontal mode using two signals and the circular configuration three signals. The signals pass a dedicated switching panel, which groups the outer electrode connections into the two different configurations. This is discussed in more detail in section 5.4.8.

It has to be noted that it is possible to operate the microscope with four independent positioning signals if one is willing to switch the positioning signals between the electrodes of the outer actuators and the tip-carrying actuator.

Both scanning and moving the probe head can be done with four positioning signals. The crucial operation is the tip-sample approach, which could be done with three positioning signals to move the probe head and a single control signal for vertical tip movement, since horizontal tip movement is not needed during tip approach. However, keeping the positioning signals for the scanning and the outer actuators separate makes operation more convenient and improves system versatility, which is a major objective of this project.

Data Processing

The operator can view the image data by transferring it from the DSP to the PC, which then displays the data on a screen. The image data can be processed by the PC, while the DSP acquires more data, although it is also possible to preprocess the image data with the DSP before transferring the recorded image to the personal computer. Images can be stored on standard digital storage media.

The operator can adjust the scan parameters of the microscope on the PC. The new parameter values are read by the DSP and the old values are altered accordingly.

Chapter 3

Mechanics

3.1 Moving the Microscope and the Scanning Tip

Fine mechanical movement of the microscope probe head is generated with the help of piezoelectric actuators. Since the movement is over very small distances (nm) and needs to be under electronic control, an electromechanical actuator with very high resolution is required. The most widely used actuators are those using the converse piezoelectric effect. Although other actuators exist, such as those of the electro- or magnetostrictive type [103], these are rarely used in SPM[1], mainly because both effects are small in comparison to the piezoelectric effect [185].

3.1.1 Piezoelectric Effect

Definition

Some materials develop surface charges when a force is applied. As a consequence of these charges a voltage is seen across the material or, if the material is short-circuited a current can be observed [79]. This is known as the direct piezoelectric effect and was discovered on quartz by Pierre and Jacques Curie in 1880. If a voltage is applied to such a piezoelectric material, then the material will distort mechanically i.e. it will be strained. This is the converse piezoelectric effect [79]. Piezoelectric materials can therefore be used as sensors (input transducers) or as actuators (output transducers). Sensors exploit the direct piezoelectric effect, while actuators use the converse piezoelectric effect. Both effects intertwine electrical and mechanical properties and are jointly referred to as the piezoelectric effect.

The piezoelectric effect is found in materials that have an asymmetric crystal structure [54, 79]. It is impossible for a crystal with central symmetry, such as NaCl, to be piezoelectric. However, crystals belonging to classes that are not centrosymmetric are piezoelectric [54], for example quartz or zincblende (ZnS). Among piezoelectric crystals a sub-group can be distinguished. These

[1] An interesting account of using magnetic levitation in AFMs was reported by Gauthier-Manuel and Garnier [74]

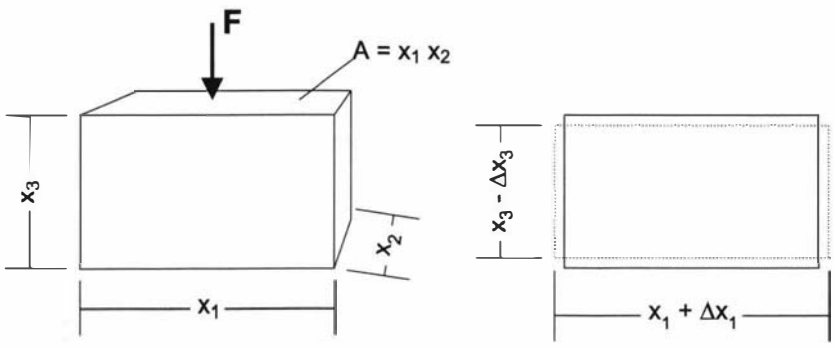


Figure 3.1: **Block of Piezoelectric Material:** A rectangular block with dimensions x_1, x_2 and x_3 is squeezed by a force **F** in the x_3 direction. The stress on the material is $X = F/A$, where $A = x_1 x_2$ is the area facing the force. The length change is shown to the right. The block experiences a compressive strain $\Delta x_3/x_3$ and a tensile strain $\Delta x_1/x_1$ and $\Delta x_2/x_2$.

are crystals with permanent surface charges[2] and are termed pyroelectric[3]. If a crystal is pyroelectric only below a certain temperature, its Curie temperature, then it is termed ferroelectric. This is a sub-group of high relevance, since almost all commercial piezoelectric actuators are made from ceramics of ferroelectric crystallites.

Stress Polarization

Let a rectangular block have the dimensions x_1, x_2 and x_3 as depicted in figure 3.1. If a force F is applied in the x_3 direction then the block experiences a stress (X) of

$$X = \frac{F}{A} = \frac{F}{x_1 x_2},$$

which causes a strain ($\Delta x_3/x_3$). To first order, the developed strain is proportional to the stress and both are related by Hooke's law

$$X = Y \frac{\Delta x_3}{x_3}, \tag{3.1}$$

where Y is Young's modulus or the elastic stiffness constant.

If the block is made of piezoelectric material, then extra charges (Q) will build up on both of the end surfaces. These charges give rise to a polarization (P) of the material[4]. The intrinsic asymmetry of piezoelectric materials has the consequence that the polarization (P) and the strain ($\Delta x_3/x_3$) are related,

[2]The charges can not normally be detected since they are neutralized by a layer of attracted ions from the atmosphere. However, if a change in charges is generated, by heating the crystal or applying a force, then surface charges can build up faster than they are neutralized.

[3]Ashcroft and Mermin [13] define piezoelectric crystals as the non-centrosymmetric crystals that are not pyroelectric. The other cited references [54, 136, 185] however define the pyroelectric crystals as a sub-group of the piezoelectric crystals.

[4]The charges on one end of the block will be equal and opposite to the charges on the other end, since the block must remain electrically neutral. The surface charges are $Q = \sigma A$, where σ is the surface charge density. The stressed block has a electric dipole moment Qx_3 given by the surface charges Q separated by the block height x_3. The polarization (P) of the block is given by the dipole moment of the block divided by its volume $V = Ax_3$. The magnitude of the polarization is equal to the charge density [61, 136] $P = (Qx_3)/(Ax_3) = \sigma$.

which is the crucial property of these materials. To a first order approximation the polarization is proportional to the strain, i.e.

$$P \propto \frac{\Delta x_3}{x_3}. \tag{3.2}$$

Combining the proportionality 3.2 with Hook's law 3.1, the piezoelectric effect can now be stated more precisely as a polarization (P) of certain materials caused by an applied stress (X) [54, 136, 185]

$$P = Xd,$$

with d being the piezoelectric strain constant. The polarization will cause an electric field within the material and the piezoelectric block acts as a generator.

Converse Piezoelectric Effect In scanning probe microscopy piezoelectric actuators are used as positioning devices. The actuators consist of a piece of piezoelectric material equipped with electrodes across which is applied a voltage. This voltage gives rise to an electric field

$$E_0 = \frac{V}{x_3},$$

where x_3 is the thickness of the material. The field E_0 is the field due to the free charges on the electrode of the actuator and causes a polarization (P) of the material. The resulting field (E) within the ceramic is

$$E = E_0 - \frac{P}{\epsilon_0}, \tag{3.3}$$

which is smaller than E_0. Often the polarization is proportional to the exciting field E_0 and equation (3.3) can be expressed as

$$E_0 = \kappa E, \tag{3.4}$$

where the dielectric constant κ has been introduced as the proportionality constant. From equation (3.3) and equation (3.4) the relation between the polarization and the electric field within the material can be found

$$P = \epsilon_0(\kappa - 1)E. \tag{3.5}$$

According to equation (3.2) the polarization of the material causes a strain, which in turn either elongates or compacts the material. The converse piezoelectric effect can be written as

$$\frac{\Delta x_3}{x_3} = Ed, \tag{3.6}$$

where the proportionality constant d is the same piezoelectric strain constant as defined above. The dimension of d is $^{C}/_{N}$ or $^{m}/_{V}$.

Piezoelectric Equations The direct and converse piezoelectric effect are summarized together with Hooke's law and equation (3.5) as the piezoelectric equations

$$P = Xd + \epsilon_0(\kappa_X - 1)E, \tag{3.7}$$

$$\frac{\Delta x_3}{x_3} = s_E X + Ed \tag{3.8}$$

where s_E is the compliance of the ceramic, which is equal to the reciprocal of Young's modulus. The subscript E indicates explicitly that this is the compliance at constant electric field. In the same way the constant κ_X is the dielectric constant at constant stress.

Direction In the example above, when stress is applied along the z-axis, the block will shorten in the same direction, but will elongate in the x and y direction. Generally speaking, stress in one direction may cause strain in other directions as well. This is explicitly described in tensor notation, where strain is a second rank tensor with nine components. The relation between polarization and applied stress is in tensor notation

$$P_i = d_{ijk} X_{jk},$$

where i is the direction in which polarization is generated, j is the direction of the strain and k is the direction in which the stress is applied. The variable d_{ijk} is the piezoelectric modulus, which is a tensor of rank three [84]. Usually the assumption is made that the strain tensor (X_{jk}) is symmetric, and the tensor components with two indices are rewritten as matrix elements with only one index [54, 185]. This index is conventionally numbered from 1 to 6. The first three values indicate the three normal directions of stress (x, y and z), while the values 4 to 6 indicate shear stress. Thus the three indices of the piezoelectric modulus are reduced to two indices, i.e.

$$P_i = d_{ij} X_j,$$

where i indicates the same direction as above and j is the direction of the applied stress. For the converse effect an equivalent expression holds, which is given as

$$\frac{\Delta x_i}{x_i} = d_{ij} E_j,$$

where i is the direction in which strain is generated and j is the direction in which the electric field is applied.

Piezoelectric actuators are most commonly used with the strain in the same direction as the field (longitudinal effect), or perpendicular to the field direction (transverse effect). The polarization direction of a piezoelectric material is conventionally denoted 3, so that the two most commonly needed piezoelectric strain constants are d_{33} and d_{31}. The most popular actuators are tube shaped actuators discussed in section 3.1.2. They use the transverse effect, which is described by d_{31}.

Piezoelectric Ceramics

Actuator Production Actuators made from piezoelectric ceramic can be shaped into any form. The actuators are then supplied with electrodes and finally poled. Without poling the ceramic would remain macroscopically unpolarized. Since the piezoelectric ceramic materials are usually ferroelectric, they consists of domains, which orientate along the field direction during poling. After the poling field is removed a remanent polarization remains in the ceramic. This process is analogous to the magnetization of a ferromagnet. Since the ceramic is polycrystalline the material not only consists of domains but also grains, which may be equal to one or more domains. Reorientation of the domains under the influence of the poling field occurs only within each grain.

The most widely used piezoelectric ceramic is a mixture of lead titanate ($PbTiO_3$) and lead zirconate ($PbZrO_3$), both of which have a perovskite type of crystal structure [185]. The ceramic exhibits very high piezoelectric strain constants for the right mixture, which is the reason why this type of ceramic material has become dominant. The ceramic is abbreviated as industry type PZT. For the center actuator the ceramic EBL2 from Staveley Sensors [175] was used, which is equivalent to the industry type PZT-5A. The piezoelectric strain constants d_{31} for this material is $-0.173\,\mathrm{nm/v}$ at 293 K.

Producing actuators from ferroelectric ceramic has several consequences, which are:

- The actuators have a maximum operating temperature.
- The actuators age.
- The actuators have a maximum operating field strength.
- The actuators exhibit hysteresis.
- The actuators exhibit creep.

Maximum Operating Temperature and Aging A ferroelectric material has two different phases depending on temperature. Only below its Curie temperature is the material piezoelectric. At higher temperatures the material turns paraelectric. The Curie temperature for PZT-5A is 350°C. Already at temperatures below the Curie temperature the material starts to degrade. This is because it is energetically more favorable for the domains to be randomly orientated, so as to show macroscopically no polarization. Uchino [185] defines aging as the slow strain degradation caused by depoling. At normal temperatures piezoelectric ceramics remain poled for many years. However, while soldering to the piezoelectric electrodes the surface temperatures of the ceramic will inevitably be elevated. The electrodes should be soldered at the lowest possible temperatures for a brief time only. To avoid heating altogether, electrical connections can also be made using conductive glue.

Maximum Field Strength and Hysteresis At small electric field strengths the converse piezoelectric effect is approximately linear. At higher field strengths the ferroelectric domains begin to orientate themselves along the field. At small field strengths the orientation is reversible, at higher field strengths the aligned

domains grow at the cost of non-aligned ones and the orientation becomes irreversible. Hence, the greater the field strength the larger the strain hysteresis of the ceramic. The field strength to depole a piece of previously poled ceramic is the coercive field. It is not desirable to depole the ceramic, so the applied field must not be larger than a maximum field strength given in the data sheet. A typical dc depoling field is 10 V per 25 μm thickness of the ceramic [44, 173], which is equivalent to an electric field strength of 400 kV/m. The manufacturer of the center actuator only specifies the ac depoling field, which is the ac field strength needed to completely randomize the ceramic. The stated rms value is 700 kV/m. The actuator used as a tip-carrying actuator in this microscope has a wall thickness of 0.51 mm and the highest possible voltages produced by the control electronics are ±190 V. The highest possible field across the ceramic is therefore 370 kV/m, which is smaller than the rms ac depoling field, but close to the previously mentioned dc depoling field.

Fields opposing the original polarization contract the ceramic. Since such electric fields tend to depole the ceramic, it is recommended that the ceramic not be contracted over a long period of time. On the other hand the ceramic may be expanded indefinitely, since this will pole the ceramic.

Creep The phenomenon known as creep is the long term response of a piezoelectric actuator to a step in applied field. The response of the actuator consists of a more or less immediate change in strain followed by an additional slower change in strain, which can last for several minutes, depending on the material. An example of creep measured with this microscope is shown in figure 3.2. Creep is due to the interaction of the ceramic as a piezoelectric material and as a capacitor. Kaizuka [109] suggests that creep is mainly caused by the change of actuator capacitance due to the very strain that is generated by the applied voltage. Marti [131] suggests the same cause, but additionally stresses that the domain walls of the ferroelectric ceramic rearrange themselves in the applied electric field, which changes the capacitance. This is due to a change in the polarization (P) and can be described as a change in the dielectric constant (κ) of the ceramic:

$$C(t) = \kappa(t) C_0, \tag{3.9}$$

where C_0 is the capacitance without the piezoelectric ceramic. Since the applied voltage is constant after the step, a change in capacitance causes a change in charge via $Q = CV$. Note that the charge Q is the free charge on the electrode as opposed to the bound surface charge of the polarization. For a constant voltage the electric field E_0 that causes the polarization of the ceramic depends only on the actuator thickness t with $E_0 = V/t$. The field in turn generates a change in strain according to the converse piezoelectric effect described in equation (3.6). The process can be summarized as

$$\frac{\Delta l + \mathrm{d}l}{l} \propto Q = Q_0 + Q(t),$$

where $\Delta l/l$ is the immediate strain, which is proportional to the initial charge Q_0, and $\mathrm{d}l/l$ is the creeping strain generated by a slow domain reorientation, which is here expressed as a time-dependent charge $Q(t)$. When the long-term response of an actuator is measured experimentally, a logarithmic behavior is

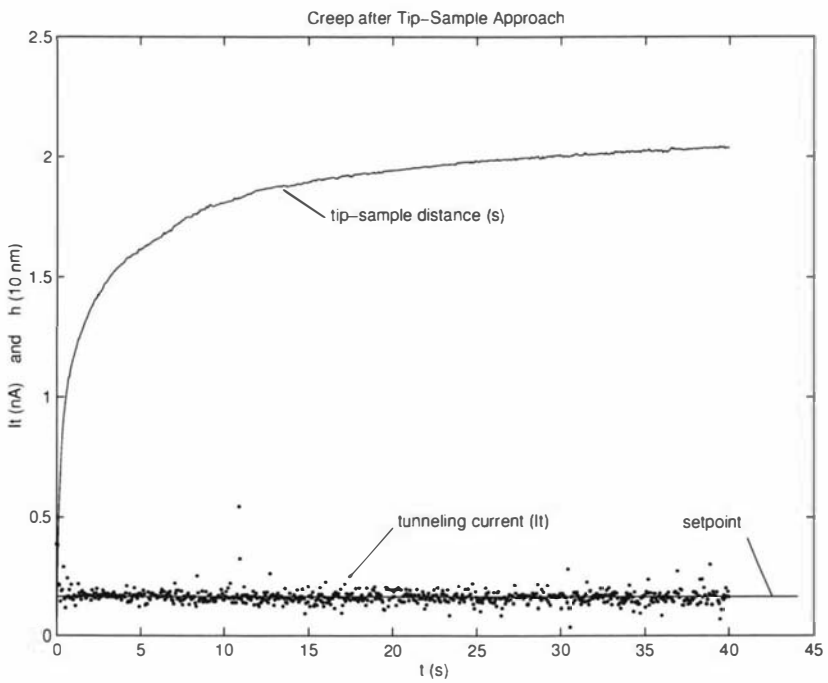


Figure 3.2: **Creep of Center Actuator:** The tip-sample distance (s) and the tunneling current (I_t) are shown simultaneously just after the approach mechanism has encountered tunneling current and switched on feedback control. The tunneling current is kept at the desired setpoint, while the controller adjusts the voltage of the center actuator so as to keep a constant tip-sample distance. Creep follows approximately a logarithmic behavior. After 40 s the center actuator expanded by approximately 20 nm.

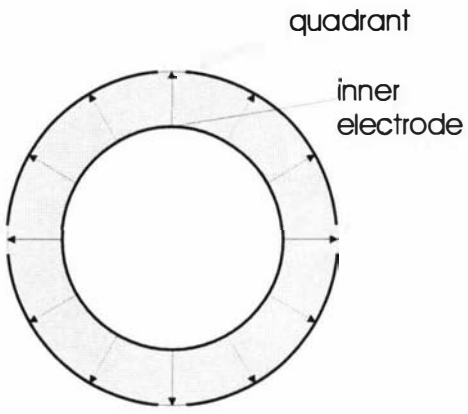


Figure 3.3: **Polarization of the Ceramic:** The piezoelectric ceramic of the actuators has been polarized with a positive voltage on the outside electrodes. After poling a permanent polarization is left, shown in the diagram. With the inner electrode grounded, the ceramic wall will contract if a negative voltage is applied to any of the quadrants.

observed [84, 156, 190]. Vieira [190] reports that the length change follows the equation

$$\frac{\Delta l}{l} = \alpha + \beta \ln t$$

very accurately, where α and β are parameters that depend on the electric field and the temperature. The logarithmic behavior can qualitatively be seen in figure 3.2. According to Dirscherl [59] all three quantities: field, strain and temperature, interact with each other. Vieira continues by summarizing the experimental findings of other researchers who found logarithmic time dependence. He suggests that creep is associated with thermally activated processes having different activation energies.

3.1.2 Piezoelectric Actuators

The actuators used in the probe head are tube shaped following the design first published by Binnig and Smith [29]. A diagram of a typical tube-shaped actuator is shown in figure 2.9. The microscope tip is attached to the lower end of the tube. Applying voltages to the tube quadrants generates strains in the tube and make it possible to move the attached tip in all six directions of space.

Instead of pressing or extruding the actuators, the manufacturer [175] machined them to achieve high precision. The tubes are electroplated on the inside and outside with a thin film of silver. The outer coating is segmented into four quadrants, which serve as four electrodes. The coating on the inside is used as a counter-electrode. The counter-electrode is kept at ground potential, while each of the quadrants is independently driven by a high-voltage control signal. The tube design uses the transverse piezoelectric effect and so the important piezoelectric strain constant is d_{31}. The piezoelectric ceramic has been poled with a positive voltage on the outside, so that the resulting permanent polarization of the ceramic is pointing from the inside of the tube radially outwards as shown in figure 3.3. An applied *negative* voltage on any of the quadrants results in an electric field opposed to the poling field and in line with the permanent

polarization. The electric field causes the tube wall to contract and the tube length to elongated. The opposite is true for positive voltages.

Since opposite fields tend to depole the ceramic, it is generally recommended that the tube be expanded rather than contracted[5], which means the actuator should preferably be used with the tip retracted from the sample rather than extended towards the sample. This is indeed possible if the actuator is used with a certain offset.

When a common voltage is applied to all four electrodes, the actuator tube elongates or contracts, which lowers or lifts the tip. This way the tip is moved vertically, which is denoted as the z direction.

When a voltage is applied between the counter-electrode and only one of the four outer electrodes, then the tube will *bend* towards or away from this electrode. In the current design, equal and opposite voltages are applied to opposing electrodes to improve linearity. In addition, with just one electrode activated the tip would move horizontally as well as vertically. The vertical movement is cancelled, if two opposing quadrants are driven simultaneously with equal and opposite voltages. In this manner the tip is moved horizontally in either the $\pm x$ or $\pm y$ direction.

When calculating the horizontal and vertical deflections of the two different actuators used in this microscope the actual lengths of the electrodes of the actuator and the attached pieces have to be taken into account. A detailed diagram of the center actuator is shown in figure 3.4, where it can be seen that the electrode length is only part of the overall actuator length. When a voltage is applied to all outer electrodes of the center actuator only the ceramic covered by the electrodes experiences an electric field. The resultant strain

$$\frac{\Delta l_{\mathrm{el}}}{l_{\mathrm{el}}}$$

relates only to the electrode length. For a horizontal deflection of the tip the actuator bends, carrying the tip holder and tip with it. The bending is discussed in appendix A.4. Figure A.1 in the appendix shows the bent actuator.

Elongation or contraction of the piezoelectric actuator can be calculated from the vertical sensitivity, which is defined as

$$r_z = \frac{v}{V_z},$$

where v is the amount of vertical length change and V_z is the applied common voltage. The z direction is arbitrarily defined as the direction along the tube axis, which is more or less perpendicular to the sample. Cross-coupling between the different axes is disregarded. The sensitivities of the actuators are approximately constant for small deflections. The vertical sensitivity is calculated in appendix A.4, equation (A.12) as

$$r_z = (-3.36 \pm 0.23)\ \mathrm{nm/V}.$$

The horizontal sensitivity r_h of the actuator is the amount of horizontal deflection of the tip h per unit of applied voltage V_h. That is

$$r_h = \frac{h}{V_h}.$$

[5] I am indebted to Kai Dirscherl [59] from the Danish Institute for Fundamental Metrology for this information.

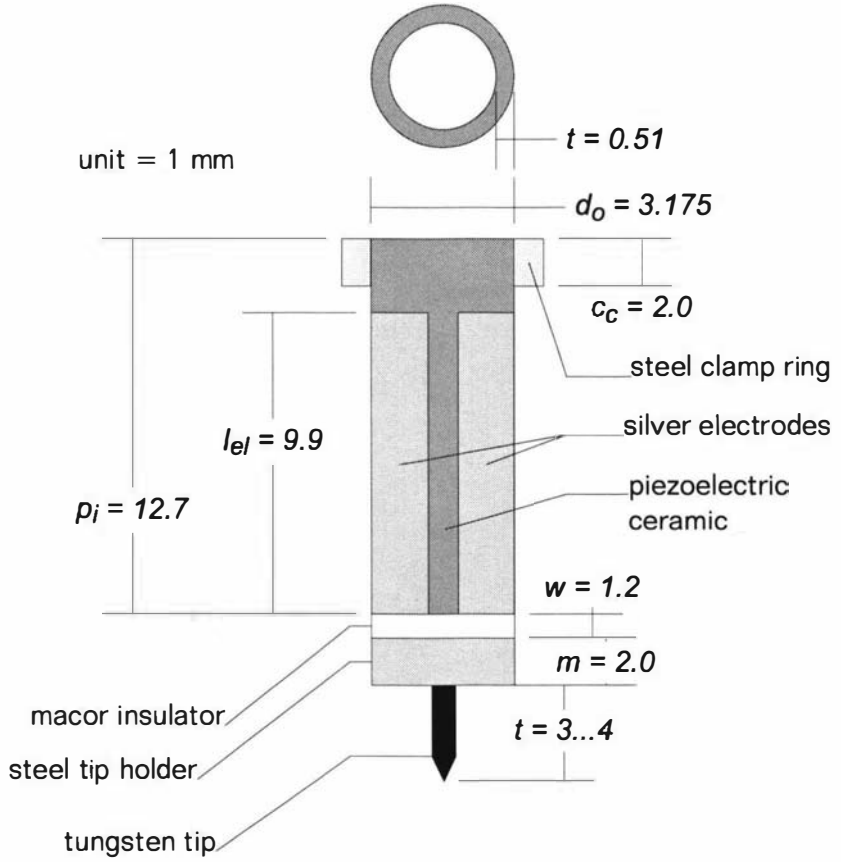


Figure 3.4: **Center Actuator:** The complete actuator is shown in cross-section and as a side view. The compound actuator consists of the piezoelectric ceramic tube, the tip holder plus tip and an electric insulator between tube and tip holder. The dimensions of each actuator component are shown in the diagram and are also listed in table A.1 of the appendix along with other tube properties. This information will be referred to when calculating the sensitivity, vibrational modes and thermal drift of the actuator. A diagram of the complete probe head plus sample holder is shown in figure 3.16.

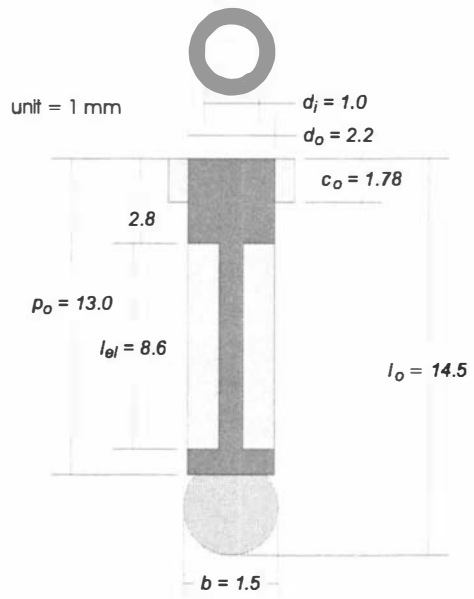


Figure 3.5: **Outer Actuator:** The complete actuator is shown in cross-section and as a side view. A small steel ball is glued to the bottom of the outer actuator. The dimensions of the tube and the ball are given in the diagram. The electrode covers only a part of the tube. Probe head properties such as vibrational modes and thermal drift are calculated from the dimensions and material properties, which are listed in appendix table A.2. For the calculation of thermal drift the same labels for the lengths of the pieces are used in table A.6. The diagram of the outer actuator is a detailed view of a more complete diagram, figure 3.16.

Because the tube is cylindrically symmetric the horizontal sensitivity is valid for both x and y directions. The voltage is applied to opposing quadrants as equal and opposite voltages, for instance $+V_h$ to quadrant $+x$ and $-V_h$ to quadrant $-x$. In appendix A.4 a horizontal sensitivity of

$$r_h = (-27.0 \pm 2.6)\,\mathrm{nm/V}$$

was found (equation (A.10)). The horizontal sensitivity is dependent on the tip length and a typical value of 3 mm was assumed here.

For the outer actuators only the horizontal deflection is of interest. A detailed diagram of the outer actuator is depicted in figure 3.5. The horizontal sensitivity of the outer actuator was calculated in the appendix in equation (A.14), and was found to be

$$r_h = (-18.9 \pm 5.7)\,\mathrm{nm/V}.$$

Soldering Since the piezoelectric ceramic is in danger of depoling when overheated, it is challenging to solder to the piezoelectric actuators. The incurred heat of the soldering iron will increase the temperature of the ceramic and accelerate aging processes. Another manufacturer of piezoelectric ceramics, Philips [152], recommends that the soldering should not last more than 3 s at temperatures between 250 and 300 °C. Tin-lead solder (60% tin, 40 % lead) is recommended together with standard non-corrosive flux.

For the electrodes in this project standard electronic tin-lead solder and

Comweld 965 solder flux was used with a 50 W iron (Antex XS-D) at temperatures of around 300°C.

Soldering was made more difficult in the case of the center actuator, since the silver electrodes were tarnished. It was impossible to solder to the first actuator, apparently the electrodes had completely turned into sulphide, since the silver electrodes are only micrometers thick. Soldering to a second actuator was then attempted, in the hope that complete tarnishing had not yet occurred. To remove the tarnish it was treated with a 3% ammonia (NH_4OH) solution, this being recommended by a manufacturer of silver oxygen probes (YSI 5331 Oxygen Probe from Yellow Springs Instruments). The actuator electrodes were then soldered, although still with some difficulty. It cannot be guaranteed that the ceramic has not been damaged.

Piezoelectric ceramics should not be exposed to excessive humidity. This is probably the reason why Philips and another manufacturer, Sensor Technology, recommend the use of a mild abrasive or india-rubber eraser to clean tarnished silver electrodes instead of using a wet chemical method. It is highly advisable to store piezoelectric actuators in air tight containers as was the case for the actuators delivered from Siemens. This prevents moisture from reaching the ceramic and sulfur from tarnishing the silver electrodes. Silver as an electrode material is discussed in appendix A.2.

Problems with Unequal Quadrants When doing a linescan in constant-current mode, a vertical tip movement of about the same magnitude as the horizontal movement of the tip was recorded. In constant-height mode the current decreases or increases along with the horizontal movement, depending on electrode polarity. This could not have been the slant of the sample, since the same effect was observed when the sample was turned by 90°.

It seemed likely that the problem was due to either the electronics or the tube scanner. However, the electronics was functioning correctly since there was only a common mode voltage of 1% of the differential signal applied to the actuator. Also, the sample bias voltage and the virtual ground at the tip stayed constant throughout the scan.

We seeked suggestions and comments from the community of SPM workers by posting a message on the Scanning Probe Microscopy Email List. The references cited in the following paragraphs are answers to our message.

Collins [47] pointed out that measuring the capacitance is a good way of ensuring that the ceramic is in good working order, since the capacitance between the outer and inner electrodes is proportional to the polarization (equation (3.9)). If the actuator has aged with time, been heated or suffered a strong depoling field, the capacitance will be reduced from the original value. The manufacturer states a value of dielectric constant (κ) for the ceramic just after poling. The calculated capacitance can therefore be compared to the measured one and any difference be attributed to a deterioration in poling due to aging. Alexander [4] measured five different piezoelectric actuators before and after depoling them deliberately with a hot air gun and reported an average reduction of capacitance of -30.8%. Lægsgaard also reported a reduction in capacitance by 30%. The capacitance data given by Lægsgaard, before and after re-poling, show an increase of capacitance by $(+31 \pm 8)\%$, which matches very well with the above value of Alexander's.

C (nF)					
$+x$	$-x$	$+y$	$-y$	$\sum$	
0.658	0.615	0.645	0.630	2.55	before poling
0.666	0.623	0.654	0.630	2.57	after 8 h 19 min re-poling at +180 V
0.674	0.640	0.654	0.623	2.59	after 13.5 min re-poling at +500 V

Table 3.1: **Measured Capacitance of Quadrants:** The capacitance of each quadrant of the center actuator was measured with a digital multimeter (Dick Smith Q-1426). All quadrants have equal capacitance to within 7%. The capacitance was measured before and after re-poling the piezoelectric ceramic. Re-poling only changed the measured capacitance by small amounts. The total rise in capacitance was 1.5%. In all cases the total capacitance of all quadrants agreed with the calculated value of (2.50 ± 0.43) nF from equation (A.2). It can be assumed that the ceramic was well poled before re-poling, since a depoled ceramic measures capacitances that are about 30% lower than in a poled state.

To find out whether the piezoelectric actuator was properly and evenly poled, the capacitance of the individual quadrants was measured. This was also recommended by Hines [96] and W. Smith [173]. The capacitances between each electrode and the grounded inside electrode was found to be (0.64 ± 0.02) nF for all electrodes. The measured values are listed in the first row in table 3.1, the total value of the electrode capacitance is listed in the fifth column (headed $\sum$). In the appendix, equation (A.2), a total capacitance of (2.50 ± 0.43) nF was calculated, which agrees with the measured result of 2.55 nF. Table 3.1 also shows the result of re-poling the ceramic of the center actuator, discussed below. These results suggest that the actuator has not been degraded.

Groos [81] drew attention to proper grounding and shielding. If the tip had a resistive path to ground, while not being adequately shielded from the sample, then influence charges are induced, which can be measured as a voltage[6]. This is also discussed in several electronics and instrumentation books [135, 19]. R. Smith [172] suspected a short circuit in the electrodes of the actuator, while Hipps [97] suggested to make certain that there was no leakage current between tip and actuator electrodes as well as clean the insulator between them if necessary[7].

As a result of these suggestions the resistance between any electrode of the center actuator and the tip was re-examined. The resistance was found to be not less than $6 \times 10^{13}\,\Omega$. Short circuits between individual electrodes and tip were also ruled out. This did not solve the problem however.

Other suggestions to test the sensitivity of the actuators came from Truscott [184], W. Smith[8] and Ramaswamy[9].

[6] Groos also suggested to turn the actuator or the sample by 90° or 180° and to estimate the necessary offset voltages that would cause the observed behavior.

[7] Hipps also made the interesting point that a tip that is bent due to impact on the sample may exhibit the problematic behavior. The impact would make a pit in the sample, which could have a steep slope. The bending of the tip gives a certain direction to the effect. Hipps noted that a bad approach algorithm could produce a bent tip every time that an approach is attempted.

[8] W. Smith [173] suggested the application of a low-frequency (2 Hz) sine wave to the actuator and observation under a optical microscope with a magnification of about 40×.

[9] Ramaswamy [160] devised a capacitive method to measure the length change of the piezoelectric actuator, thereby measuring their sensitivity. The technique seems to be similar to the one described in an article of Vieira [190].

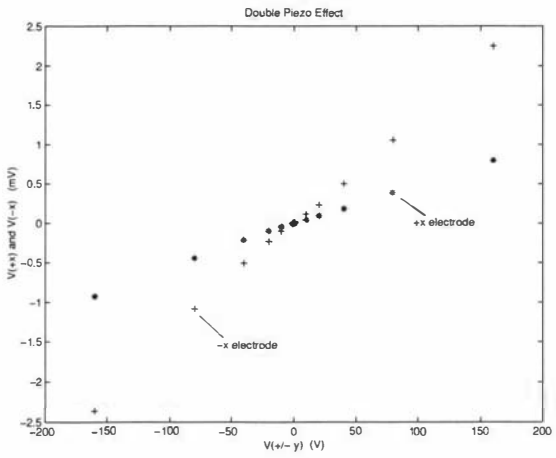


Figure 3.6: **Double-Piezoelectric Effect (y):** Following the method of measuring the double piezoelectric effect advised by Truscott [184] and also described by Chen [44], a voltage was applied to two opposing quadrants of the center actuator. Resulting voltages, generated via the double piezoelectric effect, were measured individually at the remaining quadrants. The applied voltages are ± 160 V on the two y quadrants and are shown here. The data when applying voltages to the two x quadrants is shown in figure 3.7. The strain created in the remaining quadrant is symmetric even if the quadrants with voltages applied to them do not response equally to these voltages. The resulting double piezoelectric voltage is in the range of a few millivolts. This measurement provides data that allow comparison between the individual quadrants. In principle the values can be interpreted quantitatively if the input resistance of the voltmeter is known. As a superior technique for a quantitative measurement, however, a current measuring method should be employed, since the equivalent output impedance of the quadrants is very high. The output voltage measured shows the sensitivity differences between actuator quadrants. In the two measurements the $-x$ quadrants exhibit a double piezoelectric effect with almost twice the strength of the $+x$ quadrants.

Following a recommendation by Truscott the double piezoelectric effect for the center actuator was measured. This effect is the result of using the piezoelectric effect to cause a strain and the converse effect to generate a voltage. A voltage is applied to one or more but not all electrodes of the actuator. The strain in turn generates a voltage on the remaining electrodes. The measurement should be done symmetrically, i.e. a voltage should be applied to an opposing electrode pair, for instance the $\pm x$ pair, and the resulting voltage should be measured on the orthogonal pair, which is the $\pm y$ electrode in this example. The procedure is described in the introductory book of Chen [44]. The recorded data are graphically presented in figures 3.6 and 3.7. The measurements were meant to provide data that would allow comparison between the individual quadrants. The graphs show that two quadrants, $-x$ and $+y$, respond almost twice as strongly to the applied voltages as the other quadrants.

After finding unequal response from the quadrants, the actuator was re-poled. Firstly the ceramic was re-poled at $+180$ V, which agrees with a recommendation in [44], where a re-poling field of about $E = 10\,\mathrm{V}/25\,\mu\mathrm{m} = 400\,\mathrm{kV/m}$ for more than 8 hours is suggested. Lægsgaard [122] described a technique of re-poling depoled piezoelectric actuators, where the actuator ceramic is poled at 1000 V via a 10 MΩ resistance[10] for a few minutes. His actuators had wall thickness between 0.5 mm and 0.7 mm. For the actuator used in this micro-

[10] Lægsgaard used a voltmeter.

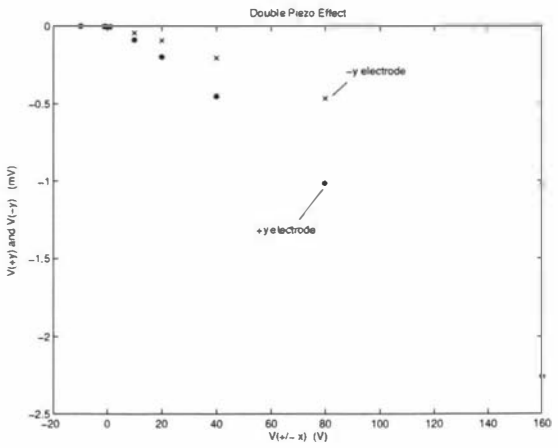


Figure 3.7: **Double-Piezoelectric Effect (x):** The graph shows the second half of the results presented in figure 3.6. The applied voltages are between -10 V and 160 V for the two x quadrants. Similar to the results in figure 3.6, the $+y$ quadrants exhibit a double-piezoelectric effect twice that of the $-y$ quadrants.

scope, which has a wall thickness of 0.51 mm, the applied voltage[11] should be approximately +200 V. The ceramic of the center actuator was also re-poled at high voltages over several hours. For the second re-poling a high-voltage supply was used with a voltage of +500 V as mentioned in table 3.1. The capacitance of the actuator electrodes increased by only 1.5%. Since all electrodes show relatively even capacitance and the re-poling is very small, it can be assumed that the ceramic is poled as well as it can be.

Although the observed tilt during a linescan appeared to be smaller after re-poling, the actuator behavior remained asymmetric. By using a graphite sample and scanning an area of roughly one hundred nanometers squared, the center actuator was calibrated. A scan was taken and displayed without slope correction. After a few repetitions, multiplication factors for the quadrants were found that rendered a reasonably flat image of the sample. The control voltage for the $+x$ and the $-y$ quadrant had to be multiplied by 2.2, which agrees well with the above finding of the double piezoelectric effect.

The fact that re-poling had hardly any effect on the actuator may be due to irreversible changes in the piezoelectric ceramic. Probably the only solution is to replace the actuator. It is not clear why a change in polarization should not be measurable as a accompanied change in capacitance.

3.1.3 Movement of the Probe Head

The probe head can be moved sideways by a concerted 'skip' of the outer tubes: all outer tubes contract, bend in the wanted direction and then elongate again to their initial length. A sideways movement can be either circular or horizontal relative to the sample holder. When the microscope moves in the circular direction, the three outer actuator tubes 'walk' up or down the ramps they sit

[11] The field strength mentioned is also the one that will cause depoling of the ceramic if applied opposite to the polarization direction of the ceramic, i.e. if -200 V is applied to the outer electrodes.

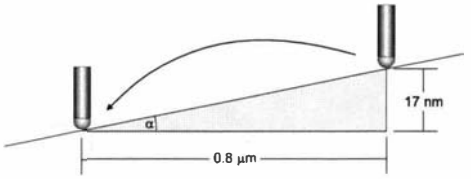


Figure 3.8: **Typical Step Size:** By moving horizontally along the sample holder ramp, the probe head is slowly lifting or lowering itself. The typical horizontal step size is approximately $0.8\,\mu$m, which results in a vertical shift of 17 nm.

on. As a result, a clockwise circular movement lowers the microscope and an anti-clockwise movement raises the microscope.

The center actuator can also be moved horizontally and vertically. Movement in all three dimensions are needed when scanning the tip over the sample surface and adjusting the distance between the tip and sample surface. In general one would move the microscope as a whole to cover greater distances of hundreds of micrometers and use the center actuator alone to position the scanning tip on a finer scale of less than $4\,\mu$m. The coarse approach and the search for a region of interest would require movement of the microscope as a whole, whereas the fine approach and all scanning activities of the tip is done by the center actuator.

Typical Step Size of the Probe Head

Any circular movement of the tunneling unit standing on the sample holder ramps gets translated into a vertical lowering or rising of the tunneling unit by:

$$\Delta y = \Delta x \tan \alpha,$$

where Δx is the horizontal movement, Δy the vertical movement and α the slope of sample holder ramp. The ramp has a small inclination of $\alpha \approx 1.3°$. A diagram is shown in figure 3.8, which shows the typical step size. Note that moving the probe-head down or up the ramp is usually done in many small steps rather than a single big step. The outer tubes have a horizontal sensitivity of

$$r_h = (-18.9 \pm 5.7)\,^{\mathrm{nm}}/_{\mathrm{V}},$$

which was deduced in appendix A.4 as equation (A.14). With a voltage of typically 40 V the typical horizontal step size turns out to be $18.9\,^{\mathrm{nm}}/_{\mathrm{V}} \times 40\,\mathrm{V} = 760\,\mathrm{nm}$. Such a step will ideally lower the probe head by $760\,\mathrm{nm} \times 0.022 = 17\,\mathrm{nm}$.

With the experimental set-up described here the outer actuators are currently driven from just one high-voltage control signal, which has been distributed via a resistive network to the electrodes of the outer actuators. The step size in this case is only half of the above, i.e. 9 nm for a 40 V voltage step. The voltage can of course be increased to compensate for this.

Step Size Calibration of the Probe Head

The vertical motion of the probe head has been calibrated using the vertical sensitivity of the center actuator. The probe head was brought into a position where tunneling current could be detected. The tip was retracted from the sample and the probe head moved either up or down the ramp. After moving the

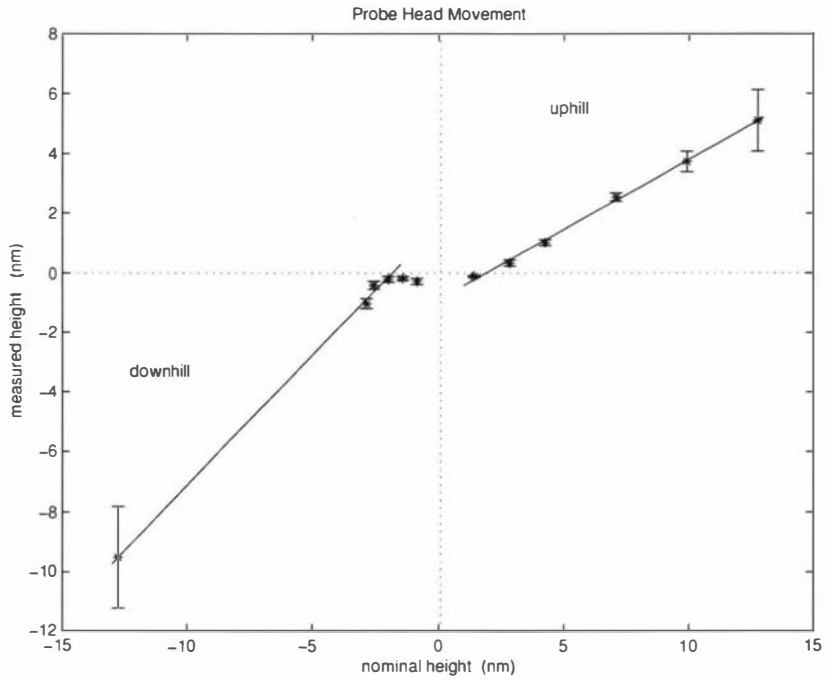


Figure 3.9: **Calibration of the Probe Head:** The amount by which the probe head moves up and down the ramp was calibrated from the center actuator. The was brought into tunneling distance before and after moving the probe head. The distance in height control voltage was noted and the change in height calculated from the vertical sensitivity of 3.4 nm/v. The measured height was compared to the ideal height calculated from the horizontal sensitivity of the outer actuators (18.9 nm/v) and the inclination α of the ramp with $\tan\alpha = 0.022$. On the right hand side of the graph, with positive ideal heights, the probe head is moving up the ramp, for negative ideal heights it is moving down. The fitted straight lines show threshold values of +1.38 nm of ideal height going up and -1.34 nm of ideal height going down the ramp. The slope for the downwards movement (1.19) is almost double as large as the one going upwards (0.64).

probe head, the tip was brought into tunneling distance again and the new tip height noted. The result is shown in figure 3.9. In the overview in section 2.3.1 the movement of the probe head was explained as stick-slip movement. During the slip phase of the stick-slip movement the outer actuators of the probe head have to overcome the static friction between ramp and ball bearings. To exert a strong enough lateral force, a voltage step is applied to the outer actuator. The steeper the change in voltage to the actuator, the larger the force will be. The height of the voltage step is programmable, while the rate of change depends on the speed of the analog electronics. Up to a maximum value the voltage rate is proportional to the desired output voltage level. If this level is too small then the force generated by the actuators will not be enough to break lose from the surface and the probe head will not move. This can be seen in figure 3.9, where a certain threshold is needed to move the probe head. To move up the ramp a voltage has to be applied to the outer actuators that corresponds to at least +1.38 nm of ideal height. This is a voltage of

$$\frac{+1.38\,\mathrm{nm}}{0.022 \times (-18.9\,\mathrm{nm/V})} = -3.3\,\mathrm{V}$$

To move down the ramp the threshold value lies at -1.34 nm of ideal height, corresponding to a voltage of $+3.2$ V.

The measured height change is less than the ideal height. This effect is more pronounced when the probe head goes up the ramp than going down, which is not surprising. The corresponding slopes of the fitted straight lines in figure 3.9 are $+1.2$ and $+0.64$. A typical horizontal step size of 760 nm moving up or down would have resulted in an ideal height change of 17 nm. Taking the threshold and the two different slopes into account this means that for such a step the probe head would actually be lowered by $1.2 \times (17 - 1.3)$ nm $= 19$ nm and raised by $0.64 \times (17 - 1.4)$ nm $= 10$ nm.

3.2 STM Tip

The tip is an important part of the microscope, since the recorded image ultimately depends on the shape, size and material composition of the tip. To image a sample surface with a very high resolution, the information obtained from the sample has to be very localized. A small point of interaction between the tip and sample is favorable, which means that the tip apex should be as small as possible. Apart from the size of the tip, its symmetry is also important as a more symmetric tip will facilitate the interpretation of recorded images.

It is possible to produce sharp tips with radii of curvature of about 100 nm [67], 50 nm [51] and even 25 nm [116, 146]. Tips of this size are sufficiently small to image relatively coarse samples such as holographic gratings or calibration samples for electron microscopy. For samples of medium roughness, having size features between 10 and 100 nm, the image reflects the tip shape as much as the sample topography and differentiating between the two becomes difficult. As the tip becomes sharper the image will reflect the sample profile more and the tip shape less.

On an atomic scale a 25 nm tip is still rather large. Most tips, however, have one or more protrusions at or near the apex. Sometimes these minitips consist of clusters of only a few atoms. Due to the exponential tunneling current relationship with tip-sample distance, the minitip closest to the sample will carry most of the tunneling current [28].

Even with atomic scale minitips protruding from the tip, it is preferable to have sharp tips, since with a sharp tip the chances are lower that there are multiple protrusions. The presence of more than one minitip can lead to multiple or shifted images being recorded, especially when the tunneling interaction switches from one minitip to another, as for example occurs when approaching an edge.

Apart from being sharp and symmetric, the tips also have to be stiff, thick and short to achieve a high resonance frequency. This is true for the etched tip-wire as a whole as well as the tapered part of the tip alone. A low-frequency vibration of the tip will cause a blurring of the acquired image. If the resonance frequency is high enough then lateral vibrations will not be excited by the tip movement.

In appendix A.5.5 the longitudinal and transverse vibration of a beam is discussed. The beam is assumed to have a uniform circular cross sectional area, consisting of one material only. Apart from the area being uniform, the assumptions are valid for the tip. The equations derived in section A.5.5 will suffice for the qualitative discussion that follows.

The lowest natural frequency for transverse vibration in a beam is

$$\omega \propto c\frac{d}{l^2},$$

where c is the speed of sound in the material, d the diameter of the beam and l its length. A similar proportionality is found for longitudinal vibration, where the lowest natural frequency is

$$\omega \propto c\frac{1}{l}.$$

To achieve high transverse and longitudinal natural frequencies, the tip should have a high speed of sound, a large diameter and short length.

3.2.1 Tip Material

An important selection criterion for the tip material is mechanical stiffness and lightness. The tip should have a resonance frequency that is as high as possible, so as not to blur the recorded image. As discussed above, the lowest natural frequency for transverse as well as longitudinal vibration are determined by the speed of sound. The two material properties that influence the speed of sound are Young's modulus (Y) and the density of the tip material (ρ). The speed of sound is

$$c = \sqrt{\frac{Y}{\rho}}.$$

The tip material chosen should have a high speed of sound, which means it should have low density and a large value of Young's modulus. The stiffer the material, the higher Young's modulus will be and with it also the resonance frequency. Material stiffness also has the advantage that the tip is less likely to deform while operating.

Another issue for material selection concerns the chemical nature of the tip. The tip has to be conductive to be used as a tunneling electrode, which means that metals are considered first as suitable tip materials. Also the chemical reaction of the tip with its surroundings has to be taken into consideration when the microscope is to be used in air, which was the case in this project. Finally, for an instrument in the making such as the one in this project it is also important that tips can be produced easily and cheaply.

The two most popular materials used for scanning tunneling microscopy are tungsten and platinum-iridium.

Noble metals are chemically very inert. They do not form an oxide layer and are unlikely to interact with the sample. Platinum is most commonly used in tip making; to enhance its stiffness, a small amount of iridium is added. Tips are prepared with relative ease by cutting a thin wire with scissors, although with more effort tips can also be electrochemically etched. Apart from making sharp tips from platinum-iridium [127, 142] other noble metals such as silver [56, 77] and gold [142, 126] have been used. The disadvantage of noble metals is their high price.

Tungsten has long been used to make sharp tips since Erwin Müller invented the field ion microscope in the 1950s. Tungsten provides a very high stiffness, can easily be produced by electrochemical etching and is a relatively inexpensive material. The drawback of using etched tungsten tips is a thin oxide layer, which is formed during the etching process.

Generally speaking, platinum-iridium tips are recommended for work in air, and tungsten tips for work in vacuum. However, since a new instrument was being developed, cheap, easily produced tips were more important than long life and so tungsten tips were chosen.

3.2.2 Tip Production

Introduction

There are in principle many ways to make tips, as reviewed by Melmed [132] and also Fotino [70]. The two most popular methods are mechanical cutting of the wire, which is often done for platinum-iridium tips and electrochemical

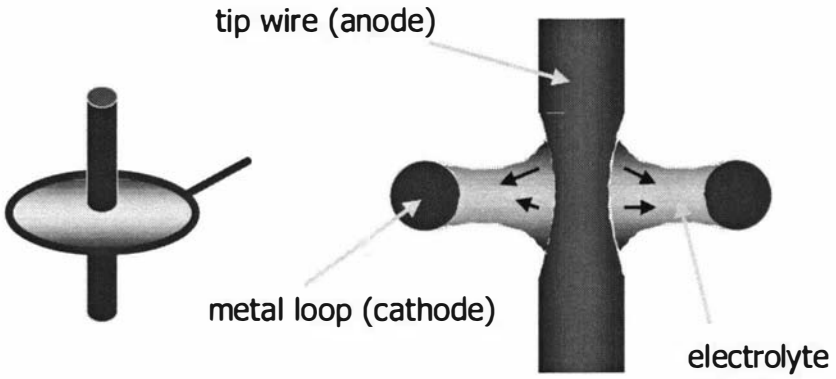


Figure 3.10: **Tip Production:** A tunneling tip is electrochemically etched from a piece of tungsten wire. Anode and cathode for the etching process are the tungsten wire and a platinum loop. The wire has a diameter of 0.25 mm, the loop is about 1 cm wide. The electrolyte used is 1 M KOH and is suspended as a lamella in the metal loop. The wire is etched with an ac current of between 30 and 3 mA.

etching, which is the preferred method for tungsten tips. In this project tips have been made from tungsten wire using the electrochemical-etching method. Electrochemical etching can produce tunneling tips that are very sharp, short and symmetric.

For electrochemical etching two electrodes and an electrolyte are needed. Of the two electrodes, anode and cathode, the tip wire is used as the anode. Typical cathode materials are platinum, graphite, stainless steel and nickel. When a small voltage is applied between the wire and the loop, the etching starts. Wire material is dissolved in the electrolyte and transported away.

Tip-Making Set-Up

In the tip-making set-up of this project, tungsten wire is used as the anode, while a metal loop made from platinum serves as the cathode. Two similar electrolytes, NaOH and KOH, are commonly used to etch tungsten, which was also the case in this project. During etching, tungsten (W) is oxidized to tungstate (WO_4^{2-}). The overall chemical reaction taking place is

$$W + 2OH^- + 2H_2O \rightarrow WO_4^{2-} + 3H_2 \uparrow,$$

with a reaction potential of 1.43 V. The etching product, tungstate, is soluble in water and is dispersed within the electrolyte by diffusion. Hydroxide (OH^-) is consumed during the reaction, which means that the electrolyte has to be replaced regularly.

The electrolyte is suspended in a small metal loop and the tip wire is placed in the middle of the electrolyte lamella as seen in figure 3.10. The etching process was controlled in two ways, the etching current was monitored and the etching region was changed by moving the loop up and down. The wire was etched with ac current, which means that the wire is the anode for only half a cycle. During the other half of the cycle, the wire is the cathode and small hydrogen bubbles are formed on the wire surface. As the etching proceeds, the wire becomes thinner and thinner until the bottom part of the wire is hanging from a very thin thread, which finally breaks.

In principle both parts of the wire can be used in the microscope. The top part is shorter, but not as sharp as the bottom part, since the electrolyte still forms a bridge between wire and loop until the loop is pulled away. The lower part is sharper, but has longer shanks since the electrolyte wets a larger part of the bottom part of the wire due to gravity. Klein and Schwitzgebel [116] suggest turning the wire over in the middle of the process to achieve more equal etching. Bryant *et al.* [36] use a similar drop-off technique and consider the bottom part to be preferable, since etching ceases very rapidly after drop-off.

Electrolyte As an electrolyte both KOH and NaOH in various concentrations have been used. Other researchers [98, 71, 67, 33, 82, 116] also report both types of electrolyte in concentrations between 1 M and 4 M. Oliva *et al.* [146] analyzed the conditions of etching tungsten tips in detail. They found that KOH at a concentration of between 2 M and 3 M works best for the immersion technique they used. For the loop technique presented here, good results were achieved with KOH at concentrations of roughly 50 g/kg. With KOH weighing 56.1 g/mol this corresponds to concentrations of about 1 M. The exact concentration is not critical other than being larger than a minimum value, such that the electrolyte remains suspended in the loop.

Etching Voltage The reaction potential is 1.43 V. To drive the reaction a voltage equal to the reaction potential plus a certain overvoltage of a few volts is needed [16]. The V-I curve during etching is non-linear. For small voltages the etching current increases linearly. This is the ohmic regime. After a voltage of the reaction potential plus the overvoltage is reached the etching current rises only very slightly with increasing voltage. This is the plateau regime of the etching process. Only at high voltages does the etching current increases again at a rapid rate.

The etching process reaches a plateau regime, because the reaction is then limited by mass transport. The ions traveling in the electrolyte can only move at a certain maximum speed, which hardly changes with applied voltage.

At small currents in the ohmic regime the etching takes a very long time, several tens of minutes, while at high currents beyond the plateau regime the etching proceeds violently. Oliva *et al.* [146] report cavities and other irregularities on the tip in this regime. In the set-up used in this project the current cannot be set too high, since the electrolyte lamella will burst when too much hydrogen is generated over too short a time. It is of no advantage to use voltages higher than the one at the beginning of the plateau regime. Much higher etching voltages will only result in electrical power being dissipated as heat in the electrolyte. In the present set-up the plateau regime begins at a voltage of roughly 2 V which results in a current of about 6 mA.

In an extensive review of tip-sharpening Melmed [132] defines the ohmic regime at small voltages as the true etching regime and the plateau as a regime where the wire is electrochemically polished. Melmed defines polishing as a process that is insensitive to the local properties of the wire surface, causing a fast and uniform removal of wire material, which is desirable in the making of sharp tips. In this project, currents of between 3 and 30 mA were used during the etching process. The etching process was started with the highest current to remove the bulk of the wetted part of the wire quickly. After the wire has

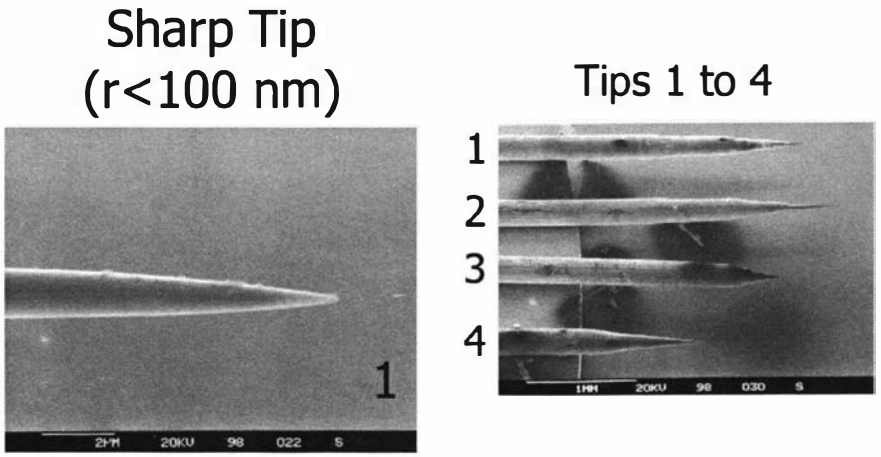


Figure 3.11: **Electrochemically Etched Tips:** Scanning electron microscope (SEM) photos of several etched tips are shown. The tips were etched from 0.25 mm tungsten wire in an electrolyte of 1 M potassium hydroxide at an ac current of 3 to 30 mA. The photo to the right is a low resolution overview shot from four tips. The photo to the left shows a good tip. 'Good' in this context means that the tip has a sharp apex in the SEM photo and also produced a good STM image.

thinned considerably, the current was decreased to about 3 mA in order to etch a well-shaped tip gently and precisely.

The use of either dc or ac voltage depends on the materials involved [132]. DC voltages allow a rather precise removal of material, since only little gas evolves at the wire. The fact that a layer of dissolved material may build up if the dissolved material is not transported away from the wire, can make dc etching difficult, because electropolishing is inhibited in the area where the layer forms. Electropolishing with ac voltages is the easier method due to gas evolution at the wire, which disperses the dissolved material. AC etching was employed in this set-up.

Etching times were several minutes depending on the age of the electrolyte, the etching current used and the immersed area of wire. Other groups report etching times of between 45 s at 25 to 30 V [67] and 20 to 50 min in the plateau regime [146]. Most groups use voltages of 3.5 to 10 V with etching times of about 5 min.

Results

Several tips, made with the technique mentioned above, are shown in figure 3.11. The photos have been made with a scanning electron microscope at different magnifications.

Tips have been produced that fulfill the requirements of being short, sharp and symmetric. The tips have a typical shank length of 400 μm, while the radius of curvature appears to be less than 100 nm.

Unstable Tunneling Junction

Problem tips, that is those which do not produce good surface images, normally have an unstable tunneling current. This could be due to two reasons; the tip

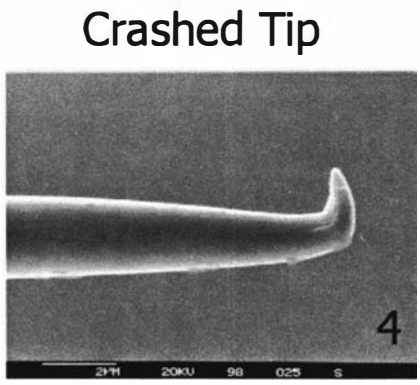


Figure 3.12: **Crashed Tip:** The photo, made with a scanning electron microscope, shows an example of a crashed tip.

may have an oxide layer covering it or it may have multiple protrusions.

Hockett and Creager [98] report current spikes due to untreated oxide layers. When etching tungsten electrochemically the build-up of an oxide layer (WO_3) during the etching process cannot be avoided. The oxide layer has a thickness of about 5 to 10 nm, which is thicker than the native oxide layer on tungsten. The layer can be removed by high vacuum annealing at high temperatures [51] or by chemical dissolution with hydrofluoric acid (HF) [98]. Since vacuum equipment was not available and HF is a very hazardous substance, the oxide layer was not removed. Before tunneling with atomic resolution is possible, the tip still has to be treated, which can be done in situ by gentle collision with the sample or by high-field treatment. Both techniques will not completely remove the oxide layer and as a consequence tungsten tips formed in this way should not be used for the recording of characteristic I-V curves, since the oxide layer will make the curve look like that of a semiconductor.

Tips that are too blunt or which have been crashed, such as the one shown in figure 3.12, are likely to have more than one minitip close to the sample surface. When the tip moves horizontally, during a scan or after an induced vibration, tunneling can shift from one minitip to another, depending on the sample roughness. The speed of current change will often lie outside of the controller bandwidth. In the best case the observed tunneling current will appear noisy, in the worst case highly erratic. A crashed tip has to be replaced.

Review of Other Etching Techniques

The etching technique used in this project evolved from the immersion method. Here the bottom part of the wire is completely immersed in the electrolyte. This immersion technique is still frequently used [11, 67, 98, 146].

Fotino [70] has reviewed all common tip-sharpening techniques extensively. He dismisses mechanical tip preparation as being poorly reproducible and producing asymmetric, broad tips with possibly more than one protrusion at the apex. Fotino considers electrochemical etching to be a superior technique and discusses the immersion method at great length. Due to bubble dynamics the tips produced by the standard immersion technique all have sharp cones on the μm scale but are relatively blunt at higher resolutions with a typical radius of curvature of 100 nm. To produce sharper tips, Fotino suggests that the wire be etched upside down so that instead of blunting the tip, the bubble dynamics help in sharpening the tip.

Apart from upside-down etching the standard immersion technique can be improved in two other ways. The etching region can be limited and the etching process can be rapidly terminated. Several methods to limit the etching region have been described in the review article of Melmed [132]. More recently, Hacker *et al.* [82] protected the immersed part of the wire from etching with wax, while Bourque and Leblanc [33] used shrink-tube for the same purpose. The loop technique used here naturally limits the region over which etching occurs and is convenient and reasonably reliable. A sophisticated set-up using the loop technique has recently been described by Klein and Schwitzgebel [116].

If the etching region is limited, the bottom part of the wire drops off. The etching of this part stops naturally. The top part is still part of the etching circuit and is etched further until the etching current is shut off. This can be done rapidly by means of an electronic circuit. Electronic etching-current shut-off has been mentioned by Nagahara [140], Anwei *et al.* [11], Fainchtein and Zarriello [67], Oliva *et al.* [146], Dickmann *et al.* [56] and Ibe *et al.* [105]. Just before rupture of the thread of etched wire, the etching current starts to decrease rapidly. This sudden decline is used to trigger an electronic circuit to shut off the etching current. This prevents further etching of the wire and a sharp tip remains.

The loop technique with an electronic shut-off circuit combines the advantage of short shanks production for the top part of the wire and rapid interruption of the etching current just after the bottom part drops off. An automated shut-off circuit similar to the one reported by Anwei *et al.* [11] has recently been devised and built by an MSc student in this laboratory, Mark Hunter. The tips he has produced are indeed superior in shank length, symmetry and radius of curvature. Detailed results will be reported elsewhere [104]. A combination of loop technique and electronic shut-off has not been published previously.

3.3 Vibration Isolation

3.3.1 Necessity of Vibration Isolation

Unwanted tip-sample displacement can occur when the probe head is mechanically excited by external sources of vibration, such as building movements and acoustical noise. Normal floor movements in buildings can be up to several micrometers in size. While scanning, the microscope cannot distinguish between the movements of the floor and a change in the sample topography. Despite the ground movement the tip-sample distance should ideally not change by more than about 1 pm [156]. The influence of such external movement can be minimized through the use of vibrational damping.

An external mechanical perturbation can be temporary in form of a shock, or continuous as an ongoing vibration. The damping system has to minimize both types of influence. Generally speaking, vibration and shock can be prevented from reaching the probe head and the sample by shielding the source of the perturbation, isolating the probe head or desensitizing the probe head to vibrations [91].

Containing external sources of vibration is not always practical[12], especially since many vibrations are caused by wind-induced swaying of the building, people walking or using elevators in the building or the operation of rotating machines, such as drills and centrifuges. Operating the microscope after hours will help to reduce the amplitude of unwanted vibrations, but cannot be the only means of vibrational damping.

The usual way to protect the probe head from vibrations is to isolate it, by impeding the transmission of vibrations to it, as well as minimizing the sensitivity of the probe head to vibrations. Both measures combine to form the vibrational damping system, which can be modeled as a mechanical system resonating at high frequencies isolated by a mechanical low-pass filter.

In contrast to the simple low-pass filter, commonly used in electronics, neither compliance of materials nor mass can be avoided with a mechanical low-pass filter. This means that the vibrational damping will always be a second-order system, which may resonate, if it is underdamped. A certain amount of damping in the system ensures that large resonance amplitudes are avoided. To maximize isolation more than one stage of vibration isolation is usually used.

3.3.2 Simple Spring-Mass System

A simple spring-mass system can be used as a model when designing a vibrational damping system. For the time being a spring-mass system without damping is considered. The natural frequency (ω_0) of this system is

$$\omega_0 = \sqrt{\frac{k}{m}}, \tag{3.10}$$

where k is the spring constant and m the suspended mass. Physically small and rigid construction elements will have a high spring constant k and when paired with low mass, give rise to a high natural frequency. Choosing large masses and soft springs on the other hand will give a rather low natural frequency.

[12] An extraordinary elaborate vibration isolation system involving an especially built acoustic isolation chamber buried into the building foundation was reported by Ferris *et al.* [68].

The probe head and the sample holder are designed to be small and rigid so as to give a high mechanical resonance frequency. If this approach is taken probe head resonance frequencies of several kilohertz can be achieved. In our apparatus the probe head and sample holder are isolated by two stages of vibrational damping; a pneumatically damped table and a stack of metal plates damped with rubber o-rings (viton). The resonance frequency of the isolator has to be set to a low value in comparison to the eigen-frequencies of the probe head and sample holder. Since most external vibrations are in the 1 to 100 Hz frequency band, the resonance of the vibrational damping should be at the lower end of this range [156].

The vibration isolation used in this system is primarily designed to isolate the probe head from vertical movements. A vertically mounted spring-mass system can be viewed in force balance as $mg = ku_0$, where u_0 is the static spring extension. In equation (3.10) the ratio k/m can now be replaced to yield

$$\omega_0 = \sqrt{\frac{g}{u_0}}, \tag{3.11}$$

which gives the natural frequency in terms of the static spring extension. As a consequence of equation (3.11), a spring with a static extension of 25 cm is needed to achieve a low natural frequency of 1 Hz [168]. It is not uncommon to place the probe head of an SPM on a heavy plate suspended from bungi ropes of one or more meters lengths [145]. The bungi ropes provide a very low eigenfrequency but are not well damped.

To describe the vibrational damping system more accurately the damping force must be taken into consideration. Often a damping force can be assumed that is proportional to the velocity of the oscillation. The damping force is then expressed as

$$F = -\beta \dot{u} \tag{3.12}$$

with a damping constant β.

The ground movement is driving the damped spring-mass system. The movement (u_g) is assumed to be cyclic in the following derivations with

$$u_g = \hat{u}_g e^{j\omega t}.$$

In this equation, $\hat{u}_g$ is the amplitude of the ground movement, j is the imaginary unit, ω is the assumed frequency of ground movement and t is time. Accounting for the relative movement between the ground (u_g) and the moving mass (u_m), the force on the mass is

$$F = m\ddot{u}_m = k_{gm}(u_g - u_m) + \beta_{gm}(\dot{u}_g - \dot{u}_m),$$

where k_{gm} and β_{gm} describe the ground-mass system. This results in the equation of motion given by [31, 88, 143]

$$m\ddot{u}_m + k_{gm}(u_m - u_g) + \beta_{gm}(\dot{u}_m - \dot{u}_g) = 0. \tag{3.13}$$

If a damped spring-mass system is excited into vibration then it will oscillate at its damped natural frequency ω_d, which is lower than the natural frequency of the undamped system [31, 120]. The sinusoidal motion decays exponentially

with a decay constant δ, which is related to the mass m and the damping constant β by

$$\delta = \frac{\beta}{2m}. \tag{3.14}$$

Oscillatory motion only occurs if the damping is less than a critical damping[13], i.e.

$$\beta < 2\sqrt{km}.$$

This can also be expressed as

$$\delta < \omega_0.$$

The damped natural frequency is related to ω_0 by

$$\omega_d = \omega_0 \sqrt{1 - \left(\frac{\delta}{\omega_0}\right)^2}. \tag{3.15}$$

As well as the decay constant δ, a quality factor (Q) is sometimes quoted to describe damped spring-mass systems. The quality factor is defined as the ratio of the spring extension at resonance to the static extension [120]. The quality factor Q indicates how fast a excited vibration dies away. A high Q means that the oscillatory motion continues over a long period, whereas a low Q indicates rapid decay of the oscillation. The quality factor Q can be expressed by δ and ω_0 as

$$Q = \frac{\omega_0^2}{2\delta\sqrt{\omega_0^2 - \delta^2}} \approx \frac{\omega_0}{2\delta}. \tag{3.16}$$

The approximation of $\delta^2 \ll \omega_0^2$ in equation (3.16) is often used, but is only valid for light damping. Nelson [143] uses the approximate relationship between the decay constant and the quality factor in equation (3.16), i.e.

$$Q \approx \frac{\omega_0}{2\delta} = \frac{m\omega_0}{\beta}. \tag{3.17}$$

as the definition of the quality factor.

3.3.3 Transmissibility

Transmissibility is defined as the transfer function of the isolation system [143]. The overall transmissibility (V) of the vibration isolation system plus the probe head is the product of the vibration isolation transmissibility (T) and the tip-sample transmissibility (S),

$$V = TS. \tag{3.18}$$

Vibration Isolation Transmissibility

Equation (3.13) can be solved and re-grouped to find the transmissibility (T) of the spring-mass system, which is the ratio of the extensions of the driven mass ($\hat{u}_m$) and the driving motion ($\hat{u}_g$) at a given driving angular frequency (ω). Using the natural frequency of the ground-mass system (ω_{gm0}) defined

[13]The authors in Harris' book [91] define a fraction of critical damping (ζ) as $\zeta = \frac{\beta}{2\sqrt{km}}$. The condition of less than critical damping is then expressed as $\zeta < 1$.

in equation (3.10) and the quality factor (Q_{gm}) defined by equation (3.17) the transmissibility is found to be [50, 143]

$$T = \frac{\hat{u}_{\mathrm{m}}}{\hat{u}_{\mathrm{g}}} = \frac{1 + j\frac{\omega}{\omega_{\mathrm{gm0}}}\frac{1}{Q_{\mathrm{gm}}}}{1 - (\frac{\omega}{\omega_{\mathrm{gm0}}})^2 + j\frac{\omega}{\omega_{\mathrm{gm0}}}\frac{1}{Q_{\mathrm{gm}}}} \tag{3.19}$$

Equation (3.19) can be analyzed at the three frequency bands of interest; the extremes of very low and very high frequency and the resonance frequency. For low frequencies $\omega \ll \omega_{\mathrm{gm0}}$ the transmissibility is 1, which means that there is no vibration isolation and the mass follows the ground movements. At resonance and for high quality factors $Q_{\mathrm{gm}} \gg 1$ the transmissibility is

$$T \approx -jQ_{\mathrm{gm}}. \tag{3.20}$$

The system amplifies the ground movements by a factor of Q_{gm}. The transmissibility in equation (3.20) is complex and the factor $-j$ means a phase shift of $-90°$.

In the band of frequencies

$$\omega_{\mathrm{gm0}} \ll \omega \ll Q_{\mathrm{gm}}\,\omega_{\mathrm{gm0}}$$

the transmissibility is found to be

$$T \propto -\left(\frac{\omega_{\mathrm{gm0}}}{\omega}\right)^2, \tag{3.21}$$

while at very high frequencies $\omega \gg Q_{\mathrm{gm}}\,\omega_{\mathrm{gm0}}$ the transmissibility only decreases like

$$T \propto -j\frac{\omega_{\mathrm{gm0}}}{\omega}\frac{1}{Q_{\mathrm{gm}}}. \tag{3.22}$$

Schmid and Varga [168] mention that elastomers have internal damping, which makes them not as stiff as the pure viscous damper assumed in equation (3.12), so that equation (3.21) also holds for these materials at very high frequencies.

The vibration isolation is most effective at high frequencies. A compromise has to be made when choosing a quality factor, since a high quality factor increases the transmissibility at resonance but decreases it at high frequencies. A small transmissibility at high frequencies is highly desirable, but with a high Q factor an accidentally excited vibration of the system will not die away rapidly.

To reduce the transmissibility further, more than one damping isolation stage was used to isolate the probe head from the ground. The individual transmissibilities of the stages are the multiplied to give the total transmissibility of all stages [88]. At very high frequencies the total transmissibility will vary with the number of stages as

$$T = T_1 T_2 \ldots T_n \propto \frac{1}{\omega^{2n}}, \tag{3.23}$$

which has been also found by Schmid and Varga [168]. Schmid and Varga conclude that multiple stages improve the vibration isolation at high frequencies, but that they may also increase the transmitted vibrations at the resonance frequencies of the individual stages, depending on the location of these frequencies and the quality factors.

Tip-Sample Transmissibility

Although there are many possible vibrational modes for the probe head and sample holder the system can be approximated by a simple spring-mass system with a resonance at the lowest vertical vibrational frequency of the tip-sample system. The transfer function can again be found from equation (3.13). For the probe head, however, it is the difference between tip movement (u_t) and sample movement (u_s) that is important. This difference is used to define the tip-sample transfer function (S) as

$$S = \frac{u_t - u_s}{u_s}.$$

This type of transfer function is known as relative transmissibility as opposed to absolute transmissibility, which was the definition of T [50]. The sample movement (u_s) is equal to the mass movement (u_m) of the last isolation stage (T_n).

If in equation (3.13) the ground and mass extensions are replaced by those for the tip and sample then a simple relation between S and T is found, which is

$$S = \frac{u_t}{u_s} - 1 = T - 1.$$

Note that both transfer functions are complex.

From equation (3.19) the tip-sample transmissibility is found to be

$$S = \frac{(\frac{\omega}{\omega_{ts0}})^2}{1 - (\frac{\omega}{\omega_{ts0}})^2 + j\frac{\omega}{\omega_{ts0}}\frac{1}{Q_{ts}}} \tag{3.24}$$

For low frequencies $\omega \ll \omega_{ts0}$ the transmissibility is proportional to the square of the driving frequency ω, i.e.

$$S \propto \omega^2. \tag{3.25}$$

At high frequencies $\omega \gg \omega_{ts0}$ the transmissibility tends to unity, while it is approximately equal to the quality factor Q_{ts} at resonance. In summary, the tip-sample system behaves like a mechanical two-pole high-pass filter.

Overall Transmissibility

The overall transmissibility (V) is

$$V = ST = \frac{u_t - u_s}{u_s}\frac{u_s}{u_g} = \frac{u_t - u_s}{u_g}.$$

This can be interpreted as the product of the transfer functions of a low-pass filter and a high-pass filter. A plot of the overall transmissibility is shown in figure 3.13. For an effective vibrational damping system the natural frequency of the high-pass filter, ω_{ts0} of the tip-sample system, should be much higher than the natural frequency of the low-pass filter, which depends on the vibration isolation stages.

At low frequencies the overall transmissibility rises with the square of the frequency in accordance with equation (3.25) until the low-frequency cut-off at ω_{gm0} is reached. If this frequency and the resonance frequencies of the damping stack are different, which is the case in the system described here, then

there is a flat area of overall transmissibility between these two frequencies. Between the stack-resonance frequency and the high frequency cut-off at ω_{ts0} the transmissibility falls with

$$V \propto \frac{1}{\omega^{2(n-1)}} = \frac{1}{\omega^6}.$$

The number of stages is $n = 4$ in the current design with three plates comprising the damping stack plus the air table. After ω_{ts0} the overall transmissibility V falls even more steeply with

$$V \propto \frac{1}{\omega^8}.$$

The most critical frequency region is the low-frequency cut-off point, where the overall transmissibility reaches its highest values at the resonance frequencies of the air table and the damping stack.

Schmid and Varga [168] analyzed vibration isolation systems and calculated the overall transmissibility. According to equation (3.23) the transmissibility of a damping stack with n plates falls with ω^{-2n}. Schmid and Varga point out that due to the maximum amplitude being at the resonance of the stack, already two plates should provide enough damping. Two plates means the number of isolation stages is $n = 3$, which is only one less than in this project. The third plate will increase the overall transmissibility at around the resonance frequency of the damping stack, which is not desirable. However, at the same time it provides an extra amount of damping at higher frequencies. The number of plates can easily be changed to suit the encountered noise spectrum of the laboratory; several experiments to that extent would need to be made in the future.

3.3.4 Air Table and Damping Stack

A pneumatically damped table, or air table, is the first stage of the vibrational isolation system. A stack of metal plates as a second isolation stage sits on top of the air table. The probe head including the sample holder sits on top of this damping stack. A photo of the air table and damping stack can be seen in figure 2.13.

The air table is a commercial vibration isolation table, model Micro-g 63-531 from TMC [180]. It fulfills the soft spring and large mass requirements. The elasticity is provided by pneumatically damped membranes in the table legs with a spring constant of approximately $4\,^{kN}/_{m}$. The heavy table top weighs about 100 kg. The manufacturer states a vertical natural frequency of 1 Hz and also specifies a horizontal frequency of 1.2 Hz. The pneumatic isolators installed by the manufacturer have two air chambers interconnected with a small orifice to provide damping. A quality factor of $Q \approx 3$ is cited for such a system.

The manufacturer specifies a vertical isolation efficiency at 5 Hz of 60% to 95% and at 10 Hz of 80% to 97% .

The damping stack consists of three concentric brass plates spaced apart by viton o-rings. The stack is part of the vibration isolation and should in principle have large mass and soft springs. The mass of each brass plate is 800 g. The o-rings are of the type 100 × 3 V80 from Engineering Plastics [66], with an outer diameter of 100 mm and a thickness of 3.0 mm. Judging from equation (3.11) the 3.0 mm thick viton rings cannot provide a very low resonance frequency.

However, the criteria are not as stringent as they are for the first stage, since the stack is already isolated by the first stage. The main purpose of the rings is to absorb a passing shock or vibrational movement by the high internal damping.

Viton is a fluorocarbon elastomer that has been designed to be incompressible to fulfill its purpose as a seal [158]. Under pressure the o-ring as a whole deforms and consequently the viton o-rings do not have a large compliance. To be able to seal fluid-containing containers under high pressure the rubber even has to have a minimum hardness [66]. Viton behaves like a viscous fluid with high surface tension, the viscosity providing high internal damping.

In their analysis, Schmid and Varga [168] point out that the spring constant of viton changes with the load placed upon the o-ring. They approximate the change as being a linear increase with the load and come to the conclusion that it is advantageous to have a damping stack with plates decreasing in mass from bottom to top. However, in this project three plates of equal mass have been used so far. Viton o-rings provide a spring constant of the order of $100\,^{\mathrm{kN}}/_{\mathrm{m}}$ [168, 147]. The static extension of the o-rings was measured and found to be less than 0.05 mm, which means a spring constant of less than $160\,^{\mathrm{kN}}/_{\mathrm{m}}$. With a mass of 0.8 kg the resonance frequencies of the damping stack will therefore be at around 70 Hz, although each plate will vibrate at a slightly different frequency due to the different loading. Using a softer type of rubber will shift the resonance of the damping stack to lower frequencies, but possibly afford less internal damping. The longer the total length of rubber is, the stiffer the equivalent spring constant. To decrease the stiffness of the damping elements shorter pieces of rubber can be used as well as pieces of a tube instead of a full-bodied ring.

3.3.5 Vibrational Modes of Probe Head and Central Actuator

Since the probe head is a mechanical structure it can be excited into vibration. It is important to know the longitudinal vibration frequencies of the probe head assembly, since the lowest order vertical vibration determines the maximum possible acquisition speed of the microscope. The resonant frequency should be as high as feasible, which was assured by designing the probe head and the sample holder so as to be small and rigid.

Figure 3.16 shows a section through probe head and sample holder. The probe head consists of five pieces, the disk, three outer actuators and one center actuator, where the actuators are clamped into the disk. The disk is supported by the outer actuators, while the center actuator is free.

The probe head can be seen as a simple spring-mass system with the disk as the mass and the three outer actuators jointly as the spring. Against the mass of the disk the masses of the actuators are negligible.

Consider a piece of material to be used for the outer actuator with length l, section A and Young's modulus Y. To achieve a high spring constant it is favorable to have a short thick piece of high elasticity Y. This can be seen from equation (A.19)

$$k = \frac{AY}{l},$$

derived in the appendix. The mass of the probe head should be as small as possible, while still providing sufficient static friction for the ball bearings.

The outer actuators can make the disk as a whole vibrate laterally, longitudinally and torsionally. The disk itself can vibrate, too, and finally the center actuator can vibrate laterally and longitudinally.

Center Actuator The center actuator operates the tip and so its mechanical behavior strongly influences the whole system. The actuator is clamped in the center of the disk, while the opposite end has the tip holder attached to it and is free to move. The free end is designed to move in all six directions of space, hence the actuator can also vibrate laterally as well as longitudinally. A drawing of the actuator is shown in figure 3.4.

The vibrational modes of the central actuator have been discussed in detail in appendix A.5.1, where the expected numerical values have been calculated. There is more than one natural frequency for the actuator when vibrating longitudinally, i.e. there is a fundamental frequency and higher harmonics. The frequencies have been calculated using different assumptions. In the first case a standing wave in the whole length of the central actuator is assumed, while in the second case the actuator is seen as a vibrating spring-mass system.

Which one of the assumptions is more likely depends on the reflection coefficient of sound waves at the border between the insulator and the tip holder. Multiple reflections will occur. Their magnitude depends on how much of the incident sound wave penetrates into the other material. The reflection coefficient is proportional to the speed of sound. The speed of sound for the various materials used in the actuator is listed in table A.3. Macor and steel share similar speeds of sound, while the speed for the actuator material is about half this value. The penetration of the sound waves is therefore regarded as weak and the actuator is regarded as a spring-mass system with the ceramic as an elastic spring and the macor insulator plus steel tip holder as a mass. This conclusion is supported by the fact that the natural frequencies of a vibrating beam only depend on one material property, the speed of sound, which is discussed in appendix A.5.5.

When the actuator vibrates longitudinally its length changes due to Hooke's law. The border between actuator and macor is moving and accelerates and decelerates the mass given by the macor and the steel tip holder. The acceleration of the mass takes longer than the acceleration of a free end. Hence, the natural frequency of the longitudinal vibration is expected to be lower than it would be if the actuator was free. In equation (A.23) the fundamental frequency was found to be

$$f_{\text{long}} = (47 \pm 3)\,\text{kHz}.$$

The resonance frequencies of longitudinal vibration of the center actuator were measured. The resonance frequencies that were found are listed in table 3.2. The

f (kHz)					
21.2 ± 0.5	22.5 ± 0.6	25.6 ± 0.6	44.4 ± 0.6	82.4 ± 2	375 ± 9

Table 3.2: **Measured Longitudinal Resonance Frequencies:** A sine wave generator was connected to all electrodes of the center actuator and the capacitively coupled output of the current-to-voltage converter was observed on an oscilloscope. In resonance the output of the converter would rise in amplitude.

measured frequency at (44.4 ± 0.6) kHz matches well with the calculation from above where the system is seen as a spring-mass system. The lower frequencies are probably due to vibrational modes of the probe head rather than ringing of the center actuator.

For lateral vibrations the center actuator has to bend and again there is a lowest natural frequency and higher order vibrations. The calculated spring constant k in the appendix used for these calculations depends on geometric and material properties and following the same line of argument as above the lowest frequency in the bending mode is likely to be

$$f_{\text{lat}} = (7.9 \pm 0.3)\,\text{kHz}$$

from equation (A.26). Here the macor insulator is included as part of the mass load. The same experiment as mentioned above was also repeated to find possible lateral ringing at around 7 or 8 kHz by applying voltages to only one quadrant of the center actuator. No resonance was found. This is probably due to the fact that the measurement method is not adequate, since one part of the actuator is stretched, while the other is compressed, resulting in an overall capacitively coupled current of about zero.

The resonance of piezoelectric ceramic tubes has also been discussed by Tapson and Greene [178], who find similar values to the ones calculated above. Tapson and Greene underline the importance of clamping the center actuator rigidly into its base, so as to avoid actuator-base vibrations. In this design the center actuator is indeed firmly clamped in the disk by means of a steel ring and a small screw.

Taylor [179] did a more detailed calculation of the longitudinal and lateral vibration modes of tube shaped piezoelectric actuators. He introduced a mass ratio μ between actuator mass and tip holder mass and found, by numerical methods, the resonance frequencies of an actuator. Taylor found that the higher the mass ratio μ is, the lower the resonance frequency will be. The center actuator in our system has a mass ratio of $\mu = 0.49$. Here only the ceramic has been taken as the elastic element, while the macor insulator was taken as part of the mass load. From one of Taylor's graphs the reduction factors relative to the normalized frequency can be found. They are 0.675 for longitudinal vibration and 0.57 for lateral vibration. The resonance frequencies are

$$f_{\text{long}} = 0.675 \times \frac{c}{4l} = 45.7\,\text{kHz},$$

for the longitudinal (equation (A.17)) and

$$f_{\text{lat}} = 0.57 \times \frac{1}{2\pi}\left(\frac{1.88}{l}\right)^2 \frac{d}{2\sqrt{2}}c = 9.1\,\text{kHz},$$

for the lateral (equation from Taylor [179]).

The longitudinal value closely matches the finding in equation (A.22) of (44.8 ± 2.7) kHz. The lateral frequency calculated in equation (A.26) of (7.9 ± 0.3) kHz also match reasonably well with the above finding. It has to be added that Taylor made the assumption of small wall thickness relative to the inner radius of the actuator tube, which is not the case here, since 0.5 mm is not much smaller than 1.1 mm.

Probe Head The outer actuators have a certain elasticity and act like stiff springs on which the mass of the disk rests. The outer actuators may bend and thereby causing the disk as a whole to vibrate laterally or torsionally or they may compress or tense, vibrating the disk longitudinally.

The calculated frequency for longitudinal vibration of the probe head was

$$f_{\text{long}} = (12 \pm 2)\,\text{kHz}$$

(equation (A.30)), for the lateral vibration it was

$$f_{\text{lat}} = 1.06 \text{ to } 4.24\,\text{kHz}$$

(equations A.32, A.33) and for the torsional vibration it was

$$f_{\text{tors}} = 1.50 \text{ to } 6.0\,\text{kHz}$$

(equations A.34, A.35). The different frequencies for the last two types of vibration depend on the connection between probe head and ramp. If the probe head is strongly attached to the ramp the higher frequency is the better estimate, otherwise the lower one is. An experiment similar to the one of Behler *et al.* [22] could elucidate this situation.

Behler *et al.* present an extensive examination of the vibrational modes of the same type of probe head used in this experiment. They calculate a horizontal frequency of 4.2 kHz for the probe head in their set-up. Experimentally they found a resonance near this frequency together with many more resonances. Specifically in a frequency band between 0.5 and 1.7 kHz they find what they call rattling resonances, which they attribute to the contact properties between the supporting balls of the probe head and ramp.

3.3.6 Conclusion

The lowest vertical natural frequency of the tip-sample system is the longitudinal vibration of the probe head at 12 kHz. The overall transmissibility (V) starts to drop after the natural frequency of the damping stack at about 70 Hz. The condition that the natural frequency of the tip-sample system be much higher than that of the isolation system is therefore fulfilled. The overall transmissibility (V) as well as the transmissibilities of the damping stages and the tip-sample system was calculated and a plot is shown in figure 3.13. At the resonance frequency of the damping stack the overall transmissibility experiences its maximum. Schmid and Varga [168] point out that although it seems at first glance better to remove the damping stack altogether, it is an important part of the isolation system, since it dampens the acoustically coupled vibrations.

The peak in figure 3.13 will in reality be not as pronounced as shown, since all three plates of the stack will vibrate at slightly different frequencies. In the future it is recommended that these eigen-frequencies be spread even further by using shorter pieces of hollow viton to achieve softer spring constants for the lower plates, and by using unequal masses for the plates with lighter ones on top.

At frequencies higher than 70 Hz and below 12 kHz the overall transmissibility V experiences a steep drop proportional to ω^{-6}. Hence, the vibration isolation system achieves a theoretical isolation of 120 dB at less than 1 kHz. If

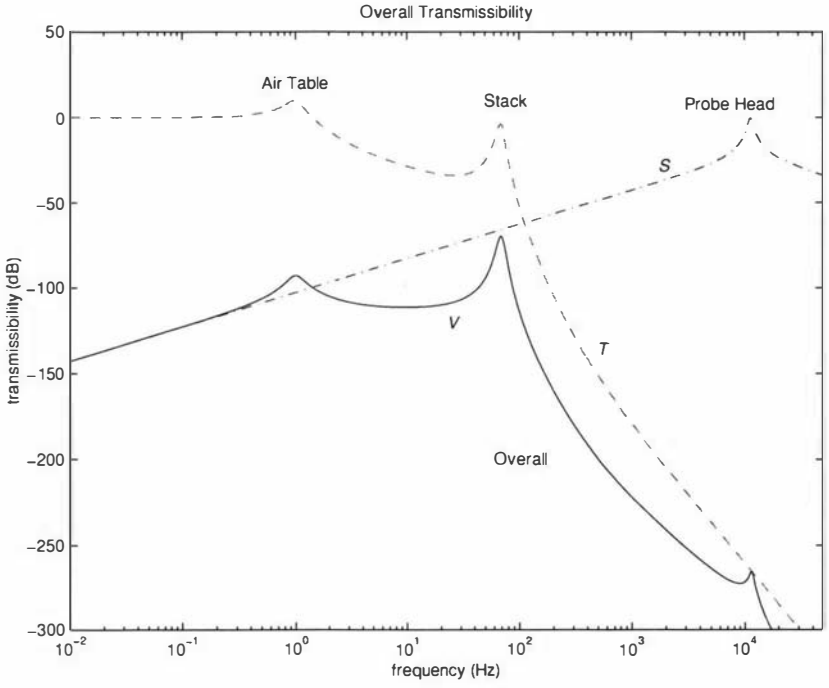


Figure 3.13: **Transmissibility:** The transmissibilities from equation (3.18), V, T and S, are plotted. The total transmissibility T includes all four damping stages. Each transmissibility was calculated from equation (3.19). The resonance frequencies were taken as 1 Hz for the air table and 70 Hz for the damping stacks. The quality factor of $Q = 3$ was given by the manufacturer for the air table and a value of $Q = 5$ for the viton rings of the stack was given by Schmid and Varga [168]. The tip-sample transmissibility S was calculated from equation (3.24) using the vertical resonance frequency 12 kHz of the probe head and a (guessed) quality factor of $Q = 12$. The maximum of the overall transmissibility is -70 dB at the resonance frequency of the damping stack. The value of overall transmissibility at the air table resonance is -93 dB, the transmissibility value at the tip-sample resonance is -265 dB.

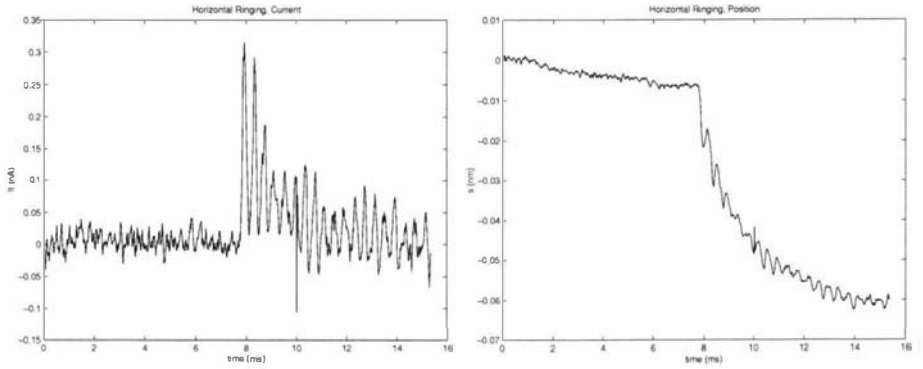


Figure 3.14: **Induced Ringing:** The graphs show recordings of tunneling current and height control voltage. The closed-loop control was active, but set to low gain. The tunneling current setpoint was 0.1 nA. At about 8 ms a voltage step of 0.027 V was applied to the center actuator to observe ringing. The voltage step is a disturbance to the closed-loop control. The disturbance dies slowly away within about 10 ms. Apparently the tip encountered a different tip-sample distance after having been moved to a different horizontal position, since the height control voltage settles to a different level.

the tip is scanned at high enough frequencies, this is sufficient isolation to reduce ground movements of 1 μm to 1 pm. Such an isolation is required if atomic resolution is to be obtained [147].

3.3.7 Horizontal Ringing

When the tip-carrying piezoelectric actuator is stepped abruptly in the horizontal direction, lateral vibrations of the center actuator and the probe head are excited. The lowest lateral vibration frequency is that of the probe head, somewhere between 1.1 and 4.2 kHz. Vibrations at higher frequencies, such as the lateral vibration of the center actuator at around 8 kHz will also be excited. The amplitude of the lateral vibration or ringing depends on the step size.

The ringing is seen as a disturbance to the closed loop controller. The suppression of the disturbance is determined by the loop gain as discussed in section 4.2.1 and expressed in equation (4.7).

To observe ringing, the tunneling current and the height control voltage were recorded while moving the piezoelectric tube sideways. This was done by applying a voltage step to the electrodes of the tube. Figure 3.14 shows the bending-mode ringing after a step occurred at mid time of recording, just after 8 ms. The tip was tunneling at a low tunneling current ($I_t = 0.1\,\mathrm{nA}$). The gain of the tip control was set low enough to allow the ringing to show in the recorded data. The attempt to correct the tip height is reflected in the height-control voltages also shown. The control voltages settle at a new height, since the sample was not atomically flat and the tip was moved to a different horizontal position. The voltage step was 0.027 V, which corresponds to 0.73 nm using the horizontal sensitivity of $r_h = 27\,\mathrm{nm/V}$.

To measure the lateral frequency, several peaks have been counted and the periods measured. Several repetitions of this experiment show the frequency to be (2.340 ± 11) Hz. Calculating the discrete Fourier transform (DFT) reveals a bending-mode frequency of 2.34 kHz, confirming the above result. The

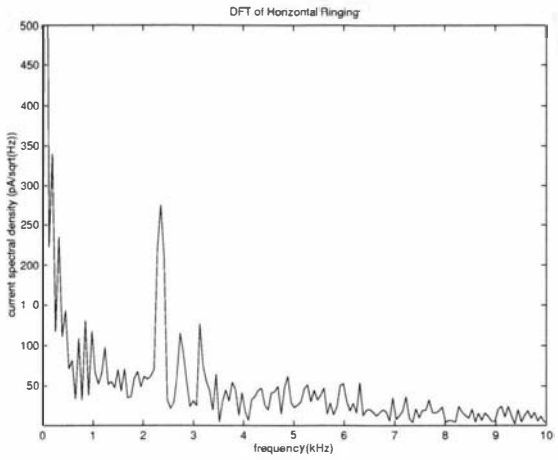


Figure 3.15: **Power Spectrum of Ringing:** Ringing of the probe head was induced by applying a voltage step to the center actuator. The power spectrum of the ringing shown here has a peak at 2.34 kHz. The number of data points was 1024.

power spectrum of the DFT is shown in figure 3.15. Whether induced ringing is observed in the tunneling current or not, depends on the closed loop gain. If the gain is high enough to compensate for the fluctuation in tunneling current, then an averaging of spatial data occurs. The controller can only adjust the tip height, whereas the tip is oscillating horizontally. The height control in constant-current mode keeps the tip at a certain constant distance from the sample, while the tip is oscillating across its surface. The observed tunneling current will be constant, while an oscillation is seen in the height control voltages. If there is no tip height control, then there will be no change in the height control voltages and an oscillating tunneling current will be seen. If the controller gain is low, then oscillation will show in both the tunneling current data as well as the control voltages, which was the case here.

As a conclusion, the horizontal vibration frequency is not altered by the feedback loop, the height controller has only influence on the vertical tip position. The oscillation will show up either mainly in the current or the control voltage. Excitement of horizontal vibrations will generally prolong the time the tip has to spend at a particular sampling point to get a sufficient number of valid tunneling current values. Hence moving the tip horizontally should be done as smoothly as possible, i.e. with gentle acceleration and deceleration instead of stepping.

3.4 Thermal Compensation

Different materials make up the probe head and the sample holder. The materials and their lengths are depicted in figure 3.16. For the microscope design a low thermal drift of the tip-sample distance is important. The microscope design has to fulfill two requirements, the lengths in figure 3.16 have to match to within the elevation of the coarse approach mechanism (Δh) and the thermal drift of the gap has to be as small as possible. When the temperature changes, the different materials that constitute the probe head and the sample holder change

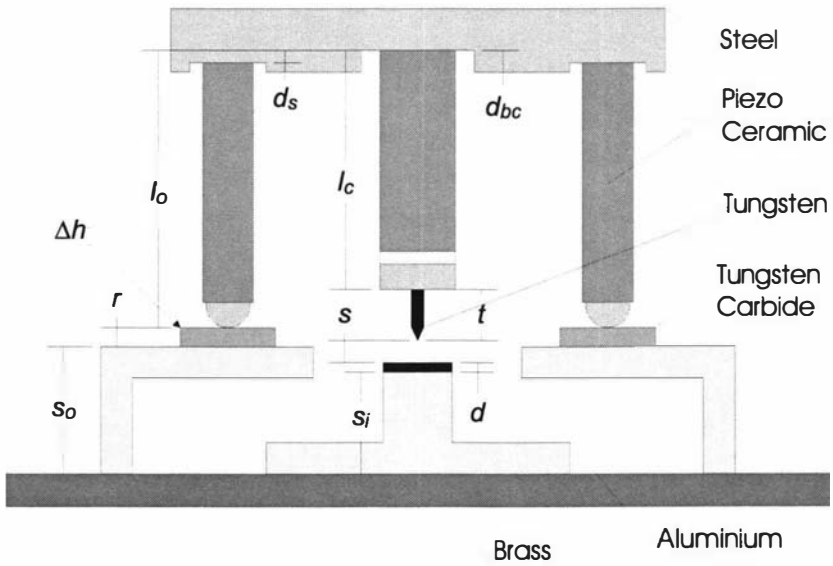


Figure 3.16: **Dimensions and Materials of Probe Head and Sample Holder:** The vertical lengths of all pieces constituting the probe head and the sample holder are summarized in this figure. The size of the probe head is grossly exaggerated to make the overview easier. Equal materials are shaded equally. The central actuator consists of three different materials, which are not labeled here, but shown in more detail in figure 3.4. The probe head consists of the outer actuators (l_o), central actuator (l_c) and head disk on top. A detailed drawing of the outer actuator is shown in figure 3.5. The tip (t) extends down from the central actuator. The sample holder consists of the sample holder cover (s_o), the sample stand (s_i), and the sample (d). Both cover and stand are mounted on a brass plate belonging to the vibrational damping. The stand is electrically insulated by a thin sheet of PTFE, which is not shown in the figure. All lengths added up leave a small gap between tip apex and sample surface (s), which for thermal drift design purposes is equal to zero. The gap can be adjusted by the coarse approach system of the probe head, since the ramp (r) has a elevation of Δh. If this elevation is not enough for a given sample, then extra steel washers (not drawn) can prop up the sample stand.

their lengths at different rates. The thermal expansion equation describes the length change to a first approximation as:

$$\Delta l = l\alpha\,\Delta T, \tag{3.26}$$

where the original length is l, Δl its change and ΔT the temperature change. The material constant α is the thermal expansion coefficient.

The gap is calculated by adding all lengths in figure 3.16, counting those that increase the gap opposite from those that decrease the gap. For thermal design purposes the original gap (s) is set to zero since it is smaller than 1 nm. The balance in lengths is

$$s = \sum_{\text{inc}} l_{\text{inc}} - \sum_{\text{dec}} l_{\text{dec}} = 0,$$

where the subscript 'inc' stands for increasing and 'dec' for decreasing the gap. In figure 3.16 the balance reads

$$s = (l_{\text{o}} + r + s_{\text{o}}) - (l_{\text{c}} + t + d + s_{\text{i}}) = 0.$$

thermal drift is generated by the difference in the individual thermal expansion coefficients. Using equation (3.26) the drift of the gap is compactly written as

$$\frac{\Delta s}{\Delta T} = \sum_{\text{inc}} l_{\text{inc}}\alpha_{\text{inc}} - \sum_{\text{dec}} l_{\text{dec}}\alpha_{\text{dec}}. \tag{3.27}$$

The gap-increasing lengths are counted positively, so a positive number for thermal drift ($\Delta s/\Delta T$) means a widening of the gap. The detailed accounting of all lengths and thermal expansion coefficients has been presented in appendix A.6, where a gap drift of

$$\frac{\Delta s}{\Delta T} = (+34 \pm 30)\,^{\text{nm}}/_{\text{K}},$$

has been found for a graphite sample with $d = 1\,\text{mm}$. The gap drift for a very thin sample is estimated as

$$\frac{\Delta s}{\Delta T} = (+13 \pm 30)\,^{\text{nm}}/_{\text{K}}. \tag{3.28}$$

In this case the sample thickness d is negligible and the sample stand is supported by two steel washers.

The overall thermal drift of the gap is rather sensitive to the lengths and materials used in the design of probe head and sample holder. While the probe head design [73] specifically addresses this problem and is discussed further below, the sample holder was initially designed with convenience in mind instead of thermal compensation.

The thermal drift was measured with a very thin sample in place. The result is shown in figure 3.17. The measured thermal drift of $-7\,^{\text{nm}}/_{\text{K}}$ lies within the estimated thermal drift range. The drift will change when a different sample is mounted or a different tip length is used.

It is worth noting that a much shorter sample holder cover and stand would result in a more predictable thermal drift. For a more detailed discussion see appendix A.6.

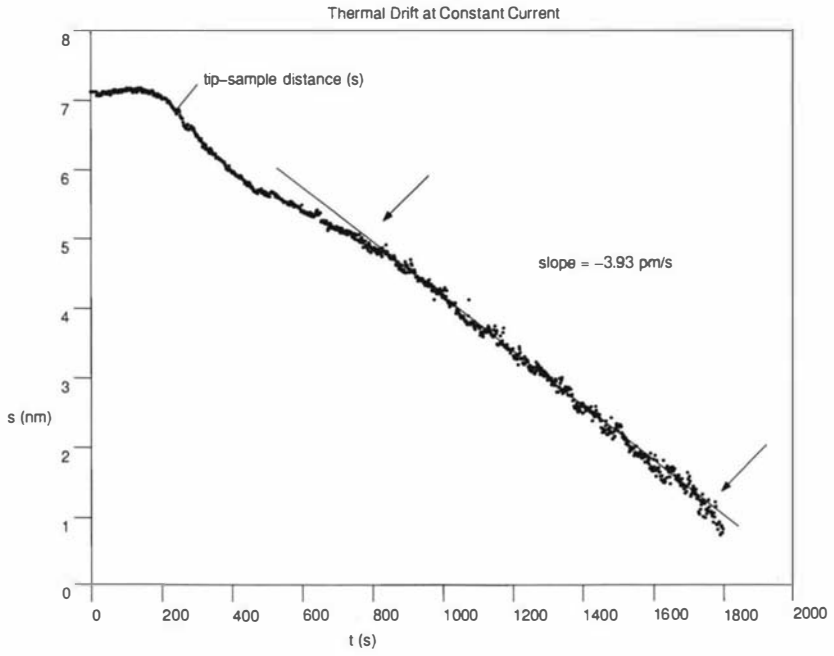


Figure 3.17: **Measured Thermal Drift:** The change of the tip-sample distance (s) with time is shown after cooling the laboratory by opening a window. The tunneling current was held constant while recording the height-control voltage for the center actuator. The tip-sample distance was calculated from the vertical sensitivity $r_z = 3.4\ \mathrm{nm/V}$. The graph shows the necessary height adjustment of the controller while the tip-sample distance changes due to thermal drift. After 800 s the tip-sample distance changes at a comparatively constant rate of -3.93 pm/s, indicated by the two arrows. At the same time the temperature change was recorded and found to be 0.56 mK/s during the period of constant thermal drift. The resulting drift is -7 nm/K.

Resulting Image Drift To get an estimate of the desired temperature stability a worst case temperature drift of $\dot{T}_{\max}$ is assumed. Over a typical image acquisition time t_{im}, the resulting temperature difference is given by

$$\Delta T_{\max} = \dot{T}_{\max} t_{\text{im}}.$$

The temperature difference $\Delta T_{\max}$ can be calculated by estimating the maximum temperature drift with

$$\dot{T}_{\max} = \frac{10\,\text{K}}{10\,\text{h}} = 2.8 \times 10^{-4}\,\text{K/s}. \tag{3.29}$$

Assuming an acquisition time of one minute, the resulting temperature change $\Delta T_{\max}$ is 0.02 K. The mechanical temperature drift of approximately 30 nm/K from appendix A.6 results in a change in tip-sample distance of

$$0.02\,\text{K} \times 30\,\text{nm/K} = 0.4\,\text{nm}.$$

The thermal drift with time is roughly

$$2.8 \times 10^{-4}\,\text{K/s} \times 30\,\text{nm/K} = 8\,\text{pm/s},$$

which is in the same order of magnitude as -4 pm/s from figure 3.17.

Not many researchers report thermal drift values. If they do, often it is a drift with respect to time without stating the change in temperature at the same time. This makes those values hard to compare to other systems, although a rough estimate as in equation (3.29) can be used. Aketagawa *et al.* [3], for example, developed a thermo-stabilized cell for their STM. They report a drift of 1.4 pm/s and a thermal stability of 1.4×10^{-5} K/s, from which a thermal drift of 100 nm/K can be inferred, which is comparatively large. Grafström *et al.* [78] mention a thermal drift of 2.5 pm/s, which is roughly

$$\frac{2.5\,\text{pm/s}}{2.8 \times 10^{-4}\,\text{K/s}} = 9\,\text{nm/K}.$$

Olin [145] reported a drift of 1.7 pm/s, which gives roughly 6 nm/K using the same temperature change as above. This is a comparatively low drift. Lyding *et al.* [128] designed an STM with temperature drift in mind. They report a drift value as low as 1 nm/K.

Chapter 4

Control

4.1 Microscope as a Control System

One of the major tasks of the microscope electronics is feedback control. While the microscope is scanning a sample, the tip has to be brought into very close proximity with the sample and subsequently kept at a distance of around 1 nm.

Electronic control of the tip-sample distance is necessary if safe approach of the tip to the sample is to be achieved, along with positioning and movement on the same scale. The only feasible way of doing this is by closed-loop or feedback control. The controller is primarily implemented in software.

4.2 Feedback Control

Figure 4.1 illustrates the relationship between controller and the controlled process in a simple feedback control system. Generally speaking, the control loop consists of the process under control and the controller. In figure 4.1 the controller is denoted by C. The process under control is described by two blocks, the process itself (G_1) and a sensor (G_2).

The blocks are considered to be linear and time invariant. Each block is described by a transfer function, which assigns gain and phase to the block as a function of frequency. The symbols within each block are also used for the transfer functions. With knowledge of the transfer function the output of a block for any given input signal can be calculated. The output is given by the convolution of the input signal and the impulse response of the block. The impulse response is the time domain representation of the transfer function.

The convolution theorem states that convolution in the time domain is equal to multiplication in the frequency domain. Working in the frequency domain the transfer functions can be thought of as operators. The convolution theorem allows one to calculate the overall transfer function of more than one block quite simply [125]. Blocks, representing transfer functions, can be arranged in series, in parallel and as a closed loop. The transfer functions of blocks in series are multiplied, the functions for parallel blocks are added. A closed loop is shown in figure 4.1.

The action of the closed feedback loop is to keep the process variable (v) close to a desired value, the setpoint (w). The process variable is compared to

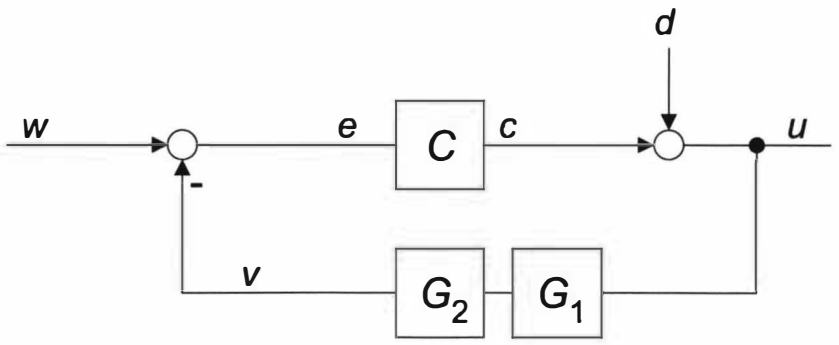


Figure 4.1: **Control Loop:** The block diagram shows a simple control loop, which employs negative feedback. Disturbances are included. The blocks stand for controller (C), process (G_1) and sensor (G_2). The last two blocks can alternatively be seen as a joined process block with $G = G_1 G_2$. The symbols in the blocks denote the transfer functions. The input to the controller is the error term (e), which is the difference between the setpoint of the controller and the process variable (v). The controller determines the direct controller output (c) from the error term. The overall controller output (u) is the sum of the disturbance (d) and the undisturbed controller output (c). The controller output is related to the setpoint by $u = [C/(1 + CG)]w$.

the setpoint and the difference (e) is fed into the controller. The output of the controller (c) drives the process. Overall the loop provides negative feedback, which means that the controller acts to decrease the process variable when it is higher than the setpoint and vice versa.

In the diagram in figure 4.1 the two transfer functions of the process and the sensor can be combined as $G = G_1 G_2$, since they are in series. The process variable is related to the overall controller output by $v = Gu$. The overall controller output is the sum of the direct controller output and the disturbances, $u = c + d$. The direct controller output is $c = Ce$, where the error term is $e = w - v$. The term C is called the forward gain of the system. The forward gain might be constant, but in general it is a frequency dependent function. Summing all terms around the loop gives

$$v = Gu = Gc + Gd = GCw - GCv + Gd \tag{4.1}$$

Solving for process variable v, the result is

$$v = v_w + v_d = \frac{CG}{1 + CG} w + \frac{G}{1 + CG} d . \tag{4.2}$$

Only the first term of the process variable (v_w) is related to the setpoint. The second term, describing the influence of disturbances (v_d) is discussed below in equation (4.7). The overall transfer function G_{cl} is the ratio of the first term of the process variable (v_w) to the setpoint (w). The ratio is known as the closed-loop gain. It is found to be

$$G_{cl} = \frac{v_w}{w} = \frac{CG}{1 + CG} . \tag{4.3}$$

The product of all transfer functions around the loop (CG) is known as the loop gain of the system.

Three important properties of feedback control can be identified from equation (4.3). They are the regions of good control, decreasing control and instability.

- **Good Control:** If the loop gain is much bigger than one then the closed-loop gain tends towards unity,

$$G_{\text{cl}} \overset{CG \gg 1}{\longrightarrow} 1. \tag{4.4}$$

  This means that the process variable is equal to the setpoint, which is the aim of the control system.

- **Decreasing Control:** If the loop gain is not large enough then the process is not well controlled,

$$G_{\text{cl}} \overset{CG \ll 1}{\longrightarrow} CG. \tag{4.5}$$

  The process variable is now much smaller than the setpoint.

- **Instability:** Under the condition that $CG = -1$ the process is not controlled but unstable. The closed-loop gain tends towards infinity,

$$G_{\text{cl}} \overset{CG \to -1}{\longrightarrow} \infty. \tag{4.6}$$

  In this case the process variable can assume any value even with the setpoint set to zero.

To summarize; for good, stable control it is desirable to have high loop gain with as little phase shift as possible. The higher the gain of the controller, the closer the process variable will be to the setpoint, while the smaller the phase shift the less likely the chance of instability.

It is the possibility of instability which usually limits the performance of closed-loop control [125].

4.2.1 Stability

Under normal operating conditions oscillations should not occur in a stable feedback loop. The loop can be called robust, if the controlled process does not oscillate even under adverse conditions. In other words the process is always stable. The feedback loop is stable as long as the feedback is negative. Taking the difference value between the actual and desired process variable is equivalent to shifting the process variable signal by 180° and adding it to the setpoint. Feeding the controller with the difference signal thus ensures negative feedback.

Stability will be lost when the controller receives an error signal e with a phase equal to the controller output c, i.e. when the phase shifts of the various feedback elements accumulate to 180° thereby giving an overall phase shift of 360°. The feedback loop will amplify these signals, oscillations will occur and the loop is then unstable. This is the situation described in equation (4.6).

Gain and phase are closely related [125]. For first and second order feedback loop elements, the phase shift exhibited by the elements increases with frequency. The frequency at which a 180° phase shift is reached is called the critical frequency. A simple strategy to prevent oscillations from occurring is to decrease the gain of the controller with increasing frequency, such that the loop gain (CG) at the critical frequency is less than one. In this case the system satisfies the Nyquist criterion of stability [17, 185].

The strategy can be realized by implementing a controller with an integrating characteristic, by incorporating a low-pass filter into the loop or by a combination of both. The gain of the control loop can be made very high at low frequencies and still meet the stability requirements.

Overall the strategy can deliver a stable controller with good accuracy at low frequencies although the bandwidth must be limited to achieve stability.

Feedback in Electronics

Feedback control is well-known in electronics. Operational amplifiers are servo controllers that follow a given input signal. The non-inverting terminal of an operational amplifier can be interpreted as the setpoint input and the inverting terminal as the input for the process variable. Stability is maintained by internally compensating the amplifier, which is to say that the forward or open-loop gain of the amplifier has a low-pass filter characteristic with a low cut-off frequency. Very good accuracy is achieved by the exceptionally high forward gain of the amplifier, which leads to the common assumption of both input terminals being at equal potentials.

Disturbances

The effect of the disturbance in figure 4.1 can also be taken into account. The transfer function of the disturbance is the ratio between process variable and disturbance. It can be found from equation (4.2) and is

$$G_{\mathrm{d}} = \frac{v_{\mathrm{d}}}{d} = \frac{G}{1+CG}. \tag{4.7}$$

The influence of the disturbances on the process variable are suppressed by the loop gain plus one. A high loop gain is therefore not only desirable so that the process variable follows the desired setpoint, but also to suppress disturbances[1].

4.2.2 Bandwidth

In this project it is desirable to have an instrument with a fast response time and a correspondingly high bandwidth. The stability condition, however, limits the bandwidth of the feedback loop. The Nyquist criterion for stability mentioned above uses the properties of the loop gain to determine the stability of the closed loop. The strategy to prevent instabilities is to decrease the gain with increasing frequency and thereby avoiding the -1 point in the complex plane of magnitude and phase. This point is shown in figure 4.2. The smaller the loop gain at 180° the more stable the closed loop, but the smaller the bandwidth. Gain and phase are often displayed in a Bode diagram. With increasing frequency the loop gain crosses the unity gain line. The difference between 180° and the phase at the gain cross-over is the phase margin of the loop, which is indicated as γ in figure 4.2. At higher frequencies the phase reaches 180°. The difference

[1] Another example from the field of electronics of a device to suppress disturbances is the linear voltage regulator. A certain fraction of the desired output voltage is the setpoint and is given by a reference voltage. The process variable is the actual output voltage while the unregulated input voltage represents a disturbance to the system. By means of high forward gain in the regulator the voltage ripple in the regulated output is almost eliminated.

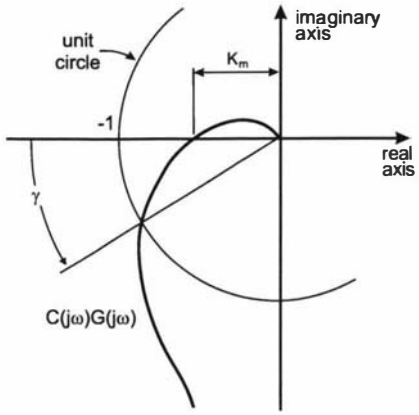


Figure 4.2: **Gain Margin:** The loop gain CG decreases with increasing frequency ω, while the phase angle increases. This is depicted on the complex plane. The decreasing loop gain CG intercepts the unit circle at phase angle γ, which is called the phase margin. Once the gain function has a phase shift of $-180°$ it has a magnitude of K_m. The fraction $1/K_m$ is the gain margin, which is usually expressed in decibels as the difference between 0 dB and CG (Figure adopted from Raven [162]).

in decibels between 0 dB and the loop gain at this point is the gain margin. The gain margin is set as small as possible without compromising stability. The chosen margins are determined by the desired robustness of the loop. As a design rule a phase margin of 30° and a gain margin of 8 dB are recommended [17].

One design goal for the feedback electronics is not to reduce the bandwidth below that imposed on the system by the piezoelectric actuators. The probe head can resonate in several horizontal and vertical modes. At resonance the phase shift is already 90°. If only an integrating controller is used, then the overall phase shift is 180° at resonance. The loop gain at the resonance frequency should be about −8 dB for a reasonable gain margin. Adding proportional and derivative terms to the controller decreases the phase shift of the controller and allows the gain cross-over to be set at higher frequencies. In section 3.3.5 the first horizontal resonance of the probe head was measured to be at 2.4 kHz. The first longitudinal resonance of the probe head was calculated to be at 12 kHz. Pohl [156] quotes a rule of thumb of control theory that states that the closed-loop control frequency should be a third of the critical frequency. If excitation of the bending mode resonance can be avoided, the bandwidth of the feedback control loop could therefore be in principle as high as 4 kHz.

Apart from the piezoelectric actuator, a second bandwidth limiting element in the system is the current-to-voltage converter. It is difficult to build a high-gain, low noise current-to-voltage converter that has a bandwidth of more than 10 kHz.

Although the highest possible frequency for closed-loop operation might be as high as 4 kHz, in many cases a smaller bandwidth is sufficient. The operator may use a lower scan speed thereby reducing the required data acquisition frequency. It is not unusual to performed line scans with frequencies as low as 10 Hz, depending on the accuracy needed [73].

4.3 Direct Digital Control

The control loop of this project is a computer-controlled system or sampled-data system. The values of the process variables, i.e. the tunneling current, are known to the computer only at discrete times. An introduction to computer-controlled systems is given in the book by Åström and Wittenmark [15]. Discrete time systems are also discussed in [17, 24, 26, 45, 124, 125]. Controllers based

on discrete time systems can be either supervisory controllers or direct digital controllers. In this project a direct digital controller was implemented.

With a supervisory system there are at least two loops of control, an analog inner control loop and a digital outer control loop. A supervisory system has the advantage that the dedicated inner loop can be fast.

In direct digital control a computer algorithm replaces the inner analog control loop. The algorithm is implemented as a program running on a digital processor. There may be more than one digital control loop implemented within one processor and a second supervisory processor might control the first one. The advantage of direct digital control lies in its flexibility and versatility.

In direct digital control both the process variable and the computed response of the controller are streams of numbers. The analog-to-digital converter samples and quantizes the process variable thereby turning a continuous variable into a discrete one. The controller operates on the incoming numbers and the response of the controller is converted back from a discrete form into a continuous form. This is termed reconstruction and is usually done by keeping the output constant over one sampling interval.

Historically more and more systems have been implemented using direct digital control as processors have become faster [15]. In this project a DSP was used. This type of processor is especially designed for fast number processing.

With a discrete time system several effects have to be taken into account that are not encountered in continuous time systems. The most important effects are aliasing and time delays.

4.3.1 Aliasing

The analog-to-digital converter has a certain sampling rate. Only signals with frequency content below half the sampling rate can in principle be restored. This fact is known as Shannon's sampling theorem [15, 110] and the maximum sampling frequency is known as the Nyquist frequency. If signals containing frequencies higher than the Nyquist frequency are sampled by the converter then part of the signal appears at lower frequencies in the sampled spectrum. This process is called frequency folding or aliasing.

4.3.2 Zero Order Hold

A digital-to-analog converter reconstructs the controller output. Usually the output is kept constant until a new digital word arrives, in which case the type of converter is known as a zero order hold. The term refers to the fact that the converter output can be represented piecewise by a polynomial of zeroth order. The zero order hold is very common due to its simplicity. It also has the advantages that it can be used with sample intervals of changing length [15] and that it does not introduce phase distortion, since the phase of the zero order hold increases linearly with frequency [115]. Reconstructing with a zero order hold does, however, introduce an error, which decreases with increasing sampling frequency [15]. This is one reason why the sampling frequency should be several times higher than the desired bandwidth of the closed-loop controller. The issues of choosing a sampling interval is discussed in section 4.4. As a consequence of the zero order hold the output voltages change rather abruptly

and usually a low-pass filter, the reconstruction filter, is used to smooth the output as discussed below.

4.3.3 Quantization

Discussions of quantization can be found in the books of Bateman and Yates [21], Klaassen [115], and Zölzer [196].

Let Q be the quantization step size. The quantization error q lies between $+Q/2$ and $-Q/2$. For the quantization error an amplitude probability distribution function is assumed that is constant between these limits and zero otherwise. To be normalized the distribution function has a constant amplitude of $f(q) = 1/Q$. The mean value of q is found to be $\bar{q} = 0$. The variance is [115]

$$\int_{-Q/2}^{+Q/2} (q - \bar{q})^2 f(q) \mathrm{d}q = \frac{Q^2}{12} .$$

The rms noise due to quantization (V_q) is

$$V_q = \frac{Q}{\sqrt{12}} . \tag{4.8}$$

The spectral distribution of quantization noise depends on the input signal. Often it can be assumed that the quantization noise is uniformly spread between 0 and the sampling frequency (f_s) [196]. The amount of quantization noise depends on the sampling rate. If the input signal is sampled at half the original rate, then the noise power is concentrated within half the original bandwidth. More noise enters the system and the rms noise would be increased by $\sqrt{2}$, while in the original system a low-pass filter attenuates the quantization noise above the Nyquist frequency. The argument holds for input as well as output quantization.

As a consequence, by choosing a high enough sampling rate, quantization noise can be decreased. With a sampling rate several times higher than the Nyquist frequency, most of the quantization noise can be removed by a low-pass filter set to a frequency just above the actual signal bandwidth. Since the Nyquist frequency is much higher than the signal bandwidth, this low-pass filter can be implemented digitally, which can be more effective than an analog filter. This process is known as oversampling in audio signal processing [196].

The fact that the reconstruction filter is set to a fixed cut-off frequency should also be taken into account. If a longer sampling time is used than was originally planned for, then more quantization noise enters the system. If low-pass filtering is needed for the input or output channel of the system, this should be done within the program with digital filtering, instead of simply prolonging the sampling time.

4.4 Bandwidth of Sampled-Data Systems

Choosing a sampling interval and anti-aliasing filter for a control system is discussed by Friedland [72] and in more detail by Åström and Wittenmark [15].

The sampling rate of the system should be as high as possible. The higher the sampling rate, the more information is available to the system. While too much

information can easily be reduced, too little information cannot be enhanced. There is a certain price to pay for a high sampling rate, cost and increased processing effort, but too low a rate is more damaging than too high a rate.

Apart from the fact that the quantization noise can be decreased as mentioned above, there are also other benefits that concern the bandwidth and phase of the computer-controlled system.

The sampling rate is closely related to the bandwidth of the closed-loop system. To capture all signals of interest the sampling rate has to be at least double the highest occurring frequency of interest so as to comply with the Nyquist criterion. However, for a control system the sensing signal bandwidth should be as broad as possible. The broader the sensing bandwidth, the larger the forward gain of the controller can be without risking instability of the control loop. A large controller gain with broad bandwidth ensures accurate and fast control and is hence highly desirable.

The sampling rate also affects the phase of the computer-controlled system, since the sampling interval and the reaction time of the system are closely connected.

Time Delay Sampled-data systems inevitably introduce a time delay or dead time into the feedback loop. The time delay consists of the conversion times of the analog-to-digital converter (t_{adc}), the digital-to-analog converter (t_{dac}) and the program time (t_{program}). The first two times are fixed time intervals. The program time is the time taken to read and write the input and output channels and compute the controller action. The sum of all three times is the dead time (t_{d}) of the system,

$$t_{\text{d}} = t_{\text{adc}} + t_{\text{dac}} + t_{\text{program}} . \tag{4.9}$$

The dead time of the system described in this thesis is mainly caused by the analog-to-digital converter, and is due to the sequential process of successive approximation. This type of converter was used, since apart from having a high conversion speed it also has a large dynamic range. The 16-bit converter has a minimum conversion time of 10 μs. The program time takes less than 1 μs, which is only a fraction of the conversion time of the analog-to-digital converter t_{adc}.

A constant delay causes a linear increase in phase shift with frequency. A closed-loop system using negative feedback with a constant delay (t_{d}) will reach the condition for oscillation at a frequency of not more than

$$f = \frac{1}{2t_{\text{d}}} . \tag{4.10}$$

The negative feedback already provides a phase shift of half a period, so oscillation will occur when the delay is equal to another half period [124]. Note that the possible phase shifts of other loop elements have not been taken into account in this estimate, which means that the closed-loop bandwidth might be lower still.

In conclusion, the shorter the time delay of the system the less phase shift is introduced. Choosing a sampling rate that is much higher than the closed-loop bandwidth of the sampled-data system usually means only a small time delay and an increased phase margin of the sampled-data system at higher frequencies [72].

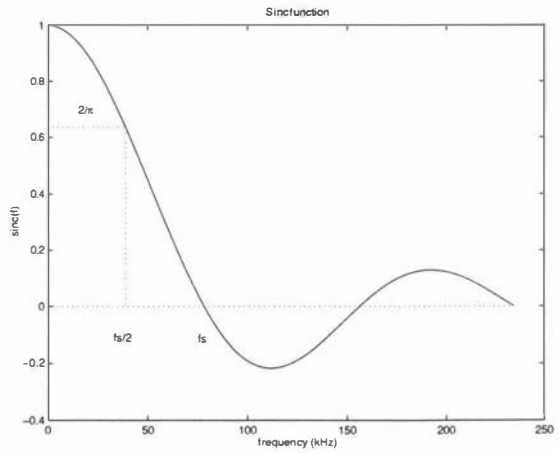


Figure 4.3: **Sinc Function:** The sinc function is shown for a sampling time of $t_s = 12.8\,\mu s$, which is the actual sampling interval used in the current system, discussed in section 7.2. The function is defined as $\mathrm{sinc}(x) = \sin(x)/x$. The sinc function is the Fourier transform of the boxcar function. If the boxcar is t_s wide, then $\mathrm{sinc}(x)(f)$ has its first zero at $f = f_s = 1/t_s$. At half of this value the sinc function is decreased to $2/\pi$ down from 1.

Once the sampling frequency has been chosen, the input to the analog-to-digital converter has to be bandwidth limited with an analog low-pass filter. This anti-aliasing filter must sufficiently suppress signals with frequencies higher than the Nyquist frequency. The anti-aliasing filter will also increase the signal to noise ratio of the system by suppressing high-frequency noise that would otherwise be aliased to low frequencies.

4.4.1 Reconstruction

Although Shannon's theorem states that a signal can be exactly reconstructed if its frequency content is below the Nyquist frequency, the reconstruction method needed is not usually used in a real-time application such as computer control, since it has several drawbacks. Firstly Shannon reconstruction would need future values as well as past values of the error signal [15]. Future values are only available if the reconstruction is delayed by several sample intervals and time delays are undesirable for closed-loop control [15, 110]. Secondly, Shannon reconstruction is complicated and can only be used with fixed sample intervals [15].

A simple and common method of reconstruction is the already mentioned zero order hold. The response of the zero order hold is given by [115]

$$\frac{\sin\left(\pi \frac{\omega}{\omega_s}\right)}{\frac{\omega}{\omega_s}} e^{-j\pi \frac{\omega}{\omega_s}}, \tag{4.11}$$

where ω_s is the sampling frequency. The magnitude response is a sinc function that has its first zero at the sampling frequency. The phase shift is proportional to the frequency and reaches 90° at the Nyquist frequency $\omega = \omega_s/2$. The sinc function is shown in figure 4.3. The sinc function is already decreasing before this frequency, thus showing a low-pass characteristic. To suppress the side lobes

of the sinc function an extra low-pass filter, the reconstruction filter, is usually placed after the zero order hold. The ideal reconstruction filter would have unity gain and no phase shift up to the Nyquist frequency [115]. At frequencies higher than the Nyquist frequency the filter would pass no signal. This filter can be an analog low-pass filter or it can be a combination of a digital and analog filter. The steepness and cut-off frequency of the implemented analog low-pass filter depends strongly on the controlled process and on the chosen sampling rate.

Choosing a sampling rate that is much higher than the closed-loop bandwidth of the sampled-data system will decrease the reconstruction error of the zero order hold [15]. In relation to this there is yet another benefit in a high sampling rate. While the bandwidth of the input channel should be as high as possible, the bandwidth of the output channel only needs to be adequate for the process under control. With a high sampling rate a digital filter can be used for the output channel, which can have the advantages of steeper roll-off beyond the frequency of interest without extra phase shift. The cut-off frequency of the analog reconstruction filter of the output channel can be placed comparatively close to the highest frequency of interest, but the filter may have a gentle roll-off, as long as signals at higher frequencies than the Nyquist frequency are sufficiently suppressed. Such a reconstruction filter can be a simple *RC* filter with the added advantage of having lower noise than an active filter. However, the filter may still be a higher order filter as described by Klaassen [115]. In this case only little phase shift would be introduced up to the Nyquist frequency.

Åström and Wittenmark [15] recommend choosing a sampling rate six to ten times higher than the highest frequency of interest. Friedland [72] suggests an even higher rate of 16 times under adverse circumstances.

The shortest possible time between two incoming data is the conversion time of the analog-to-digital converter (t_{adc}), which is the chosen sampling time for this project. To minimize the program delay time (t_{program}), the analog-to-digital converter should be read at the beginning of the sampling interval and the digital-to-analog converter should be loaded immediately after the controller output has been computed.

If a sampling time of $t_s = 10\,\mu$s without additional delay could be realized, then a closed-loop frequency of

$$f_{\text{cl}} = (\frac{1}{8} \text{ to } \frac{1}{6}) \times \frac{1}{t_s} = 13 \text{ to } 17\,\text{kHz}$$

would be possible using the recommendation of Åström and Wittenmark. The actual closed-loop frequency in this system turned out to be at roughly half that value, as discussed in section 7.2.

The possible bandwidth of this sampled-data system is adequate to control the probe head of this microscope. Additionally, the current-to-voltage converter will only have a limited bandwidth and much lower scanning rates than the highest possible are often used.

Reaction Time Wong and Welland [194] chose a sampling time for their system equal to the sum of the conversion times of the converters, i.e. $t_s = t_{\text{adc}} + t_{\text{dac}}$. This way of choosing the sample time has no clear advantage, as can be seen from the following discussion of the reaction time of the controller.

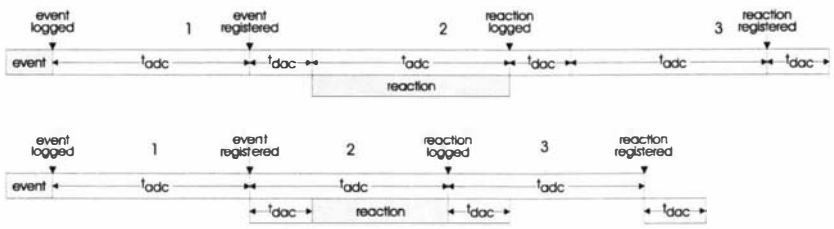


Figure 4.4: **Reaction Time:** Two timing diagrams for programs with different sample time t_s is shown. The analog-to-digital converter needs a certain time t_{adc} to convert a logged event to a datum that the program can register. The time that the program needs to react to the event consists of the conversion time of the digital-to-analog converter t_{dac} and the processing time of the analog electronics, i.e. the time lag of the output channel. In the top diagram, the sample time is equal to the total time the converters take, i.e. $t_s = t_{dac} + t_{adc}$, while in the bottom diagram the sample time is equal to the conversion time of the analog-to-digital converter, i.e. $t_s = t_{adc}$. The program time is taken as negligible in this diagram. As long as the time lag of the output channel is shorter than $t_{adc} - t_{dac}$, the program can react faster to an event if the shorter sample time is chosen.

One might regard the time between seeing an event, reacting to it and then seeing the result of the controller action as the characteristic time of the system, where a short reaction time is better than a long one.

The output rate of the system in closed-loop control is limited to comparatively low frequencies, so that this reaction time is not significant on the involved time scales. However, when the system is used in open loop configuration, such as while approaching the tip to the sample or when recording characteristic curves, the reaction time is important.

Sampling as fast as the analog-to-digital converter allows gives the program the opportunity to react as rapidly as possible. A timing diagram is shown in figure 4.4. Adding the time it takes between registering an event and registering a reaction to the event from the program, it is found that the reaction time is shorter if only t_{adc} is taken as the sample interval. In this case the reaction time is

$$t_{reaction} = 2 \times t_{adc},$$

while in the other case it takes

$$t_{reaction} = 2 \times t_{adc} + 2 \times t_{dac},$$

which is longer. For both cases there is one intermediate sampling interval where the program has already reacted but has not seen its own reaction. If this is not accounted for, the program might react, although it has already done so. This is however a matter of control strategy and can be incorporated into the program. It can only be prevented by additionally waiting a certain fraction of the time lag of the system, i.e. the step rise time.

4.5 Microscope Control

The direct digital control loop implemented in this system can be seen in figure 4.5. The distance between tip and sample is the process that is to be controlled, with the distance between tip and sample being measured by the

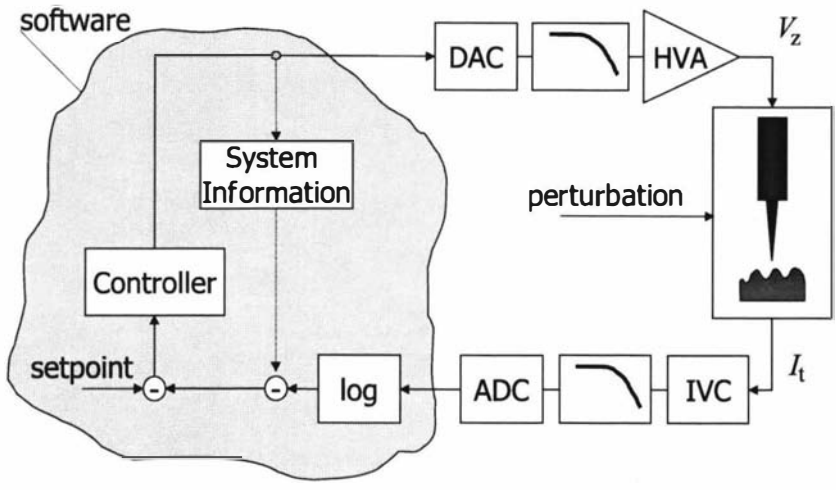


Figure 4.5: **Feedback Control Loop of the STM System:** The process to be controlled is the tip-sample distance, while the core of the controller is a control algorithm residing in DSP software. The process variable is the tunneling current and the controller output is the height-control voltage that is applied to the piezoelectric actuator carrying the tip. The electronics, interfacing between controller and process can be interpreted as part of the controller.

tunneling effect. In this case the tunneling current is the process variable. The elements of the loop starting with the tip and sample are the following:

The tunneling current is measured by a current-to-voltage converter. The output voltage is then low-pass filtered with an anti-aliasing filter. The tunneling current is then digitized and read by the software. The irregular shape on the left of the diagram indicates software residing in the DSP. The input value from the ADC module is compared to a given desired signal level, the setpoint. The controller, a software algorithm, decides what action to take and outputs a control word. The control word gets converted to analog form, is filtered and then amplified. The tip-sample distance is changed by the high-voltage signals fed to the center actuator, which holds the tip.

Overall feedback has to be negative, so that the actions taken by the controller bring the measured signal closer to the setpoint. If the tunneling current is too high, then the input signal is bigger than the desired signal and the controller pulls the tip away from the sample and vice versa.

The forward gain of the loop is set in software. The minimum sampling time is determined by the hardware as discussed in section 4.4. The system can simulate a slower sampling time. The outputs of the analog-to-digital converter are still read with the minimum sampling time but a digital filter is used to perform a time average. This way aliasing is avoided. The possible sampling times will of course be integer multiples of the minimum sampling time.

Low-Pass Filters In the design of the probe electronics active low-pass filters have been used. Overshoot, ringing and other distortions are to be minimized, so that signal shapes are amplified as undistorted as possible. Among the three common filter styles 'Butterworth', 'Bessel' and 'Chebychev', the Bessel filter offers best performance in the time domain with the shortest rise time and no overshoot and ringing [171]. Bessel filters are also commonly used in computer

controlled systems [15]. All low-pass filters employed in the probe electronics are Bessel filters.

Perturbations On the right of figure 4.5 perturbations to the tip and sample are indicated. When the tip moves horizontally in constant-current mode, the control loop treats the changing tip-sample distances like a perturbation. The corrections, which appear on the controller output, are recorded as an image. An important consequence of this is that the system cannot easily distinguish between unwanted perturbations and valid scan signals.

Unwanted effects that change the tip-sample distance are external vibrations and temperature drift. The control loop will react to these perturbations, especially the thermal drift, since it is changing at a low rate. Hence, the task to suppress vibrations and drift cannot be left to the control loop. Instead it should be addressed by the vibration isolation and the thermal compensation design of the probe head.

If the perturbations and the scan signal lie at different frequencies then they can be distinguished and a correction be attempted. This is the basis for thermal drift correction obtained by subtracting a slope and also the idea behind two dimensional post processing of images to filter out $1/f$ noise.

4.6 Imaging Modes

Running the microscope in constant-current mode requires a slow scan speed with high controller gain. With high gain the tip follows the contours of the sample accurately.

Using the constant-height mode of the microscope the tip need not follow the detailed contours of the sample surface. The control loop is open and the signal is extracted from the tunneling current directly. Along with providing a good closed-loop control behaviour this is a second reason for having a high bandwidth for the input and output path of the electronics. The tip can be kept at constant height only over a flat area of the sample, since otherwise the tip might crash. To increase the scan area the tip is often used with low gain closed-loop control. This makes it possible to avoid gross features and also removes them from the image. Whether the low gain needed for this mode has to be set directly, depends on the scan speed. At higher scan speeds the tip is automatically in this mode, since the controller gain decreases with frequency.

Chapter 5

Electronics

5.1 Design Specifications

The electronic hardware[1], which includes the DSP and the custom-made electronics, is a major part of the microscope system. The role the electronics plays in the microscopy system is discussed in section 2.3.2 and is shown, in the form of an overview diagram, in figure 5.1. The DSP is a dedicated microprocessor, which has been chosen as the main part of the electronic hardware. This is an important feature of this instrument in comparison to many other custom-made microscopes. The DSP communicates with the PC, reads the tunneling current from the microscope, controls the tip movement and moves the microscope head. The DSP exchanges data with either the PC or the probe electronics via the DSP system interface.

The interface to the microscope operator is a software program running on the PC. The image data are uploaded from the DSP via an 8-bit parallel link between a parallel data register on the DSP side and an input/output card on the PC side.

5.1.1 Speed

In the analog STM the response time of the closed-loop control is dependent on the speed of the analog electronics. By comparison, the DSP-controlled STM has a response time limited by the speed of the digital electronics and the DSP software. The sampling rate of the digital system and the bandwidth of the analog part of the system are discussed in section 4.4.

5.1.2 Noise, Resolution and Dynamic Range Considerations

To scan a sample surface with very high resolution one needs a stable tip-sample distance and bias voltage, as well as the ability to measure the tunneling current to some minimum resolution. A certain amount of noise is inevitable. Shot noise and tunneling junction noise are always part of the tunneling current and impose

[1] Electronics and computer control is important in scientific instrumentation in general. Many textbooks covering the subject are available [42, 53, 80, 115, 161].

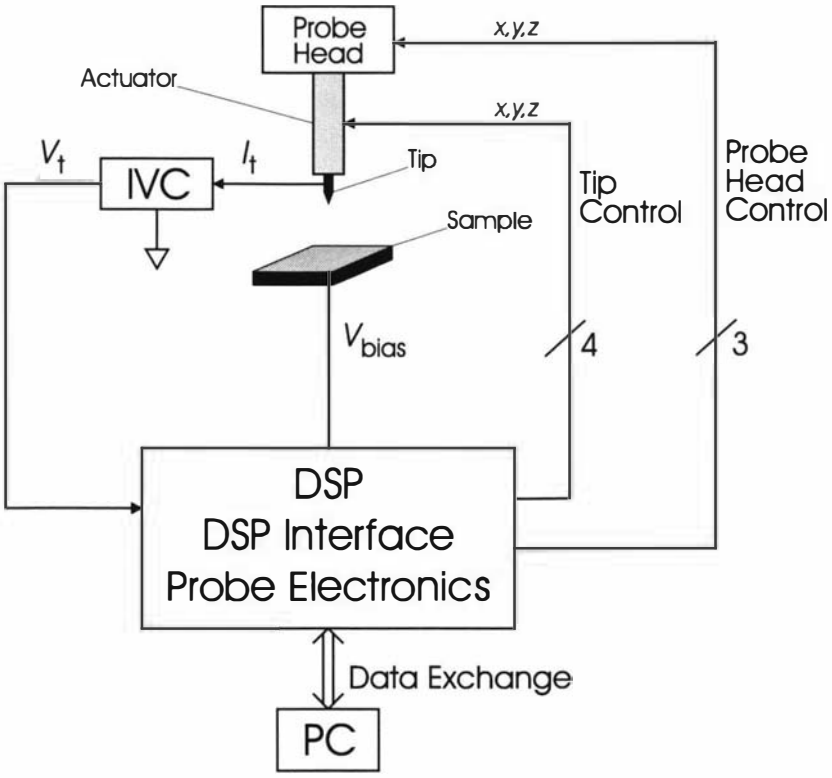


Figure 5.1: **System Overview:** The electronics is in the center of the microscopy system. It controls the vertical and horizontal position of the tip and of the probe head. The bias voltage (V_{bias}) is adjusted by the electronics and the tunneling current (I_t) measured after it is converted to a voltage (V_t) by a current-to-voltage converter (IVC).

a lower limit on the possible resolution of the microscope [118]. The tunneling equation (2.6) can be used as a guide to estimate the overall possible resolution of the tunneling current due to inevitable noise and other fluctuations. The general design considerations discussed in section 2.2 are also used.

5.1.3 Tunneling Current

The tunneling current is the input signal to the microscopy system. The tunneling current varies because of changes in the tip-sample separation (s) and bias voltage (V_{bias}) and also because of inevitable noise such as shot noise and noise from electronic components. Although s also changes with thermal drift and vibrations, this fact is ignored in this section. The relationship between tunneling current, bias voltage and tip-sample separation, is given by the tunneling equation:

$$I_t = \Lambda V_{\text{bias}} e^{-\frac{s}{s_0}}$$

originally given in section 2.1.1.

As part of the design criteria a tunneling current resolution of 10 pA was chosen. Therefore the noise produced by the electronic components in the signal path had to be much smaller than that. It is desirable to keep the digitization noise about ten times below other fluctuations in the analog part of the input path. A resolution of 1 pA was therefore aimed at for the digital step size plus component noise. With the help of the tunneling equation and assuming uncorrelated fluctuations, one obtains the following expression for the noise fraction in the tunneling current:

$$\frac{\Delta I_t}{I_t} = \left[\left(\frac{\Delta I_{\text{int}}}{I_t} \right)^2 + \left(\frac{\Delta I_{\text{el}}}{I_t} \right)^2 + \left(\frac{\Delta s}{s_0} \right)^2 + \left(\frac{\Delta V_{\text{bias}}}{V_{\text{bias}}} \right)^2 \right]^{\frac{1}{2}}, \qquad (5.1)$$

where I_t is the tunneling current, V_{bias} the bias voltage, s the tip-sample distance, ΔI_{int} the intrinsic fluctuations of the tunneling current and ΔI_{el} the noise produced by the electronic components. The fluctuations of the bias voltage are ΔV_{bias} and the magnitude of fluctuations in the tip-sample distance is Δs. Typical values used in equation (5.1) are 1 nA for the tunneling current, 50 mV for the bias voltage and 0.1 nm for the constant s_0, as derived in section 2.1.1.

To get an order of magnitude value for $\frac{\Delta I_t}{I_t}$, we first assume that intrinsic fluctuations in I_t are only caused by shot noise, which cannot be suppressed and in the best case will be the dominating noise source. Shot noise current density is given by

$$i_n = \sqrt{2eI_{\text{dc}}}$$

Here, i_n is the noise current density, e is the electron charge and I_{dc} the dc component of current under consideration. With a typical value of 1 nA for the tunneling current, one obtains

$$i_n = \sqrt{2 \times 1.6 \times 10^{-19}\,\text{C} \times 1\,\text{nA}} = 0.02\,\text{pA}/\sqrt{\text{Hz}}.$$

The noise current within a closed-loop bandwidth of $f = 10$ kHz is therefore

$$\Delta I_{\text{int}} = i_n \sqrt{f} = 0.02\,\text{pA}/\sqrt{\text{Hz}} \times 100\sqrt{\text{Hz}} = 2\,\text{pA}.$$

The relative contribution of shot noise to the tunneling current is then

$$\frac{\Delta I_{\text{int}}}{I_t} = 2 \times 10^{-3}.$$

The value of 2×10^{-3} is used below as a maximum noise criterion for the tip-sample separation and the bias voltage. The acceptable noise level of all parts of the electronic hardware should stay below the level given by the maximum noise criterion. A value of

$$\Delta I_{\text{el}} = 1 \times 10^{-3}\,\text{nA} = 1\,\text{pA}$$

has been envisaged as an achievable goal. For the electronic components of the input signal path that means a desired signal-to-noise ratio of

$$\frac{1\,\text{nA}}{1\,\text{pA}} = 1000 \equiv 60\,\text{dB}$$

5.1.4 Piezoelectric Control Voltages

Resolution

The range and resolution of the control voltages for the piezoelectric actuator tubes depend on the sensitivities of the actuators and the desired horizontal and vertical range and resolution of the microscope. This was discussed in section 2.2, where a desired vertical resolution of 1 pm was suggested.

To find the influence of a possible fluctuation of 1 pm on the tunneling current it has to be compared with s_0. As a fraction of s_0 this is:

$$\frac{\Delta s}{s_0} = \frac{1\,\text{pm}}{100\,\text{pm}} = 10 \times 10^{-3} \equiv 40\,\text{dB}, \tag{5.2}$$

which is 5 times larger than the maximum noise criterion.

The vertical sensitivity of the actuator is $r_z = -3.4\,^{\text{nm}}/\text{V}$, while the horizontal sensitivity is $r_h = -27\,^{\text{nm}}/\text{V}$. With these sensitivities the desired resolution values of 10 pm horizontally and 1 pm vertically translate to piezoelectric actuator voltages of 400 μV and 300 μV respectively. The demands on the electronic hardware are therefore to keep the noise level at or below 300 μV.

Range

The type and size of the piezoelectric actuator and the available voltages determine the possible scan area. A large scan area is desirable, but the applied voltages need to be manageable. There will always be a compromise between scan area and speed of data acquisition due to the size of the piezoelectric actuator. Keeping in mind the demands on the control voltage noise levels, a maximum voltage of just $\pm$190V has been chosen as a reasonable design goal. These voltages allow the scanning tip to be moved up to $\pm 190\,\text{V} \times 3.4\,^{\text{nm}}/\text{V} \approx \pm 0.7\,\mu$m vertically and up to $\pm 190\,\text{V} \times 27^{\text{nm}}/\text{V} \approx \pm 5\,\mu$m horizontally. It is important to bear in mind that these distances define the end points of movement. For instance, when the actuator is fully extended then the maximum voltage is applied to all electrodes and no horizontal movement is possible. The same is true for the horizontal extreme point, where no vertical movement is possible any

more. To allow room for movement in all directions a rule of thumb could be to limit the horizontal excursions to half the possible range. Note that a large vertical range is important since samples are often slightly tilted. The scan area would then be $5\,\mu\text{m} \times 5\,\mu\text{m}$, which is a relatively large scan area compared to other commercial and custom-made scanning tunneling microscopes. The availability of high voltages simplifies the tip-sample approach procedure, since the fine approach has a greater range. The required dynamic range for the high voltage signals is

$$\frac{\pm 190\,\text{V}}{300\,\mu\text{V}} = 1.27 \times 10^6 = 122\,\text{dB}. \tag{5.3}$$

Achieving such as range would be a tall order.

Considering the range of movement for the outer piezoelectric tubes of the microscope, only the horizontal sensitivity is relevant. Since the outer tubes have different dimensions and are made from a different material their horizontal sensitivity is different:

$$r_h = (-18.9 \pm 5.7)\,^{\text{nm}}/_{\text{V}}.$$

This was deduced in appendix A.4 as equation (A.14). For horizontal movement the maximum step size should therefore be $18.9\,^{\text{nm}}/_{\text{V}} \times 190\,\text{V} \approx 3.6\,\mu\text{m}$, which is more than sufficient for the movement of the probe head, since usually many small steps are taken rather than one large step.

Bandwidth

The bandwidth of the output channel was discussed in section 4.4. The bandwidth for the piezoelectric control voltages needs to be high enough so as to allow production of the pulse-shaped signals necessary for the stick-slip movement. Approximating the output smoothing low-pass filter as a simple *RC* filter, the rate of voltage change at the output is proportional to the desired output voltage level as well as the bandwidth. The higher the bandwidth, the easier the outer actuators will be able to slip during the stick-slip movement. The choice of bandwidth has to be a compromise between this and the amount of digital and other noise that finds its way into the output. A maximum bandwidth of 100 kHz has been envisaged, which corresponds to the sampling frequency.

5.1.5 Bias Voltage

Resolution

The maximum noise criterion determines the smallest usable bias voltage. A bias voltage V_{bias}, between tip and sample of less than 1 V is needed if the microscope is to be used as a non-destructive imaging device. An achievable noise level for the electronic components is $\Delta V_{\text{bias}} = 100\,\mu\text{V}$. Together with the maximum noise criterion this limits the bias voltage to a minimum level of

$$V_{\text{bias}}(\text{min}) = \frac{100\,\mu\text{V}}{2 \times 10^{-3}} = 50\,\text{mV}$$

Using smaller bias voltages is of course possible, but leads to decreased signal-to-noise ratio in the tunneling current. Note that for a topographical scan the digital step size of V_{bias} does not have to be as small as the acceptable noise level of V_{bias}. In a topographical scan, the bias voltage is held constant, and

hence no digitization noise is introduced. However, if the digital step size is kept at the same level as the acceptable noise, the possible voltage range would be determined by the number of available bits. For a 12-bit digital-to-analog converter for instance, the range would be $\Delta V_{\text{bias}} \times 4096 = 0.4\,\text{V}$. This could be desirable if the microscope is to be used as a spectroscopic instrument with high resolution. In this case one should consider the option of superimposing signals with coarse and fine step sizes. In the current design such high resolution was not considered to be of prime importance and as mentioned below higher maximum bias voltages were required. A single converter has therefore been used with a larger step size. A bias voltage step size of a few millivolts is sufficient to provide enough data points for a characteristic I-V curve.

Bandwidth

The bias voltage should be variable to enable the measurement of the current-voltage characteristics of the sample, i.e. $I_t = I_t(V_{\text{bias}})$. Since this microscope was not primarily designed for spectroscopy, it was decided to restrict the bias voltage bandwidth to reduce the overall noise in the system. The same bandwidth as that provided for the closed-loop control circuit, i.e. 10 kHz, was considered sufficient.

Range

When choosing the magnitude of the bias voltage, it is necessary to consider the work function of the tip and the sample. V_{bias} is usually less than 1 V to fulfill the assumption that the bias voltage is small with respect to the work functions of the tip and the sample.

To allow other modes of operation, where a higher bias voltage might be advantageous, it was decided to provide bias voltages in the range of $-10\,\text{V}$ to $+10\,\text{V}$. A typical mode of operation would be 'nanolithography', where the sample topography is actively manipulated by the tip. Higher electric fields between the tip and sample are necessary for this purpose [84, 159] and the assumption of small voltages is replaced by $eV_{\text{bias}} \approx \Phi$. Instead of imaging, the microscope will then be in Fowler-Nordheim tunneling mode. This mode is accompanied by higher tunneling currents.

5.1.6 Input Channel

Range

Typical values for the tunneling current, I_t, are around 1 nA. Since the system is also designed to accommodate higher, non-imaging currents, the highest expected input current is anticipated to be 10 nA.

Resolution

If the relative contributions from s, V_{bias} and the noise contribution of electronic components (ΔI_{el}) are taken into account then from equation (5.1) the overall relative variation in the tunneling current $\Delta I_t / I_t$ is about

$$\frac{\Delta I_t}{I_t} = \left[2^2 + 1^2 + 10^2 + 2^2\right]^{\frac{1}{2}} \times 10^{-3} = 10 \times 10^{-3} \tag{5.4}$$

The expected signal-to-noise ratio is therefore about 40 dB at a tunneling current of 1 nA. In the estimate above, the electronic noise of the input channel was taken as 1 pA. However, due to the greater uncertainty in the tip-sample distance, such a stringent condition is not needed. The noise could be as high as 10 pA and the signal-to-noise ratio would only be degraded by 3 dB to 37 dB.

5.1.7 Data Transfer

In a given time interval, the number of useful image data points acquired and ready to be stored depends on the communication between PC and DSP and on how much data the DSP can store on-board. Since all data acquired by the DSP have to be sent to the PC to be stored, the overall input data acquisition rate has to match the overall data transfer rate from the DSP to the PC. The available DSP on-board memory can serve as a storage buffer while acquiring new input data. The buffered data can then be transferred in packets to the PC. The design goal was to be able to transfer data from the DSP to the PC at a rate that does not compromise the use of the microscope. One might regard a total transfer time per image of order 1 s to be acceptable.

Transfer Rate

To see how long the data transfer might take, one has to look at the amount of data in one image (N_{data}) and the speed of the transfer process.

Since the DSP is a floating point processor it is appropriate to represent the image data in floating point format. For design purposes assume that an image from one scan of the microscope consists of $N_{\text{data}} = 128 \times 128$ data points, each of which is stored as a word in floating point format with a depth of 32 bits. This comes to a total of 16 kilowords per image. Since the DSP system has on-board data memory of 32 kilowords, at least one such image can be kept in memory. Under the assumption that only one byte at a time is transferred, one needs a transfer rate of greater than $(4 \times 16.4) \times 10^3$ bytes/s $= 66 \times 10^3$ bytes/s to have a transfer time of less than one second. Now the transfer rate of the input/output module in the PC is, at its very maximum, 10^6 bytes/s. As a result of the above discussion, one has to make sure that the DSP data link hardware and the software does not slow down the input/output module by more than a factor of 16 for floating point data.

Acquisition Time

The minimum acquisition time for an image, will occur when the tip is moving at a horizontal speed which still allows the control loop to follow the surface contours on an atomic scale. In other words, the frequency of the changing tunneling current should remain within the bandwidth of the control loop of about 10 kHz. Horizontal scan speed and different scan algorithms are discussed in greater detail in section 7.4.1. Assuming that each line is scanned backwards and forwards and that the horizontal step size is 0.1 nm, the maximum speed is

$$v_{x,\max} = \frac{0.1\,\text{nm}}{100\,\mu\text{s}} = 1\,\mu\text{m/s}.$$

It therefore takes 50 ms to scan a 256×0.1 nm long line backwards and forwards. The total image acquisition time will be $256 \times 50\,\text{ms} \approx 13\,\text{s}$. A transfer time of

about one second is considerably less, which means that the data transfer has little effect on the overall image acquisition time.

This result represents a minimum acquisition time since in section 7.4.1 a smaller bandwidth for the closed-loop control is found. Also, only the minimum step size of the tip across the surface was considered in this assumption. For a smaller control bandwidth and for more extended scan areas one will arrive at larger times for total image acquisition. To scan a 1 μm $\times$ 1 μm area for instance will take approximately 8 min even with a 10 kHz bandwidth using $v_{x,max}$ from above.

5.1.8 Accuracy

It is always desirable to minimize unwanted signals, such as electronic noise and interference, in a measurement system. The most important requirement on the electronics in this system is not the absolute accuracy of the signal values, but rather the need to keep the relative signal fluctuations below a certain level. This is because, without calibration, imaging is a relative measurement process. An image of the sample is formed by comparing tunneling current values at different positions of the sample surface. Component matching, and the tracking of the component's temperature coefficients is more important than absolute accuracy.

The design philosophy which has been implemented, was to use as few components as possible, choose high-precision, low-noise electrical components, and to carefully lay out the printed circuit boards. Digital and analog circuits were kept separate as far as possible, by using separate grounds and supplies and by physical separation on the boards. Every signal conditioning stage is bandwidth limited to suppress high-frequency noise. This is most important for those stages which follow the low-pass filter. When limiting the bandwidth one has to be careful not to introduce too much phase shift at useful signal frequencies, i.e. the cut-off frequency of the filters should not be too low.

Choosing only a few components of high precision minimizes errors in gains and offsets. This is the best way to achieve a low level of unwanted fluctuations and it also means that gain and offset need only slight adjustment or none at all. There is also the option to adjust the gain and offset in software within the resolution of the digital-to-analog converters and the analog-to-digital converter. An exception to the above statements is the logarithmic converter, which needs an accurate offset trim at its inputs to ensure logarithmic conformity. A lot of effort could have been saved by implementing the logarithmic function in software.

Protection circuits are only installed for very expensive components. For the remainder, the security of protection has been disregarded in favor of fewer components and therefore less noise.

Temperature Design of the Electronics

During the discussion of the probe head temperature design in section 3.4 a value for the vertical mechanical temperature drift of $D_{mech} = +34\,^{nm}/_{K}$ was found for a graphite sample. The change in the tip-sample distance due to a temperature drift of the electronics is negligible by comparison, as is shown below.

The contribution to the temperature coefficient by the few, high precision integrated circuits included in the analog circuitry, is below or around 1 ppm/K. Even the contribution of the high-voltage amplifier PA85 to the temperature coefficient is negligible, although the PA85 is a FET amplifier and has a strong temperature dependence in the bias current. However, the actual value of the bias current is so small, that the temperature drift is below 1 ppm/K.

A much bigger gain drift can arise due to the temperature drift of the resistors that determine the gain of the operational amplifiers used in the probe electronics. Only a mismatch in the temperature coefficients of these resistance values[2] gives rise to a gain drift. If the temperature coefficients were matched within 5%, as the resistance values are, then the gain drift using 5% carbon resistors would be $0.05 \times 800\,\text{ppm/K} = 40\,\text{ppm/K}$. The most critical design of all modules is the one for the HV modules. To reduce the amount of $1/f$ noise and to decrease the thermal drift, 1% metal film resistors were used for the HV modules. This reduces the calculated gain drift to $0.01 \times 100\,\text{ppm/K} = 1\,\text{ppm/K}$.

Taking the worst case drift of 40 ppm/K together with the highest possible voltage of 200 V one gets a maximum voltage drift of

$$40\,\text{ppm/K} \times 200\,\text{V} = 8 \times 10^{-3}\,\text{V/K}.$$

Now multiplying by the vertical sensitivity of the piezoelectric actuator one calculates a resulting drift of

$$D_{\text{el}} = 8 \times 10^{-3}\,\text{V/K} \times 3.4\,\text{nm/V} = 0.03\,\text{nm/K},$$

which is negligible when compared with D_{mech}.

Although minimizing the vertical temperature drift is necessary, so as to prevent the microscope probe from impacting on the sample, the horizontal drift is also important, since horizontal stability ensures undistorted images. The drift due to the electronics might have a greater role to play in this case, because the mechanical design is inherently symmetric and only the mismatch in temperature coefficients of the different segments of the piezoelectric ceramic tube contribute to horizontal drift. Since the material can be assumed to be fairly homogeneous, the horizontal drift due to mechanical origin should be very small.

The horizontal drift due to the electronics is obtained by multiplying the voltage drift by the horizontal sensitivity of the piezoelectric actuator, giving:

$$D_{\text{el}} = 8 \times 10^{-3}\,\text{V/K} \times 27.0\,\text{nm/V} = 0.22\,\text{nm/K}.$$

Using metal film resistors in the HV module design the horizontal drift is found to be only a small fraction of this value

$$D_{\text{el}} = \frac{1}{40} 0.22\,\text{nm/K} = 6\,\text{pm/K},$$

which is negligible.

Even though the thermal drift due to the electronics is very small, it is still important to keep the room temperature as stable as possible to minimize thermal drift from the mechanical components.

[2] Here, the temperature coefficient of a carbon resistor is taken as 800 ppm/K and the coefficient of a metal resistor is taken as 100 ppm/K (values from a supplier's catalog [165]).

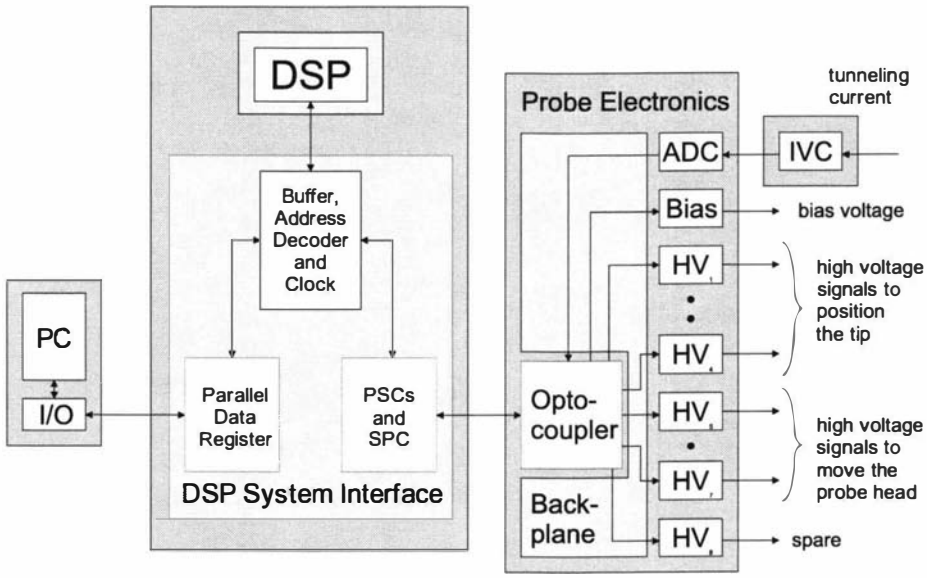


Figure 5.2: **Electronic Hardware:** An overview of all the electronic hardware is shown, including PC, DSP, DSP System Interface and Probe Electronics. The details are explained in sections 5.3 and 5.4.

5.2 Physical Organization of the Electronic Hardware

A more detailed view of the electronic hardware block given in figure 5.1 is shown in figure 5.2. Apart from the DSP there are two other building blocks: the DSP system interface and the probe electronics. The DSP System Interface is discussed in section 5.3 and the probe electronics in section 5.4.

5.2.1 Modularity

To extend the flexibility and versatility of the microscope much of the electronics has been designed in a modular fashion. This brings about the advantages that

- a module can be removed from the system easily and with minimal impact on the rest of the system,
- backup modules can be made, so as to rapidly replace any module that has been damaged,
- a design for one module can be used for multiple modules and
- a modular system can be extended relatively easily.

5.2.2 Organization

The DSP is part of a commercial piece of electronics, which is located on one board. A custom-made DSP system interface is accommodated on another board, which is directly connected to the DSP board. Both boards share the same metal enclosure equipped with one dedicated power supply.

The parallel data register on the PC side is part of the interface between the DSP and PC. The register is mounted on a separate board, which is connected to the input/output card. Both boards reside within the PC enclosure. The interface connection is made with a ribbon cable.

The probe electronics is accommodated in a separate enclosure with its own power supplies. The connection between the probe electronics and the DSP system interface is made with a ribbon cable. The connections between the probe electronics and the probe head are made with coaxial cables. The output signal cables connect to the supporting frame of the probe head. The input signal cable connects to the current-to-voltage converter, which is positioned as close as possible to the microscope probe head. The current-to-voltage converter is installed within a separate metal casing to afford maximum electromagnetic shielding.

5.3 DSP System

5.3.1 DSP

The DSP used in this project is an ADSP-21020 from Analog Devices [10]. It is a 32-bit floating-point processor with single-cycle execution and a cycle time of 30 ns. The ADSP-21020 employs Harvard architecture, which means that data and program memory are separate. This has the advantage that both memories can be accessed simultaneously. The ADSP-21020 has an on-chip instruction cache, so that within one program cycle the DSP can fetch operands from data memory, program memory and the instruction cache. The ADSP-21020 is equipped with a programmable timer, which is used to time sampling intervals for the control algorithm.

According to the manufacturer, the ADSP-21020 was designed with high-level programming languages in mind, specifically ANSI standard C. The manufacturer supplied a C cross-compiler, which made it possible to develop high level language code on the PC for the DSP. High-level language programming is convenient. The code is easy to read and can be modified and re-compiled quickly. One feature of high-level language support is the possibility of using floating point format values. This makes controller computations easy and precise. Scaling of the input and output values is not necessary as would be the case with integer processors [45]. The floating point format offers a very wide dynamic range, which is an advantage when data are accumulated to increase the signal to noise ratio [10]. For floating point values the ADSP-21020 uses the 32-bit IEEE single precision format. This format has an 8-bit exponent, a sign bit and effectively 24 bits of mantissa [10]. A high precision of 24 bits avoids limit cycles[3] in the closed control loop [15]. Additionally an extended version having 8 more bits in the mantissa and 40 bits in total can be used. The ADSP-21020 also supports 32-bit fixed-point formats, such as integers. Kernighan and Ritchie [113] define several types of integers that are used in ANSI standard C. On the ADSP-21020 all these integer types are represented with the same number of bits, 32. More information about the ADSP-21020 can be found in the data sheet [9] and User's Manual [10].

The DSP is the basis of the commercial Ez-Lab evaluation board from Analog Devices. The evaluation board, which has been used in this project, is supported by 32 kilowords of program memory and 32 kilowords of data memory. The Ez-Lab evaluation board employs a second DSP, an ADSP-2111, to handle the downloading of compiled program code and uploading of data from selected memory blocks of the ADSP-21020. The evaluation board is described in manual supplied with the evaluation board [7].

5.3.2 DSP System Interface

The system interface allows the DSP to communicate with both the probe electronics and the PC. The role the system interface plays within the digital electronics module is illustrated in figure 5.2. This section describes the DSP system

[3] In the broadest sense, limit cycles are periodic motions that can occur in non-linear systems [17]. One of the simplest limit cycle would be the temperature cycling of a refrigerator due to its on/off control. Limit cycles can also occur in digital control systems, if the resolution is not high enough [15].

interface, which links the DSP, the PC and the probe electronics. Before the finally implemented design is described, other options for the DSP to probe electronics interface will be discussed. The design of the data link between the DSP and PC is then described in the following section.

Design of the Interface between the DSP and Probe Electronics

The interface between the digital and analog domains of the system includes several digital-to-analog converters and one analog-to-digital converter. Early in the design, the decision had to be made as to how to interface the DSP to these converters. There are several different options available for connecting a processor to multiple peripheral devices. Many of these are discussed in the ADSP-21020 User's Manual [10]. The three main options are listed below and are described in more detail. There are two directions of data transmission, the input channel reading from the analog-to-digital converter and the output channel writing to the digital-to-analog converter. For the sake of clarity, the options are discussed for the output channel first. The options are

1. using a single serial data channel and a multiplexer,
2. using parallel-to-serial converters connecting to several serial data channels or,
3. using a common parallel data channel.

All three options leave the question of the digital interface of the digital-to-analog converter open. Both serial-data input and parallel-data input digital-to-analog converters can in principle be used. In option 2 for example, a parallel input digital-to-analog converter can be used in conjunction with a serial-to-parallel converter at the far end. However, it seems more straightforward to use serial data input digital-to-analog converters for option 1 and 2 and a parallel input data digital-to-analog converter for option 3.

The input channel options are equivalent to the above listed output channel options with reversed data transmission direction.

Objectives Several goals were set during the design of the DSP-probe electronics interface: to provide reasonable speed; to minimize the digital noise coupling into the analog part of the electronics; and to find a balance between the effort of building hardware components and coding software. High speed and low noise were crucial if this instrument was to be on equal footing with commercial equipment. Realizing most functions of the instrument in software with a minimum amount of hardware, ensures greatest flexibility and versatility. Following is a list of the objectives with a discussion on how well the different options fulfill these.

- **Speed:** Transmitting data in parallel is faster than transmitting them in serial if the same clock speed is used. Therefore using a parallel data channel would be the fastest way to transmit data to devices in the probe electronics while using the serial channel with multiplexer option would be the slowest. However the speed at which data are collected is a limiting factor and so the input data transmission speed need not be faster than the conversion speed of the analog-to-digital converter. Likewise the output

channel need be no faster than the digital-to-analog converter it is talking to. In fact, if a fast clock is used and only some channels are needed at the same time, then the necessary data can be transmitted serially during the conversion time of the analog-to-digital converter.

- **Digital Noise:** To minimize the coupled digital noise, it was decided to separate the digital and analog grounds by optocoupling. If only a small number of digital lines are used, both the amount of digital noise and the number of optocouplers required is reduced. When estimating how much digital noise is coupled into analog signal lines, it is important to consider the amount of activity on the data lines and the timing of the transmission. The acquisition time span of the analog-to-digital converter should be the 'quietest' time, i.e. the time of least digital noise. Since the parallel option has the most digital lines, it produces the highest amount of digital noise. The second option is better in that respect and the first option is the best. A small number of digital lines also requires the smallest number of optocouplers.

- **Hardware Effort:** Each option needs some means of decoding the address of the digital-to-analog converter or analog-to-digital converter in question. All lines connecting the DSP interface with the probe electronics have to be buffered, which is easier for fewer lines. The two options that transmit serial data need a clock. Most effort has to be put into option 2. The other options need less hardware. The option with a single serial channel and a multiplexer works best with a DSP that has a serial output. If a serial output is not available a single parallel-to-serial converter can be used.

- **Software Effort:** Minimum effort is needed in software when the different peripheral devices can be treated like memory locations, as is the case for the two options that do not transmit serial data directly (2 and 3). When the peripheral devices cannot be treated like memory locations, the software has to generate specific load commands for the different devices.

In the end option 2 was chosen, because it minimizes the amount of digitally coupled noise while still providing a reasonable transmission speed. Using a multiplexer would slow down the overall transmission rate to an unacceptable level. Using parallel data transmission would create the most digital noise and would also need the most optocoupling lines. For the chosen option all converters need to have a serial interface, while the DSP inputs and outputs parallel data. This means that parallel-to-serial converters have to be built for the output channel and a serial-to-parallel converter has to be built for the input channel. The extra effort in building parallel-to-serial converters has been considered acceptable given the gains of fast transmission and low noise. The highest speed parallel data transmission offers no considerable advantage, since the DSP needs time to compute the control values for the tip movement. During this time the parallel-to-serial converter option is fast enough to exchange all necessary data. More importantly, there are not many high-resolution digital-to-analog converters and analog-to-digital converters which are faster than those used in this project. Option 2 also needs the least effort in terms of software coding.

Clock: One design decision which had to be made, was whether to use the DSP clock for the DSP interface, or a separate local clock. Both options are similar with regard to effort in designing hardware and timing signals; the DSP clock has to be buffered on the interface, but a separate clock module has to be integrated into the interface. Since the DSP clock is fast when compared to the speed of the digital-to-analog converters, the analog-to-digital converter and the PC input/output card, it has to be divided down to give a multiple of its original period. Employing a local clock renders more control over the clock period, although it also has to be divided into several available clock signals for the different speed requirements of the different subunits.

Both options are feasible. In this design a local clock has been implemented, which is called the system clock.

Frequency of the System Clock Let T_{clk} be the cycle time of the system clock residing in the interface. In the specifications it was decided to convert data between the analog and digital domains within a time span of 10 μs. The maximum clock period that can be used is given by the maximum number of bits and this time span. An n-bit serial to parallel converter needs $n \times T_{\text{clk}}$ seconds to convert n bits of serial data. The clock period should not be more than 500 ns, so that a 20-bit conversion can be completed in the chosen time span of

$$20 \times 500\,\text{ns} = 10\,\mu\text{s}.$$

The minimum clock period on the other hand is related to the speed of the converter logic and other logic used in the serial data transfer. The converter logic has typical switching times of less than 50 ns; the minimum clock period that can be used with all converters is 100 ns. The closer the clock period is to this minimum value, the more digital noise is generated and the more likely it is that there will be logic timing problems. It is usually advantageous to use the minimum logic speed possible since this results in less stringent buffering needs, less noise and easier timing. However, if a clock period close to the maximum value of 500 ns is chosen, then the longer conversion time of the digital-to-analog converters would degrade the closed-loop bandwidth of the system.

As a compromise between these extremes a value of 200 ns for T_{clk} was chosen, so that the serial data download time for a 20-bit digital-to-analog converter would be $20 \times 200\,\text{ns} = 4\,\mu\text{s}$. The clock for the converters is derived by division from the system clock, which has a period of 100 ns. The 12-bit analog-to-digital converter, used to generate the bias voltage, needs less time to convert data and the time needed for operations other than conversion is small when compared with the conversion time span. All conversions happen essentially in parallel. The analog-to-digital converter uses an internal clock with a period of 440 ns.

Reset Concept All output voltages should be at ground potential once the electronics is switched on. A system reset in the middle of an operation should also revert all outputs to ground. Similarly, all digital systems should be found in a definite state after power has been supplied to the electronics or after a system reset has occurred.

The 20-bit digital-to-analog converters used in this project do not feature a reset input. Hence one has to explicitly load zero to reset these converters. For

the corresponding 20-bit parallel-to-serial converters, this means that the ongoing processes have to be terminated and a zero value subsequently transmitted. It seemed easiest to leave it to the software to load the digital-to-analog converters with zero, while the hardware reset brings the DSP and its extensions into a definite state, ready for the next software command.

Usually the 20-bit digital-to-analog converters do output zero volts after they are switched on. This is true when the DSP is without power. Only rarely after power is supplied, does the DSP or its extensions output unwanted data to the digital-to-analog converter. A way to ensure that the high-voltage outputs remain at ground level, is to switch on the DSP System first, set the parallel-to-serial converters to zero in software and then switch on the probe electronics.

Design of the Interface between DSP and PC

A feature of this microscope system is the use of a DSP to complete two tasks simultaneously: controlling the microscope tip and exchanging data with the PC. To offer good tip control, especially to prevent tip-crash, the data exchange has to occur rapidly. In addition, the amount of time the DSP has to spend up loading or downloading data should also be minimized. A direct connection to the PC, supported by line drivers, is therefore ruled out in favor of a buffered connection through data registers. Writing to a memory-mapped register does not take up much DSP processing time. The data transfer from the registers to the PC is managed by handshaking signals, which can be generated outside the DSP. The communication could even be handled by another dedicated microprocessor, which would provide a certain flexibility. However a simpler option is to produce handshaking signals in hardware.

The communication channel has to achieve a transfer rate of the order of 10^5 bytes/s. The connection should therefore be a parallel data link and not a serial one like the RS232. Note that this is a parallel link between DSP and PC. The decision to use serial links between DSP and the probe electronics is not affected.

To handle communications on the PC side, a commercial input/output unit was chosen, which is based on the 8255A programmable peripheral interface from Intel. The 8255A is a widely used component, which offers reasonable speed, is inexpensive and includes handshaking. The 8255A can be configured as a one 8-bit wide bidirectional bus or two 8-bit wide unidirectional communication links. The latter option was chosen here.

The data between the DSP and PC is buffered by a temporary data storage register. Although a FIFO register[4] would provide the most rapid data transfer rates, for cost reasons this was not an option. The storage register, which can only hold one datum at a time, also requires simpler support logic than a FIFO register.

The handshaking signals, required when communicating with the 8255A, are generated by dedicated hardware placed close to the data registers. The length of the handshaking signals is tailored to the requirements of the 8255A, so that a data transfer rate close to the maximum can be achieved.

[4] A common way of improving data transfer rates is to use a first-in first-out (FIFO) register. The deeper the FIFO register, the lower the chances that the faster processor (i.e. the DSP) will be held up by the slower processor (the PC in this case).

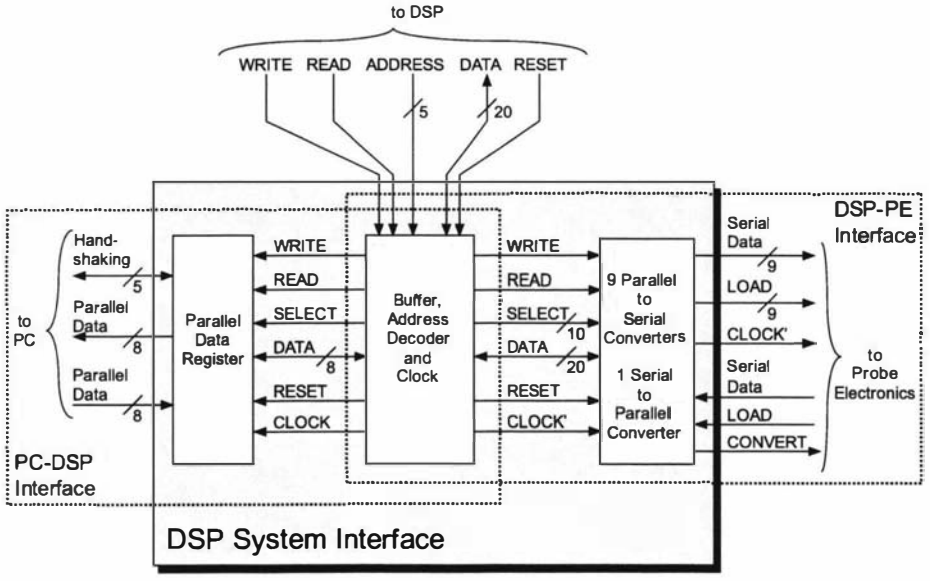


Figure 5.3: **Block diagram of the DSP system interface:** Each subunit (e.g. the serial-to-parallel converter) can be selected by one of the signals from the address decoders. The subunits are connected to the DSP data bus, to the respective read or write strobe and the reset signal from the DSP. All subunits are supplied with a clock signal.

5.3.3 Overview of the DSP System Interface

The system interface is made up from a parallel-data register, nine parallel-to-serial converters, one serial-to-parallel converter and some additional logic, including a clock, an address decoder and buffers. An overview is given in figure 5.3. To exchange data with the converters in the probe electronics, the DSP can write to the parallel-to-serial converters and read from the serial-to-parallel converter. To send data to the PC, the DSP writes to the parallel-data register. To receive data from the PC the DSP reads from this register.

Physically the DSP system interface consists of a printed circuit board, which is attached to the DSP. This interface board holds the buffer, address decoder and the clock. The interface subunits are accommodated on separate, smaller boards and plug into sockets on the interface. The connections between the DSP, PC and the probe electronics are made via ribbon cables.

Logic Family Most of digital electronics has been designed with the low power Schottky (LS) logic family. The LS family belongs to the transistor-transistor logic (TTL) class and has been chosen, because it is readily available, commonly used and robust. The switching time for a chip of the LS class is typically 10 ns, which is much shorter than the period of the system clock. Compared with the DSP cycle time of 30 ns however, the 10 ns switching time is not negligible. To interface to the DSP the FAST (F) logic family is used, which has typical switching times of 3 ns. The logic circuits are not completely designed in F logic, since the most important components, shift registers 74F595 and 74F597, were unavailable.

When mixing faster logic families with slower ones, the power supply and

digital ground connections must be carefully laid out. The logic components of the two families were therefore decoupled. The fast components do not share power supply lines with the slower components apart from one star point. All components are connected to a ground plane.

5.3.4 Buffer, Address Decoding and Clock

Several DSP system interface components are shared between the interface to the probe electronics and the PC. These are the components that buffer the control signals and data lines, decode the addresses of subunits and provide a local clock signal. These components interface to the DSP via several control signal lines and a data bus. Figure 5.4 shows a block diagram.

The shortest possible time for the DSP to access data memory is one DSP clock cycle, i.e. 30 ns. The DSP is programmed to insert two waitstates when communicating with the interface. This means the address, data and the read and write strobes are held for 90 ns in total. This provides enough time for the data lines to settle and the address to be decoded.

Buffering Both the read strobe and the bank select signal are connected directly to the drivers without buffering, to ensure the shortest possible switching time for the drivers. Drivers of 74F logic family are chosen to keep within the timing limitations of the DSP.

System Clock The system clock has a period of 100 ns. The two derived clock signals $\mathtt{CLK}_2$ and $\mathtt{CLK}_4$ have periods of 200 ns and 400 ns. All clock signals are buffered and distributed to the interface subunits and the probe electronics. The clock signal to the probe electronics is passed through a 100 Ω resistor. Together with stray capacitance, the resistor effectively forms an *RC* low-pass filter, which lengthens the signal rise time. This lowers the digital noise radiated by the ribbon cable connecting the interface and the probe electronics.

5.3.5 Conversion between Parallel and Serial Data

As previously mentioned, data are transmitted between the DSP and the probe electronics by means of parallel-to-serial converters and serial data links. A major part of the interface between the DSP and the probe electronics are the subunits of the DSP system interface that convert between the parallel data from the DSP and the serial data coming from the probe electronics. These subunits are shown in figure 5.5. Ordered by function they are

- eight parallel-to-serial converters, which connect to the eight digital-to-analog converters that subsequently generate the high-voltage control signals for the probe head actuators,
- one parallel-to-serial converter, which connects to the digital-to-analog converter that generates the bias voltage between tip and sample, and
- one serial-to-parallel converter, which receives data from the analog-to-digital converter that acquires the tunneling current signal.

The DSP exchanges data with the subunits via buffered data lines, which are an extension of the DSP data memory bus. The number of connected data lines is equal to the number of bits of the corresponding converter. After a parallel-to-serial converter receives a word from the DSP, it sends this word in serial form to the probe electronics. The DSP can input data from the probe electronics by selecting and reading from the serial-to-parallel converter.

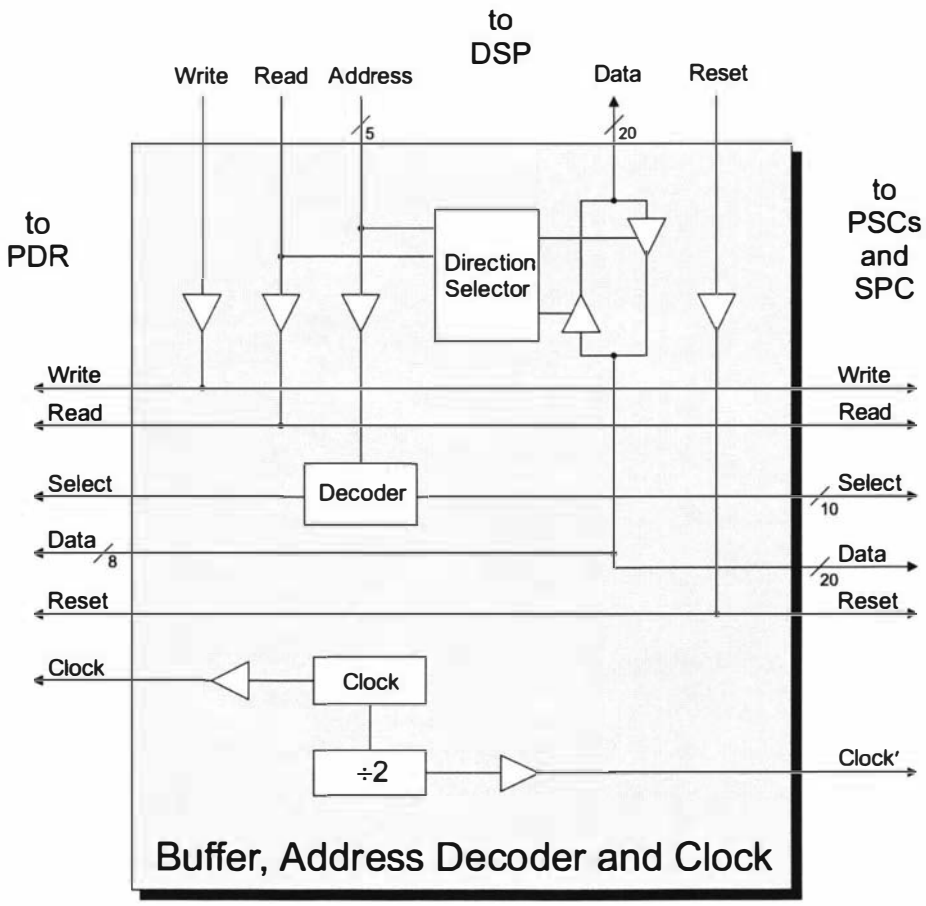


Figure 5.4: **Block diagram of the Buffer, Decoder and Clock:** The parts shown in the diagram link the DSP to the parallel-data register (PDR), to the parallel-to-serial converters (PSC) and to the serial-to-parallel converter (SPC). The DSP read strobe, write strobe and five address lines serve to select the individual subunits of the DSP system interface. The buffered strobes are directly connected to the subunits. The address lines are fed via a buffer to an address decoder. The data lines are provided by the lower 20 lines of the DSP data bus and are buffered bidirectionally with three-state devices. The data flow direction is determined by the read strobe, which is connected to a direction selector. The DSP selects the data buffers with a bank select signal, which is one of the five address lines. The buffered DSP reset signal is supplied to all subunits. The DSP system interface also carries a clock module, which is independent of the DSP. The module provides local clock signals for the subunits. The clock signals are well buffered, so that one clock line goes to a maximum of only two subunits.

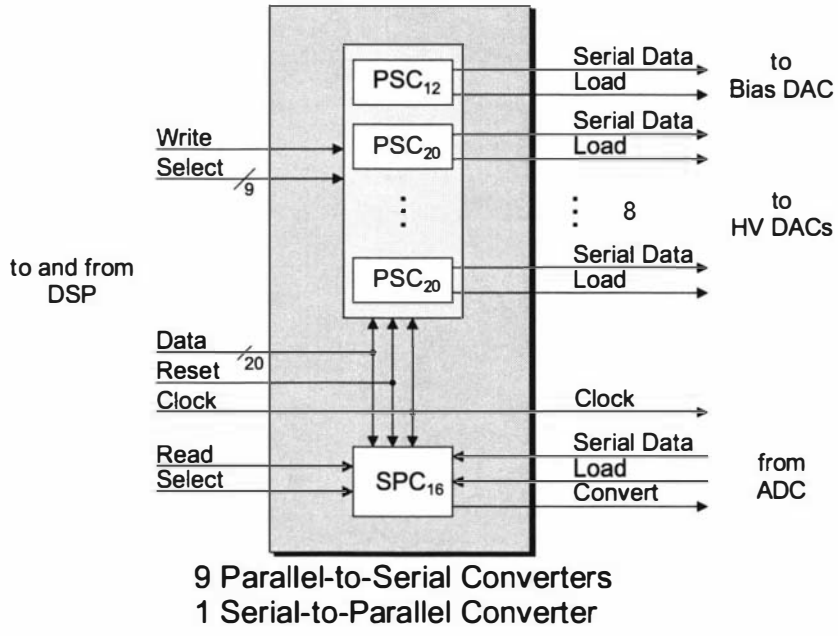


Figure 5.5: **SPC and PSCs:** Several subunits convert between the parallel DSP data bus and the serial data channels which connect with the probe electronics. These consist of eight 20-bit parallel-to-serial converters (PSC_{20}), one 12-bit parallel-to-serial converter (PSC_{12}) and one 16-bit serial-to-parallel converter (SPC_{16}). Each subunit also sees the reset from the DSP and is synchronized by a clock signal. The DSP can output data to the probe electronics by selecting and writing to any of the PSCs. Conversely, the DSP inputs data by selecting and reading from the SPC.

The DSP can access the subunits quickly, because the main components of each subunit are memory-mapped registers. While the parallel-to-serial converters are downloading serial data, the DSP can access another of the subunits, since the timing of data transmission to the probe electronics is independent of the DSP. As the DSP access time is much faster than the time it takes to upload or download serial data from the probe electronics, the DSP can practically operate all these subunits in parallel.

Now the serial-to-parallel converter receives serial data from the probe electronics without being prompted by the DSP. It is the serial-to-parallel converter that continuously and regularly triggers the conversion of tunneling current values. Since there is no acknowledgement by the parallel-to-serial converters or the serial-to-parallel converter that they have received a datum, the DSP has to make sure that it does not overwrite its last datum.

Before the individual subunits are discussed, the common principle of converting between parallel and serial data is explained.

Principle of Conversion

In this design, parallel data are converted to serial data using a shift register. The parallel data enter the shift register, are stored and leave the shift register serially. Several shift registers can be cascaded if a single component does not have a sufficient number of parallel entries. Cascading is explained in appendix B.0.1.

The register used in this design combines a storage register and a shift register. Parallel data are internally shifted between shift and storage register after they have been loaded. Using two different registers means that data can be accessed on both the parallel and the serial side. For true independence of both sides, the internal shift command must not coincide with access to either of the two registers.

The shift register needs several logic signals to handle data input and output. These signals are generated by additional hardware located on the converter board, consisting of gates, flip-flops and counters. The way the signals are generated is explained in the appendix.

Serial-to-Parallel Converter

The serial-to-parallel converter is one of the subunits of the DSP system interface. It converts serial data from the probe electronics to parallel data which is made available to the DSP. A diagram is shown in figure 5.6. The central parts of the serial-to-parallel converter in figure 5.6 are two registers, which accept serial input data and output parallel data. Both registers are located in one component, the 74LS595. The serial data are clocked into the shift register with a pulse train accompanying the data. The data are then latched into the storage register on the edge of a transfer pulse (STCP). Finally the DSP reads the data from the storage register on the lower 16 data lines (DMD[8..23]).

Analog-to-Digital Converter The design of the serial-to-parallel converter depends on the analog-to-digital converter residing in the probe electronics. The analog-to-digital converter that was used is the ADS7809P from Burr Brown.

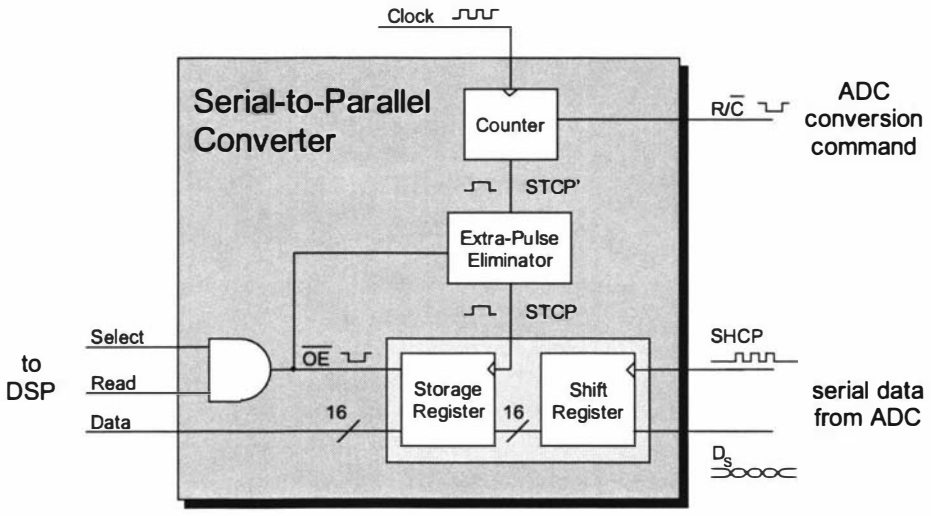


Figure 5.6: **Serial-to-Parallel Converter:** 16 bits of serial data D_S coming from the analog-to-digital converter in the probe electronics are converted to parallel data. The shift register is the central component of the converter. The serial data are clocked into the shift register with the data clock SHCP. The same data appear in parallel form at the input of the storage register, where the data are latched by the STCP signal. The DSP reads the data by asserting the read strobe and the address select. The counter circuit generates a conversion command ($R/\overline{C}$) for the analog-to-digital converter as well as STCP', a first form of the latch command STCP. The extra-pulse eliminator prevents coincidence of $\overline{OE}$ and STCP'.

The data conversion is synchronized to an internal clock, while the 16-bit output (D_{SA}) is accompanied by a corresponding train of 16 pulses (SHCP). The ADS7809P starts conversion after an external conversion command ($R/\overline{C}$) is applied. The conversion command needs to be generated at intervals of not less than 10 μs.

Conversion Command $R/\overline{C}$ The conversion command $R/\overline{C}$ required by the analog-to-digital converter is generated independently of the DSP by the serial-to-parallel converter in a sub-circuit labeled Counter. The $R/\overline{C}$ command is generated by counting a number of clock cycles, in fact it is the inverted carry pulse of the counter. The input clock to the counter circuit has a period of 400 ns. The clock period is first doubled and then counted with a 4-bit counter to give an interval of[5]

$$16 \times 2 \times 400\,\text{ns} = 12.8\,\mu\text{s}.$$

The conversion command $R/\overline{C}$ is then sent to the analog-to-digital converter in the probe electronics. After the conversion is completed, the shift register receives the serial data D_S and the accompanying pulse train SHCP. The data are then transferred to the storage register, where they can be read by the DSP.

[5]The output interval of 12.8 μs is not as short as it could be. To output a conversion command at a faster rate the serial-to-parallel converter would have to be changed to count fewer clock cycles. If the 4-bit counter counted to 13 instead of 16 then an interval of $13 \times 800\,\text{ns} = 10.4\,\mu\text{s}$ would be obtained. The first two doubling stages could be kept and counting to 13 would only be slightly more elaborate than counting to 16.

Transfer Pulse STCP The loaded serial data have to be transferred to the storage register at the same rate as the analog-to-digital conversions take place. The transfer pulse STCP is therefore derived from the same counter that also produces R/$\overline{C}$. The timing of the transfer pulse has a limitation: the shift must not occur at the same time that the storage register asserts the DSP data bus, for otherwise the DSP would upload incorrect data. This is discussed in the following paragraph.

'Spikes' The DSP accesses the shift register asynchronously with respect to the internal shift of the incoming data. Since the bus driver has different switching times for high-to-low and low-to-high transitions, the DSP would read the next datum 'mixed' with the actual datum. The uploaded data are only slightly incorrect for digital numbers that differ in the lower bits. However, when major bits are changing, the error becomes large. The changing of a major bit might be missed and the incorrect datum would appear as a spike in the data set. This happened before the extra-pulse eliminator was incorporated into the design. In the worst case of completely random data access by the DSP both events coincide once in (10 μs/100 ns) × 10 μs = 1 ms on average, which is of course intolerable.

Before it is delivered to the storage register, the transfer pulse passes an extra-pulse eliminator to ensure that only one internal shift occurs. The precursor of the transfer pulse is named STCP′ to distinguish it from the final form STCP.

Extra-Pulse Eliminator The extra-pulse eliminator prevents simultaneous access of the storage register by the transfer pulse and the output enable command $\overline{\text{OE}}$ from the DSP. The function of the eliminator is explained in appendix B.0.2.

Parallel-to-Serial Converters

Nine of the eleven subunits of the DSP system interface are parallel-to-serial converters. A block diagram of a 20-bit parallel-to-serial converter is presented in figure 5.7. The parallel-to-serial converters download serial data (D_S) to the corresponding digital-to-analog converters. They also generate and output load commands (LE) accompanying the serial data. A register is the basis of the parallel-to-serial converter. The register is divided into a storage and shift register. These registers are combined in a single component, the 74LS597. Each 74LS597 has eight inputs for parallel data. To provide a shift register with more than eight parallel inputs, the parallel-to-serial converters are built from two or three cascaded shift registers. Cascading is explained in appendix B.0.1.

Supporting circuits, a pulse tagger and a clock gater, provide load and shift commands for the register. The pulse tagger[6] generates a signal ($\overline{\text{SHLD}}$) to shift the latched parallel data internally from storage to shift register. A pulse train (SHCP) to clock the serial data out of the shift register is generated by using a pulse stretcher to gate the system clock. The clock gater generates the load command Load that accompanies the serial data sent to the probe electronics.

[6]The pulse tagger and pulse stretcher circuits are described in appendix B.0.2.

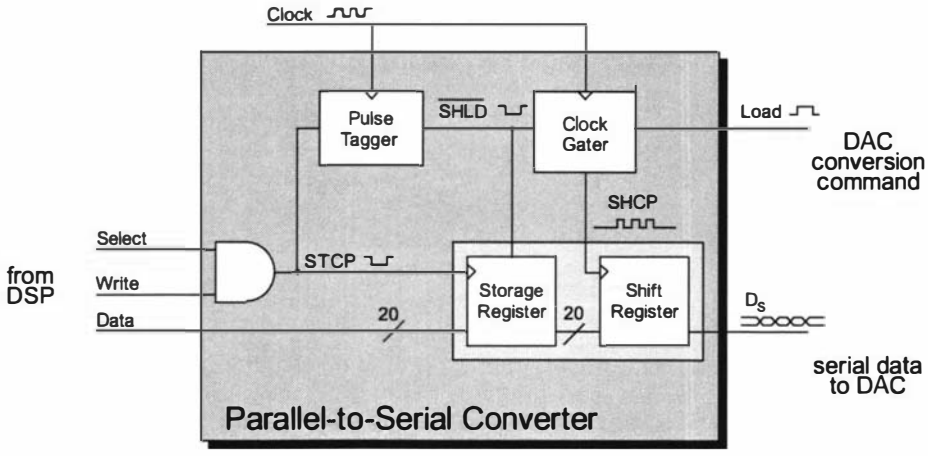


Figure 5.7: **Parallel-to-Serial Converter:** 20 bits of parallel data are converted into serial form by a shift register. The parallel data are first latched into a storage register with the combination signal STCP, transferred to the shift register by internal shift command $\overline{\text{SHLD}}$ and finally clocked out with the pulse train SHCP. The clock gater generates a load command that accompanies the serial data sent to a digital-to-analog converter in the probe electronics. To fulfill the tight timing requirements of the DSP, the component connecting to the DSP is a logic gate of the 'FAST' family (74F08).

The load command is derived from the counter at the end of the pulse train SHCP.

The operation of the clock gater is similar for both 20-bit and 12-bit parallel-to-serial converters. The main differences between the designs are the length of the pulse train and the timing and length of the load commands, which are tailored to suit the corresponding digital-to-analog converter.

Parallel-Data Register and Driver

As discussed in section 5.3.2, the DSP can exchange data with the PC via an 8-bit wide data link. The data link consists of two unidirectional 8-bit channels. At either end data are stored in registers.

PC Interface The PC uses a commercial input/output module for data exchange, the FPC-024 card from Flytech Technology, Taiwan, which is based on the 8255A programmable interface chip from Intel. The FPC-024 card in the PC has two 8255A chips, but only the one connected to connector CN1 is used. The strobed input/output type of communication of the 8255A has been chosen, since it is designed for unidirectional communication and it provides handshaking signals. The 8255A consists of three ports (A,B and C). In strobed input/output mode, Port A is programmed as strobed input and Port B as strobed output[7]. The third port (C) provides the handshaking signals.

A dedicated driver module supports the programmable peripheral interface 8255A in the PC. The driver unit consists of three drivers, one for each data channel and one for the control signals. The parallel-data register and the driver unit provide at total of 2×8 data lines and 5 control lines.

[7]The control word to achieve this combination of input and output is $B4_{16}$.

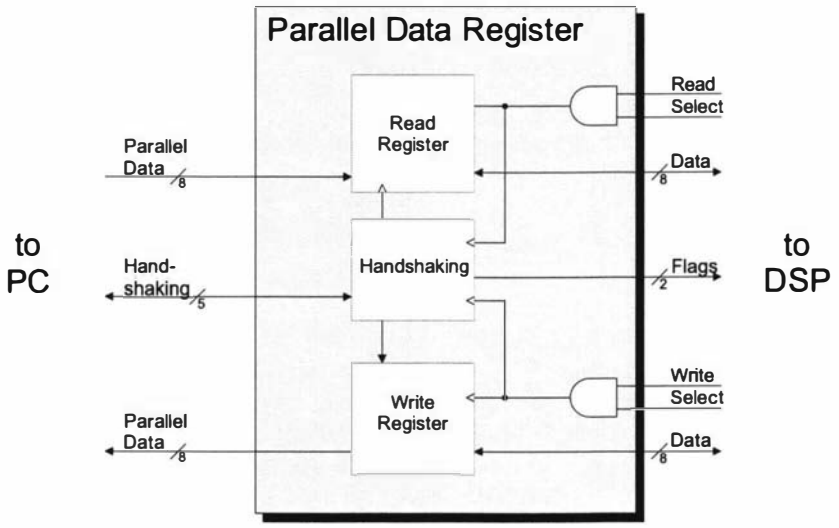


Figure 5.8: **Parallel-Data Register:** The parallel-data register is part of the interface between DSP and PC. Two unidirectional 8-bit wide parallel data links are provided for the interface. The DSP can read from the read register and write data to the write register. The data are communicated to the PC with handshaking signals. Two flags indicate whether the communication links are busy or clear.

Parallel-Data Register On the DSP a parallel-data register built to suit the 8255A handles the data exchange. A block diagram of the parallel-data register is shown in figure 5.8. The parallel-data register consists of two registers and of supporting circuits, which generate handshaking and flag signals. Handshaking signals are used to drive data lines and indicate the current status of communication, i.e. whether a channel is busy or available. The hardware to generate the handshaking signals has been built according to the specifications given by the manufacturer of the 8255A. Since the 8255A is relatively slow, several pulse stretchers, explained in appendix B.0.2, are used to generate the handshaking signals; output enable ($\overline{\mathrm{OE}}_{\mathrm{IO}}$), strobe ($\overline{\mathrm{STB}}$) and acknowledgement ($\overline{\mathrm{ACK}}$). The communication protocol between the DSP and PC is described in appendix B.0.3.

All incoming signal lines to the parallel-data register are terminated. The connection between DSP and PC is made with a 34-way ribbon cable. Most of the data and control lines in the ribbon cable are separated by either a power supply line or the digital ground, so as to minimize cross-talk.

DSP Interface Both parallel channels share the lowest 8 lines of the DSP data bus and the select signal. The read and write strobe of the DSP are connected to the corresponding channel. Each channel has a DSP flag associated with it so as to signal the communication status to the PC. The DSP can access the memory-mapped registers quickly. The necessary handshaking signals are generated by dedicated hardware, while the DSP is free to process other data.

5.4 Probe Electronics

5.4.1 Overview

The probe electronics stands between the DSP system interface and the microscope probe-head. An overview diagram is shown in figure 5.2. The probe electronics exchanges digital data with the DSP via the DSP system interface. It also forms part of the control loop for the tip-sample distance as shown in figure 4.5. The probe electronics provide output channels for the piezoelectric actuators and the bias voltage, and an input channel for the tunneling current. All output channels consist of a digital-to-analog converter followed by a low-pass filter, then an amplifier. The input channel consists of a current-to-voltage converter, a low-pass filter and an analog-to-digital converter. The digital-to-analog converters receive digital data from the DSP and convert them to analog control voltages that drive the microscope actuators, or adjust the bias voltage between tip and sample. Tunneling current values are converted to voltages by the current-to-voltage converter, converted to digital form and then sent to the DSP.

Organization

The probe electronics are organized in the following way: modular subunits plug into a backplane, which distributes data and power. The output channels and most of the input channel constitute one subunit each. The input channel subunit is referred to as the ADC module in the text; the output channels driving the piezoelectric actuators as high-voltage (HV) modules, and the output channel to change the bias voltage as the Bias module. Another subunit houses optocouplers, which provide an interface for digital signals between the DSP system interface and the probe electronics. All digital signals pass through this subunit. The signals are distributed to the remaining modules via the backplane. A separate power supply provides the necessary voltages for the probe electronics including the high voltage required by the actuator drivers. The power is also distributed via the backplane. Apart from carrying all digital signals and power supply lines, the backplane also provides a low impedance ground and adds shielding. There are no analog signals on the backplane.

The probe electronics are housed in a separate enclosure from the DSP and the DSP system interface. This has its own backplane and power supply. The current-to-voltage converter is installed within a separate metal casing, positioned close to the microscope probe head. The connection between the DSP system interface and the probe electronics is made with a ribbon cable and the connections to the microscope are made with coaxial cables.

Power Supply Lines

An important design goal was to minimize the coupling of digitally generated noise to the analog probe electronics. Because of this, there are separate supply and return lines to power the digital part of the probe electronics. The power supply for the digital components is V_{s2} and its dedicated return is the 'high-level analog ground'. Note that this is not the digital ground of the DSP system interface. All analog components are supplied with power supplies $\pm V_{s1}$. The

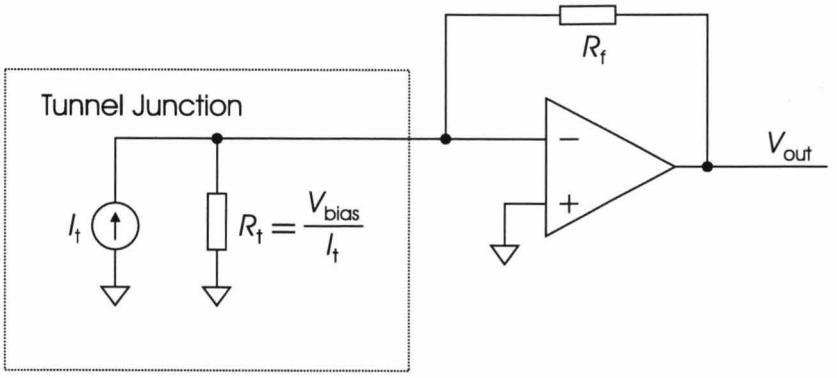


Figure 5.9: **Transimpedance Amplifier and Source Model:** An operational amplifier in transimpedance configuration is shown to the right. The tunneling junction is modeled as a current source. The tunneling junction is non-linear and therefore the tunneling resistance changes with tunneling conditions, usually between 1 MΩ and 1 GΩ. The feedback resistor has a value of $R_f = 90\,\mathrm{M\Omega}$. The gain of the amplifier is equal to this, i.e. 90×10^6 V/A.

corresponding return is the 'low-level analog ground'. The high voltage amplifiers are supplied by the high voltage supply $\pm V_{\mathrm{high}}$. This supply shares the low-level ground as a return. High-level and low-level ground are connected at only one point on the backplane. Both grounds run physically separated on the top and bottom of the backplane. The grounds cover both sides, front and back, within their respective areas. This scheme should give good ground separation while still providing low impedance grounds for both power supplies.

5.4.2 IVC Module

The IVC module is a current-to-voltage converter that turns the small tunneling current I_t into a high level voltage V_t. As the first stage of the input signal channel, it is a crucial system component. The position of the converter within the electronics is shown in figure 5.2. The converter is also an important element of the microscope control loop as is shown in figure 4.5. The converter is physically separate from the ADC module and is located as close as possible to the actual signal source, the microscope probe head.

The IVC module consists of a transimpedance amplifier and a driver. A circuit diagram of the probe and transimpedance amplifier can be seen in figure 5.9. In this circuit the tunneling junction is modeled as a current source with varying tunneling resistance (R_t) given by

$$R_t = \frac{V_{\mathrm{bias}}}{I_t}.$$

The tunneling junction is best modeled as a current source for two reasons, firstly the bias voltage is usually fixed and secondly the encountered tunneling resistance is very large. Tunneling resistances of 1 GΩ are not unusual.

An operational amplifier in transimpedance configuration is used, in which the non-inverting terminal is at ground potential and the input current is converted via a feedback resistance (R_f) to give

$$V_{\mathrm{out}} = -R_f I_t.$$

This signal is buffered by a driver to minimize the current drawn from the operational amplifier, as recommended by the manufacturer [6]. The driver is designed to handle the capacitive load presented by a coaxial cable of up to 3 m in length.

Power Supply and Shielding Both power supply filtering and shielding of the circuit are important in attenuating externally generated noise. Therefore the power supply lines to the transimpedance amplifier contain small resistors (47 Ω). The resistors together with by-pass capacitors form a low-pass filter, which reduces noise coupled via the power supply lines.

The input to the transimpedance amplifier must be shielded [112]. The IVC module is therefore mounted inside a metal enclosure and the input wire is shielded all the way from the tip holder to the transimpedance amplifier. Other effects that generate unwanted input currents are triboelectric, piezoelectric and microphonic effects in the input cable of the IVC [112]. These effects are minimized by securely mounting the module and providing acoustic isolation.

Conversion Options

It is a challenge to build a current-to-voltage converter that simultaneously possesses high gain, broad bandwidth and low noise. A loss in quality in one of these characteristics can be traded against good performance in the other two. Measurement of low level currents is discussed in detail in the handbook of Keithley [112]. The use of operational amplifiers to build current-to-voltage converters is also discussed in the application handbook of Burr Brown [39]. The kind of compromise made depends on the application. For instance, current-to-voltage converters are often used as photodiode amplifiers, where they have very high gain, but bandwidths of only a few Hertz [39]. Amplifier designs with high bandwidth and high gain have been published by Birk *et al.* [30] and Jersch *et al.* [107]. Very sensitive current-to-voltage converters are reported by Buksh [37] and Heer *et al.* [92].

In STM the current-to-voltage converter should have a very high gain of between 10^8 and 10^9 V/A, a signal-to-noise ratio of about 60 dB and a bandwidth of about 10 kHz. The high gain is needed, as signal currents are very small (0.1 to 10 nA) and an output voltage of up to 10 V is desirable. To convert a current into a voltage one can simply use a small resistance followed by voltage amplification [5, 84, 112]. The small resistance is a load for the signal source and the voltage developed across it is known as the voltage burden [112]. The voltage burden should be kept small, therefore the resistance and with it the gain has to remain small.

An alternative to the small resistor is a transimpedance amplifier, which can have a large gain without compromising the small input resistance. The input resistance of the transimpedance amplifier is given by

$$R_{\text{in}} = \frac{1}{1+A} R_{\text{f}},$$

where A is the open-loop gain of the operational amplifier. The voltage burden is clearly reduced by the same factor; $1/(1+A)$. Since the low frequency open-loop gain is very high for all operational amplifiers the input resistance will therefore be very small. An important point to note is that in the actual circuit, parasitic

capacitance cannot be ignored. This is discussed in a later section. A second reason for choosing a transimpedance amplifier is the possibility of a better signal-to-noise ratio, which *rises* with the square root of the resistance used to convert the input current.

Choice of Feedback Resistance

The size of the feedback resistance has a bearing on the two important amplifier criteria, which are noise and bandwidth. A large resistance produces the better signal-to-noise ratio but a small bandwidth, while a small resistance offers a broader bandwidth at the cost of a decreased signal-to-noise ratio.

The noise added to the tunneling current signal is mainly amplifier voltage noise and the Johnson noise of the feedback resistance. If we only consider the Johnson noise then the signal-to-noise ratio rises with the square root of the resistance value R_f, since the signal is proportional to R_f and the noise proportional to $\sqrt{R_f}$, i.e.

$$\frac{s}{n} \propto \frac{R_f}{\sqrt{R_f}} = \sqrt{R_f}.$$

A larger feedback resistance therefore increases the signal-to-noise ratio. A large feedback resistor is also of advantage when considering amplifier voltage noise, which is discussed in greater detail in appendix B.2.5. From equation (B.38) the leading term of the low-frequency amplifier noise is found to be proportional to

$$\left(\frac{R_t}{R_t + R_f}\right)^{\frac{1}{2}},$$

where R_f is the feedback resistance and R_t is the tunneling resistance as defined in figure 5.9. The voltage amplifier noise therefore decreases strongly with increasing R_f once R_f is of the same order or larger than R_t. A comparison of amplifier voltage noise and Johnson noise is deferred until section 5.4.4.

In the current amplifier design a feedback resistance of

$$R_f = \frac{V_t}{I_t} = 90\,\mathrm{M\Omega}$$

was employed. A higher value of $R_f = 1\,\mathrm{G\Omega}$ was initially considered since it should result in a better signal-to-noise ratio. However, because of the relatively high input capacitance of the current design such a high feedback resistor reduced the bandwidth to below 1 kHz. The possibility of using a higher value feedback resistor should be reconsidered if the current-to-voltage converter is to be re-designed.

Choice of Operational Amplifier

Since minute input currents have to be amplified, a FET input operational amplifier was used. This type of amplifier affords reasonable output voltage accuracy, since FET input operational amplifiers have very small input bias and offset currents. Several possible operational amplifiers are listed in table 5.1. Of the operational amplifiers listed the AD549 from Analog Devices was chosen, since neither the OPA128 nor the OPA129 were available at the time of design

Model	Bias Current (fA) max	Unity Gain (MHz)	Input Voltage Noise ($\frac{nV}{\sqrt{Hz}}$)	Input Current Noise ($\frac{fA}{\sqrt{Hz}}$)
OPA111	2 000	2	8 (1 kHz)	0.5 ((0.1 . . . 20) kHz)
OPA128[a]	75	1	15	0.22 ((0.1 . . . 20) kHz)
OPA129[a]	100	1	17 (1 kHz)	0.1 (10 kHz)
AD548	20 000	1	30 (1 kHz)	1.8 (1 kHz)
AD549	250	1	35 (1 kHz)	0.22 (1 kHz)

[a]was not available at the time of system design and construction

Table 5.1: **Low Bias Current Operational Amplifiers:** The table lists operational amplifiers that have very low input bias current. Since very small signal currents, 10 pA, should be measurable only those amplifiers which provide low offset and low input current noise are considered. The unity-gain bandwidth is very important in determining the overall bandwidth of the current-to-voltage converter and should be as high as possible. In this respect the specifications are very similar. The voltage noise contributes strongly to the overall noise, because voltage noise has a broader bandwidth than the input current signal. The listed values are the level of white noise, which is the noise most important at higher frequencies.

and construction. Since this has now changed the OPA129 would be recommended for a new design. Apart from having better specifications it is also well priced.

The AD549JN has an input bias current of $I_b = 250\,fA$ accompanied by a current noise of $0.22\,fA/\sqrt{Hz}$. To minimize additional offsets the input lead was mounted on a teflon stand-off[8]. In addition to having a very low bias current the AD549JN also has a small offset voltage, $V_{os} = 500\,\mu V$. The temperature drift of the AD549JN is $20\,\mu V/K$, which is equivalent to a temperature drift of $0.2\,pA/K$, which is very small. The first-order errors introduced by the operational amplifier can be calculated from the expression

$$V_{out,os} = V_{os}(1 + \frac{R_f}{R_t}) + I_b R_f. \tag{5.5}$$

The value of the tunneling resistance can be determined by measuring the bias voltage and the tunneling current. A minimum value of $R_t = 5\,M\Omega$ and a maximum value of $R_t = 3\,G\Omega$ are expected. The input bias current term is 0.25 pA and negligible. With these values $V_{out,os}$ will be between 0.5 and 9.5 mV, depending on the tunneling resistance. From equation (5.5) it can be seen that the transimpedance amplifier works best with high tunneling resistances (R_t), which is reasonable since an ideal current source should have an infinite source resistance[9]. Reflected to the input the offset voltage is equivalent to a tunneling current offset of

$$\frac{0.5 \text{ to } 9.5\,mV}{90\,M\Omega} = 6 \text{ to } 110\,pA.$$

The driver IC adds an equivalent current offset of 2 pA, which is comparatively

[8]Both the stand-off and the feedback resistor have to be as clean as possible to minimize leakage currents. If at all possible, the feedback resistor must not be touched with bare hands. Humidity can also be a problem.

[9]A consequence of the above equation is that the output voltage will saturate if the transimpedance amplifier input is short circuited.

small. The total offset error of the IVC module will therefore have a maximum value of 110 pA, due mainly to the AD549JN input offset voltage.

The actual measured current offset under open circuit conditions ($R_t \to \infty$) was

$$\frac{100\,\mu\text{V}}{90\,\text{M}\Omega} = 1.1\,\text{pA},$$

which is smaller than the 6 pA. However, it has to be remembered that the range of offset voltage values given above are maximum values from the data sheet.

Speed The bandwidth of the transimpedance amplifier is limited by two parasitic capacitances, the input capacitance of the current-to-voltage converter, and the parasitic capacitance of the feedback resistor. The input capacitance is interacting with the open-loop gain of the operational amplifier and is discussed below. The parasitic feedback capacitance is discussed here.

Because of the large value of the feedback resistor, even a small parasitic capacitance of a few femtofarads will result in a cut-off frequency of several kilohertz. The roll off frequency of the feedback impedance Z_f consisting of feedback resistor R_f and parasitic capacitance C_f is

$$\omega_f = \frac{1}{R_f C_f}. \tag{5.6}$$

The estimated parasitic capacitance of the resistors that have been used is 250 fF, resulting in the cut-off frequency of 7 kHz. The way the capacitance is estimated is explained below.

There are several ways to reduce this effect. One method is to include an extra amplification stage. This might be within the feedback loop to increase the open-loop gain or it might be a second stage following a low-gain current-to-voltage converter. Another method would be to add a compensation network in the feedback path so as to counter the decrease of the feedback impedance with frequency. Such a network is discussed in appendix B.2.7.

The large feedback resistance also influences the high-frequency signal-to-noise ratio of the operational amplifier. To explain this fact, both the signal and noise gain of the amplifier have to be considered.

Analysis of the Transimpedance Amplifier

Ideal Transimpedance The transimpedance amplifier used with the microscope can be modeled by the circuit shown in figure 5.10. The circuit includes two parasitic capacitances; one across the feedback resistor and one at the input to the amplifier, the latter capacitance includes the input capacitance of the input cable to the current-to-voltage converter as well as any capacitance bypassing the tunneling resistance.

In this circuit the tunneling junction is again modeled as a current source with a tunneling resistor (R_t). A general analysis of this amplifier circuit, assuming ideal components, is done in appendix B.2.4 and delivers the following result

$$V_{out} = -Z_f I_t \tag{5.7}$$

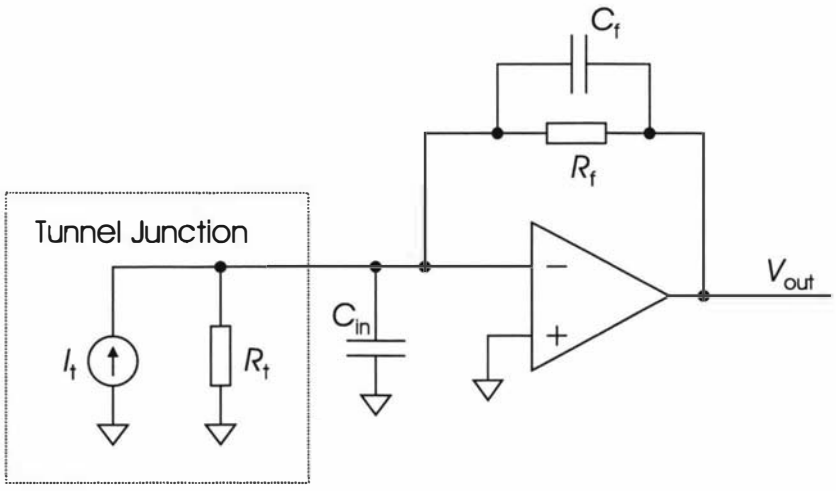


Figure 5.10: **Transimpedance Amplifier with Parasitic Capacitances:** The transimpedance amplifier from figure 5.9 is shown with two parasitic capacitances that are usually present in the real circuit. The input capacitance (C_{in}) is largely due to the shielded input cable. The feedback capacitance (C_{f}) is the very small parasitic capacitance of the feedback resistor. The source is modeled as in figure 5.9. Since the resistance values involved are very large, the capacitances have a significant effect on the signal bandwidth of the transimpedance amplifier. Expected values for the tunneling resistance are between 5 MΩ and 3 GΩ. The capacitance values are calculated from a fitted curve, as discussed below. They are $C_{\text{f}} = 0.25\,\text{pF}$ and $C_{\text{in}} = 70\,\text{pF}$.

for the output voltage, where Z_{f} is the feedback impedances and I_{t} is the tunneling current. For an ideal operational amplifier the input capacitance (C_{in}) has no influence on the signal gain.

To find the ideal transimpedance of the circuit in figure 5.10, the impedance for a resistor in parallel with a capacitor

$$Z_R \parallel Z_C = \frac{Z_R Z_C}{Z_R + Z_C} = \frac{R}{1 + j\omega RC}$$

replaces Z_f in equation (5.7). The variable ω is the angular frequency of the tunneling current and j is the imaginary unit. The ideal output voltage is found to be

$$V_{\text{out}} = -\frac{R_{\text{f}}}{1 + j\omega R_{\text{f}} C_{\text{f}}} I_{\text{t}}\,. \tag{5.8}$$

Now a transimpedance (T) can be defined for the amplifier as

$$T = \frac{V_{\text{out}}}{I_{\text{t}}} = -\frac{R_{\text{f}}}{1 + j\omega R_{\text{f}} C_{\text{f}}}\,, \tag{5.9}$$

which is equation (B.35) in appendix B.2.4. At dc, the transimpedance is

$$T_{\text{dc}} = -R_{\text{f}},$$

i.e. just the feedback resistance of the transimpedance amplifier. The frequency dependence of the transimpedance is given by the roll-off caused by the parasitic feedback capacitance, i.e.

$$\frac{T}{T_{\text{dc}}} = \frac{1}{1 + j\omega R_{\text{f}} C_{\text{f}}}\,.$$

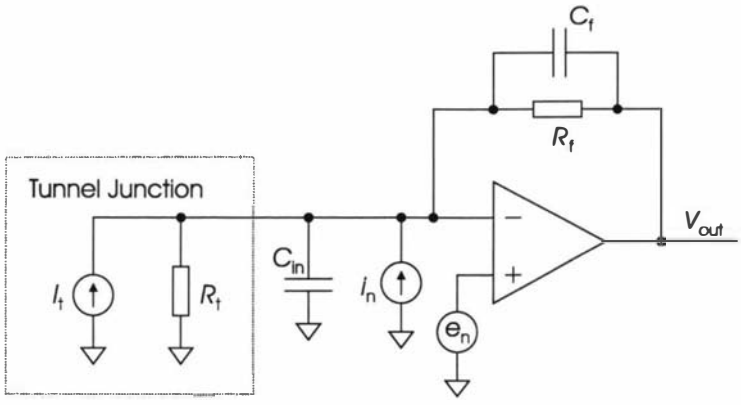


Figure 5.11: **Noise Model for the Transimpedance Amplifier:** Two noise sources, amplifier voltage noise (e_n) and amplifier current noise (i_n), are added to the transimpedance amplifier circuit, which is otherwise equal to the one shown in figure 5.10. At 1 kHz the noise values from the operational amplifier data sheet [6] are $e_n = 35\,^{nV}/\sqrt{Hz}$ and $i_n = 0.22\,^{fA}/\sqrt{Hz}$.

Ideal Noise Gain An operational amplifier configured as an inverting or transimpedance amplifier has a higher noise than signal gain. This arises, because the input voltage noise of the amplifier can be modeled as occurring between ground and the non-inverting terminal as shown in figure 5.11. The noise is therefore amplified by the gain of a non-inverting amplifier.

The ideal noise gain (G_n/G_{n0}) is derived in appendix B.2.3 as equation (B.27), and is repeated here

$$\frac{G_n}{G_{n0}} = \frac{1 + j\omega(R_t \parallel R_f)C_{tot}}{1 + j\omega R_f C_f}, \tag{5.10}$$

where

$$G_{n0} = \frac{R_f + R_t}{R_t} \tag{5.11}$$

is the noise gain at zero frequency, derived in equation (B.28), and

$$C_{tot} = C_{in} + C_f$$

is defined as the total capacitance. Note that the input capacitance does have an influence on the noise gain.

According to equation (5.10), the noise gain remains at the dc level until the first corner frequency (ω_{tft}) is reached at $\omega_{tft} = ((R_t \parallel R_f)C_{tot})^{-1}$ after which the noise gain *rises* until it reaches a second corner frequency at $\omega_f = (R_f C_f)^{-1}$. At frequencies higher than the second corner frequency the noise gain is determined by the input and feedback capacitances

$$G_{nh} = C_{tot}/C_f \tag{5.12}$$

Equation (5.10) should be compared to the ideal signal gain of equation (5.8). The ideal signal gain shares the same high-frequency cut-off point as the ideal noise gain.

The corner frequency ω_f given by the feedback impedance plays a crucial role for both signal and noise gain. While the signal gain starts falling at ω_f, the noise gain only stops rising at the same frequency. The noise gain and with

it the amplifier voltage noise can be reduced by shifting ω_f to lower frequencies, but only at the cost of decreasing signal bandwidth. A more promising way of reducing noise gain is to shift the rising corner frequency of the noise gain (ω_{tft}) towards higher frequencies. This corner frequency depends strongly on the total capacitance as defined above. In contrast to the noise gain, the ideal signal gain is not affected by the input capacitance. Reducing the input capacitance therefore increases the possible signal-to-noise ratio of the current-to-voltage converter.

The ratio of signal gain to noise gain at low frequencies is given by

$$\frac{T_{dc}}{G_{n0}} = \frac{R_f R_t}{R_f + R_t} = R_f \parallel R_t.$$

The signal-to-noise ratio is thus determined by the smaller of the two resistances. Since the tunneling resistance is very large it is beneficial for the low-frequency signal-to-noise ratio to have a large feedback resistance.

However, a large feedback resistance value decreases both corner frequencies of G_n, which is detrimental for the high-frequency signal-to-noise ratio and the signal bandwidth. On the one hand the noise will start increasing at lower frequencies while on the other hand both signal and noise gain are limited by a low value of ω_f.

Since the ideal signal gain begins to roll off at ω_f, while the ideal noise gain only reaches a constant level, the operational amplifier is mainly amplifying noise at higher frequencies. To achieve the best possible signal-to-noise ratio one should amplify only in frequency regions where most of the useful signal lies. This can either be done within the feedback loop by actively decreasing the loop gain with a second amplifier or after the feedback loop with the help of a low-pass filter.

Real Amplifier The input capacitance of the current-to-voltage converter has an effect on the noise, stability and bandwidth of the real amplifier. The effect on noise is already evident from the ideal noise gain. Additionally the input capacitance interacts with the open loop gain (A) of the amplifier.

The open-loop gain A is modeled with a single roll-off corner frequency as

$$A = \frac{A_0}{1 + j\omega \frac{A_0}{\omega_u}}, \tag{5.13}$$

where A_0 is the dc open-loop gain and ω_u is the unity-gain frequency. In equation (5.13) the approximation

$$\omega_c = \frac{\omega_u}{\sqrt{A_0^2 - 1}} \approx \frac{\omega_u}{A_0}$$

was used for the corner frequency of the open-loop gain. The approximation is valid for $A_0 \gg 1$, which is the case for all operational amplifiers and in particular the amplifier used here for which $A_0 = 10^5$.

Measuring the Frequency Response

The frequency dependence of the current-to-voltage converter output voltage was measured. The measurements are shown in figures 5.13.

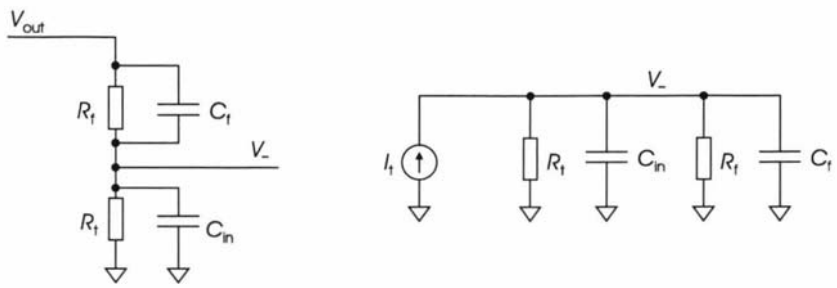


Figure 5.12: **Superposition Principle:** To find the voltage at the inverting terminal (V_-) of the transimpedance amplifier, the circuits are re-drawn from figure 5.10 under two different conditions. To the left the current source was replaced with an open circuit, while to the right the output voltage was replaced with a short circuit.

A sinusoidal input current was provided by a resistor of $R_t = 29\,\mathrm{M}\Omega$ connected to a function generator. The measurement showed a high frequency roll-off starting at a rather low frequency of around 7 kHz with an attenuation slope of $-40\,^{\mathrm{dB}}/_{\mathrm{dec}}$. This phenomenon can be explained by including the frequency behavior of the open-loop gain (A) according to equation (5.13) into the analysis of the model of figure 5.10.

To analyze the current to voltage converter circuit from figure 5.10, the voltage at the inverting terminal of the transimpedance amplifier V_-, is calculated. For an ideal operational amplifier this voltage is zero. In figure 5.12 two circuits are drawn that are used to calculate V_- with the superposition principle. The output voltage and the tunneling current are replaced by their ideal equivalent, i.e. short circuit and open circuit, the voltage V_- is calculated for both cases and added. The left hand circuit is a resistive and capacitive divider, which is a circuit that is also used for frequency compensation of oscilloscope probes. It has two corner frequencies and V_- is given by

$$V_-\,(\text{left}) = \frac{R_t}{R_t + R_f}\,\frac{1 + j\omega R_f C_f}{1 + j\omega (R_t \parallel R_f) C_{tot}}\,V_{out}\,.$$

The right hand circuit shows that the current source sees the two resistors and the two capacitors in parallel. The voltage V_- is found to be

$$V_-\,(\text{right}) = \frac{R_t \parallel R_f}{1 + j\omega (R_t \parallel R_f) C_{tot}}\,I_t\,.$$

The sum of the voltages $V_-\,(\text{left}) + V_-\,(\text{right})$, gives the actual voltage V_- at the inverting terminal according to the superposition principle:

$$V_- = \frac{R_t}{R_t + R_f}\,\frac{1 + j\omega R_f C_f}{1 + j\omega (R_t \parallel R_f) C_{tot}}\,V_{out} + \frac{R_t \parallel R_f}{1 + j\omega (R_t \parallel R_f) C_{tot}}\,I_t\,. \qquad (5.14)$$

For an ideal operational amplifier the voltage V_- is zero and the ideal transimpedance mentioned in equation (5.9) is again obtained. However, for a real amplifier the denominators of both terms in equation (5.14) do play a role. In appendix B.2.4 the finite open-loop gain of the operational amplifier is taken into account and in equation (B.37) the transimpedance is found to be

$$T = T_{dc} P\,, \qquad (5.15)$$

where $T_{dc} = -R_f$. The factor P is given in equation (B.23) as

$$P = \frac{\omega_0^2}{\omega_0^2 - \omega^2 + j\omega 2\delta}.$$

where ω_0 is defined in equation (B.21) as

$$\omega_0^2 = \frac{\omega_u}{R_f C_{tot}}$$

and δ in equation (B.22)

$$2\delta = \frac{1}{(R_s \parallel R_f)C_{tot}} + \omega_u \frac{C_f}{C_{tot}}.$$

This peak factor P is mainly a consequence of the interaction between the input capacitance of the transimpedance amplifier and the finite open-loop gain of the operational amplifier. Without either of them there is no peaking.

The data measured using a source resistor of 29 MΩ between tip and sample could be analyzed to find ω_0 and δ from the quality factor and peak of the data in figure 5.13. The undamped resonance frequency (ω_0) allows one to calculate a value for the total capacitance and the damping factor (δ) then leads to a value for the feedback capacitor. The other values are taken from the amplifier data sheet [6] or are otherwise known.

The values found are $C_{tot} = 66\,\text{pF}$ and $C_f = 254\,\text{fF}$. Equation (5.15) was then used to calculate the modeled frequency response. The result is shown in figure 5.13 as the solid line. The modeled curve interpolates reasonably well with the measured data. The data have been read off an analog oscilloscope and errors of about 5% could be included. The model is clearly good enough to predict the shape of the measured response.

The other set of data, where the signal was capacitively coupled to the transimpedance amplifier, could also be modeled. In equation (B.18) $Z_s = 1/j\omega C_s$ was inserted. The result is a similar equation to equation (5.15), but here expressed as a voltage gain:

$$G = \frac{\omega_0^2}{\omega_0^2 - \omega^2 + j\omega 2\delta} j\omega R_f C_s, \tag{5.16}$$

except that δ is without the $1/(R_s C_{tot})$ term, since there is no source resistor, which is equivalent to saying that $R_s \to \infty$. The resonance curve is now multiplied by $j\omega R_f C_s$ due to the increased capacitive coupling with frequency. To get an estimate for the coupling capacity the initial rising slope of $+20\,^{\text{dB}}/_{\text{dec}}$ was extrapolated to the 0 dB value. At this point the input and output voltages are equal and from the ideal equation $V_{out} = R_f I_{in} = R_f C_s \omega V_s$ one gets

$$V_{out}(\omega_1) = V_s(\omega_1) \to C_s \approx \frac{1}{\omega_1 R_f},$$

where ω_1 is the frequency at $G = 0\,\text{dB}$. The coupling capacitance C_s was found to be approximately

$$C_s \approx \frac{1}{2\pi \times 13\,\text{kHz} \times 90\,\text{M}\Omega} = 140\,\text{fF}.$$

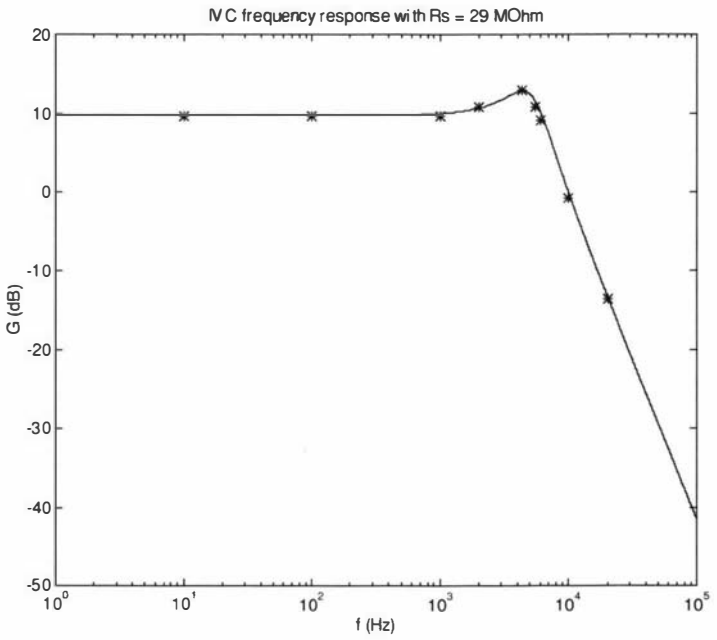


Figure 5.13: **Measured and Modeled Frequency Response:** The graph shows both measured data and a modeled frequency response overlaid. The data were recorded with the help of a signal generator and an analog oscilloscope. The signals were resistively coupled with a source resistance of $R_t = 29\,\mathrm{M}\Omega$. The solid line is the graph modeled by equation (5.15). The gain displayed here is the voltage gain, which behaves equally to the transimpedance. A comparatively good agreement can be seen for most data apart from two points just behind the peak. The uncertainties in the data have not been indicated. From the graph a peak frequency of $f_{\mathrm{res}} = 4.35\,\mathrm{kHz}$ and a quality factor of $Q = G_{\mathrm{res}}/G_{\mathrm{dc}} \approx \sqrt{2}$ is measured. The values used in the model were $R_t = 29\,\mathrm{M}\Omega$, $R_f = 90\,\mathrm{M}\Omega$, $f_u = 1\,\mathrm{MHz}$, $C_f = 0.25\,\mathrm{pF}$ and $C_{\mathrm{tot}} = 66\,\mathrm{pF}$.

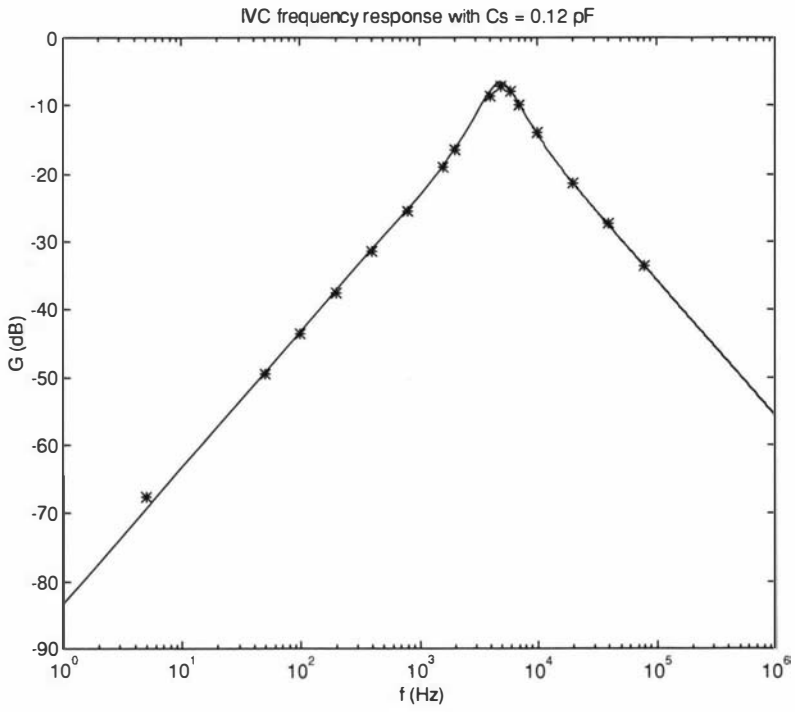


Figure 5.14: **Measured and Modeled Frequency Response:** The graphs shows the response of the current-to-voltage converter for capacitively coupled signals. Similar to figure 5.13 the data were recorded with the help of a signal generator and an analog oscilloscope. The drawn-out line is modeled by equation (5.16). Again a comparatively good agreement can be seen for most data. The tunneling tip was just out of tunneling distance. The source capacitance was estimated from the graph and found to be $C_s = 0.12\,\mathrm{pF}$. For the total capacitance a value of $C_{\mathrm{tot}} = 73\,\mathrm{pF}$ had to be used in the model, although this value seems overly high.

In the model the value was slightly modified until the modeled curve matched the measured data. The result is shown in figure 5.14. The assumed value for C_s was 120 fF and the value for the total capacitance was $C_{tot} = 73$ pF. A step response measurement of the capacitively coupled amplifier is shown in section 7.2, figure 7.2.

Input Capacitance The input capacitance (C_{in}) is the greatest contributor to the total capacitance (C_{tot}). This will still be the case even if C_{in} can be decreased to a few picofarads. The peak frequency is inversely proportional to $\sqrt{C_{tot}}$, while the damping factor δ is inversely proportional to C_{tot}. It is the input capacitance that has a strong influence on both peak frequency ω_{res} and damping factor δ.

The input capacitance was measured directly with a digital multimeter, a Dick Smith Q-1426, and again with a *LCR* meter, a B&K Precision 875A. The resulting input capacitance for the resistively coupled measurement was $C_{in} = (45 \pm 2)$ pF (multimeter) and $C_{in} = (49 \pm 1.2)$ pF (*LCR* meter). The measured capacitance values are not as high as the model predicted, although both meters are in agreement.

The individual capacitances of the components contributing to the input capacitance have also been measured. The capacitance of the tunneling current connection was found to be (5 ± 1) pF. By rearranging the transimpedance amplifier such a small value for the input capacitance should be achievable. The lowest possible value is approximately 1 pF, which is the input capacitance of the operational amplifier alone.

The interaction of the input capacitance with the open-loop gain of the amplifier has several consequences. The frequency response can have a peak and the amplifier has a phase shift of 90° at ω_0.

In a current-to-voltage converter it is not desirable to have a peak in the frequency response, so one strives to select the capacitances for an overdamped condition. There will be no resonance peak if

$$\delta \geq \frac{\omega_0}{\sqrt{2}}.$$

In the definition of δ in equation (B.22) the term $(C_f \omega_u)/C_{tot}$ is dominant. This is the largest term independent of the specific value of C_{tot}, since all three terms contain C_{tot}. To find a condition for δ with respect to ω_0 we disregard the other two terms and insert this δ into the overdamping condition. This gives a limit for the total capacitance with

$$C_{tot} \leq \frac{\omega_u R_f C_f}{4} C_f. \tag{5.17}$$

If the current feedback capacitor value of 0.25 pF is used, then the total capacitance should be

$$C_{tot} \leq 8.8\,\text{pF}$$

to avoid a peak in the frequency response. The current feedback capacitor would therefore be adequate if the input capacitance could be decreased to 5 pF, while it is currently slightly too small.

The gain of the current-to-voltage converter at $\omega = \omega_0$ is

$$G(\omega_0) = \frac{\omega_0}{2\delta}.$$

If the frequency response is critically damped with $\delta = \omega_0/2$ then the gain is $-3\,\mathrm{dB}$ below the dc value, i.e. ω_0 is then the corner frequency.

Phase Shift If the microscope is considered a control system, the chief problem with the current-to-voltage converter is the 90° phase shift that accompanies the frequency response at $\omega = \omega_0$. Since the controller is an integrator it has a phase shift of 90° already. The combination of these phase shifts will cause instability at some frequency higher than ω_0. In addition the time delay inherent in the digital system causes an extra phase shift which increases with frequency.

Increasing the Bandwidth To increase the bandwidth of the current-to-voltage converter one should first try to maximize ω_0 and then choose δ for critical damping. There are in principle three ways in which ω_0 can be made as high as possible, by: increasing the unity-gain bandwidth ω_u, decreasing the feedback resistor R_f, and decreasing the total capacitance C_tot. Once these values are known, the feedback capacitance C_f has to be adjusted according to equation (5.17)

$$C_\mathrm{f}^2 \geq \frac{4C_\mathrm{tot}}{\omega_\mathrm{u} R_\mathrm{f}}.$$

Of the three ways of increasing ω_0 not all are equally desirable. To achieve a high signal-to-noise ratio one should have as large a feedback resistor as possible, which limits the choice of operational amplifier and the value of C_tot. Having a broad bandwidth operational amplifier is desirable, but owing to design constrains there is usually a trade-off between low-noise, low-input-current and speed characteristics for operational amplifiers. The few operational amplifiers that come into consideration for measuring very low signal currents all have unity-gain bandwidths of $f_\mathrm{u} = 1\,\mathrm{MHz}$ (compare table 5.1). The unity-gain bandwidth can nevertheless be increased by using more than one amplification stage within the feedback loop [39]. The disadvantage of two amplification stages is the increased noise from the second operational amplifier and the need for a more complicated circuit. Hence, the easiest and most advantageous way of increasing ω_0 is to decrease the input capacitance of the current-to-voltage converter. In the current design, decreasing the input capacitance to 5 pF should result in a bandwidth of 19 kHz. More generally, using the definition of the peak frequency in equation (B.21) as an approximation for the bandwidth, one can say that a reduction by a factor x in total capacitance broadens the bandwidth by $\sqrt{x}$. This can seen below in figure 5.16.

In the current design an optimized input capacitance was not the main focus and higher input capacitance has been traded off with convenient prototyping. A future current-to-voltage converter design should be placed as close as possible to the tip itself and be rigidly mounted. Shielding of the current-to-voltage converter is necessary, but over a short distance the input wire need not be coaxial.

Noise Gain Peaking

Since the operational amplifier is not a broadband amplifier, its effect on the noise gain has to be taken into account. The open-loop gain A is therefore again modeled with a single roll-off corner frequency and unity-gain frequency (ω_u)

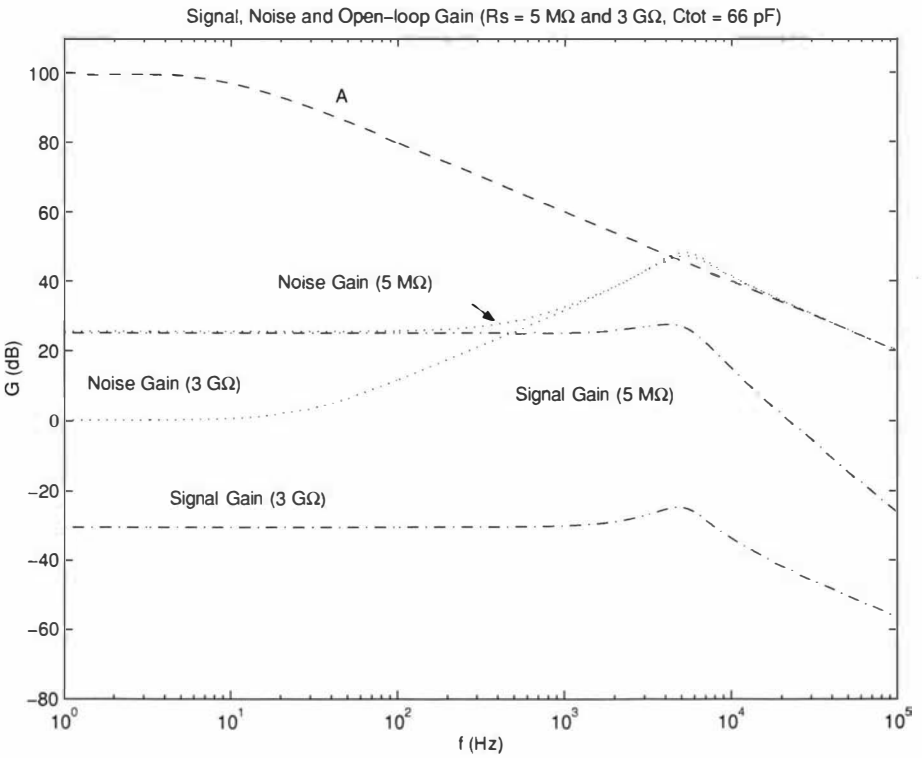


Figure 5.15: **Signal, Noise and Open-loop Gain:** The figure shows the modeled signal and noise gain for the extreme values of tunneling resistance $R_t = 5\,\mathrm{M\Omega}$ and $3\,\mathrm{G\Omega}$. The modeled gain of the transimpedance amplifier is the voltage gain, which behaves equally to the transimpedance. Additionally shown is the top graph, which is the open-loop gain A of the operational amplifier. The signal gain is the lowest graph, which is constant until it reaches a small peak and then rolls off at $-40\,\mathrm{dB/dec}$. The initial value is $G_{dc} = R_f/R_t$, where $R_f = 90\,\mathrm{M\Omega}$. The noise gain is shown in the intermediate graph. Its initial value is $G_{n0} = 1 + R_f/R_t$ and therefore it never falls below 0 dB. It has an inital, rising, corner frequency at $f_{tft} = (2\pi(R_t \parallel R_f)C_{tot})^{-1}$ that depends on R_t and is approximately 28 Hz and 510 Hz. The noise gain reaches a peak when it intercepts with the open-loop gain A at 7.1 kHz. At higher frequencies the noise gain is equal to A. For the total capacitance $C_{tot} = 66\,\mathrm{pF}$ was used. The remaining values used were $R_f = 90\,\mathrm{M\Omega}$, $C_f = 0.25\,\mathrm{pF}$ and $f_u = 1\,\mathrm{MHz}$.

as in equation (5.13). The noise gain is derived in appendix B.2.2 and repeated here as

$$G_n = G_{n0}\frac{1 + j\omega(R_t \parallel R_f)C_{tot}}{1 - \omega^2\frac{R_f C_{tot}}{\omega_u} + j\frac{\omega}{\omega_u}\frac{R_f+R_t}{R_t} + j\omega R_f C_f}, \tag{5.18}$$

Equation (5.18) describes the phenomenon known as noise gain peaking, i.e. an increase in noise gain at higher frequencies followed by a decline when the noise gain intercepts the open-loop gain of the amplifier. The noise peaking is shown graphically in figure 5.15, which also shows the open-loop gain A of the operational amplifier and the signal gain of the current-to-voltage converter.

The influence of the input capacitance is shown in figure 5.16, where $C_{tot} = 66\,\mathrm{pF}$ is assumed as has been calculated previously. A smaller total capacitance of 5 pF shifts both gain curves towards higher frequencies. The influence of the open-loop gain on the signal gain can also be observed in figure 5.16, where the signal gain changes from a peaking curve into a curve with a single corner

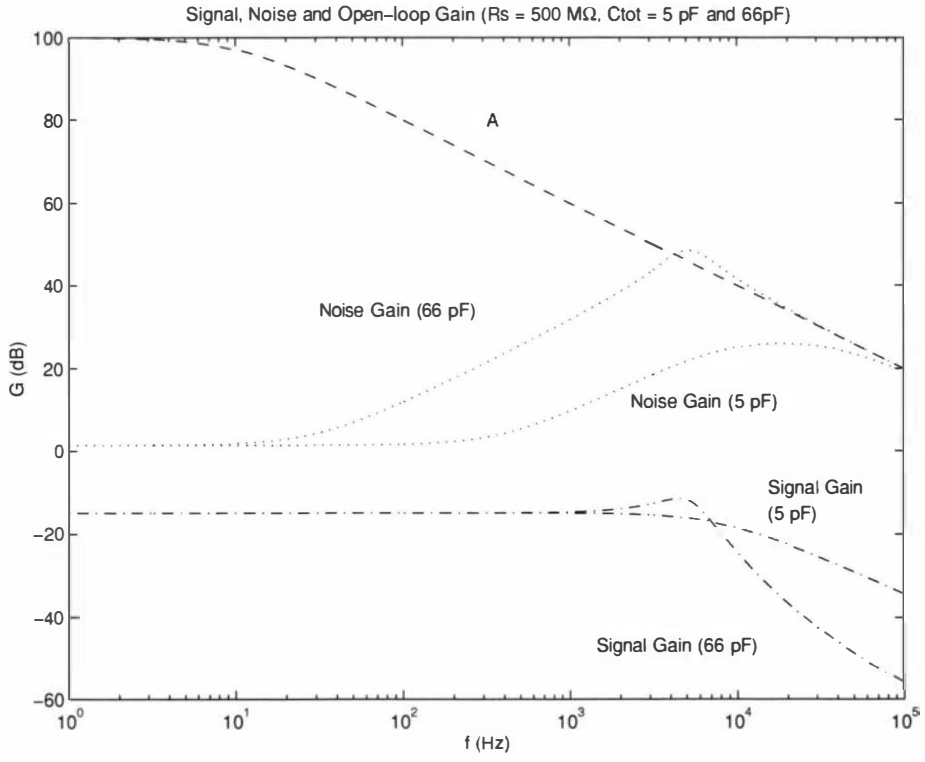


Figure 5.16: **Signal, Noise and Open-loop Gain:** All three modeled gains are shown in the diagram, where the signal gain is the lowest graph, the noise gain the intermediate and the open-loop gain A the envelope. For the input signal of the transimpedance amplifier the voltage gain is shown, which behaves equally to the transimpedance. The graph parameter is the total capacitance, which is $C_{\mathrm{tot}} = 66\,\mathrm{pF}$ for the earlier peaking graphs and $5\,\mathrm{pF}$ for the flatter graphs. The input capacitance is the dominant capacitance in the sum C_{tot}, i.e. we assume $C_{\mathrm{in}} = C_{\mathrm{tot}}$. A smaller capacitance causes the noise gain to rise later and also suppresses the peak in the signal gain, thereby broadening the frequency response of the current-to-voltage converter. As a consequence the signal-to-noise ratio is greatly enhanced. The values used are $R_{\mathrm{t}} = 500\,\mathrm{M\Omega}$, $R_{\mathrm{f}} = 90\,\mathrm{M\Omega}$, $C_{\mathrm{f}} = 0.25\,\mathrm{pF}$ and $f_{\mathrm{u}} = 1\,\mathrm{MHz}$.

frequency. Interestingly the $-3\,\mathrm{dB}$ point is also shifting to higher frequencies although the feedback impedance has not been changed.

The noise gain experiences a dramatic reduction as the input capacitance is decreased, which is perhaps not obvious from the logarithmic graph. The greatest change occurs in the region of the peak.

5.4.3 Future Options

In the current set-up the input capacitance is the largest contributor to the total capacitance. The input capacitance alone limits the bandwidth of the current-to-voltage converter and amplifies high frequency noise. The most important step in a future re-design would be to decrease the input capacitance. A value of 5 pF should be easily obtainable. Once the input capacitance is minimal the use of a larger feedback resistor can be considered, depending on whether bandwidth or low noise is preferred. The bandwidth should be limited to the useful signal bandwidth, which is ω_0. The input channel should be low-pass filtered at higher frequencies. An effective way of doing this would be to add a second operational amplifier into the feedback loop as suggested in [39]. If this is considered to be too much effort, it can also be done with a digital filter in software. The third, but least preferable option would be to use a frequency dependent T-network to decrease the high frequency gain. However a T-network will result in increased low-frequency noise. The T-network is discussed in detail in appendix B.2.7.

A compensation T-network may also be considered to extend the bandwidth of the amplifier. Because it alters the feedback fraction a of the amplifier, the compensation network increases both the signal and the noise bandwidth. This is only useful when the high-frequency level of the noise gain ($G_{\mathrm{nh}} = C_{\mathrm{tot}}/C_{\mathrm{f}}$) is small and the corner frequency ω_{f} is very low, i.e. for a very high feedback resistor (for instance $1\,\mathrm{G\Omega}$) and small input capacitance ($C_{\mathrm{in}} < 5\,\mathrm{pF}$). As soon as the noise gain intercepts the open-loop gain on a rising slope the current-to-voltage converter will oscillate, which means that the compensation network can only extend the bandwidth until that point is reached. This could be as low as 10 kHz depending on the input capacitance and feedback resistance.

When acting as a closed-loop controller the STM system requires a broad bandwidth. If by some means the signal bandwidth can be extended up to an acceptable frequency then both the signal and the noise of the current-to-voltage converter can be rolled off at that frequency and the greater noise bandwidth of the current-to-voltage converter would not play a role anymore. This can be achieved by equipping the current-to-voltage converter with a small feedback resistance and following the current-to-voltage converter with a second bandwidth-limited gain stage. The small feedback resistance decreases the low-frequency signal-to-noise ratio, but the two stages together suppress the high frequency noise. As long as the input capacitance is small this would be one way to attain reasonable bandwidth and signal-to-noise ratio.

5.4.4 IVC Module Noise

To find the overall noise contributed by the current-to-voltage converter one has to add amplifier voltage noise, the Johnson noise of the feedback resistor and the amplifier input current noise. The amplifier voltage noise is calculated in appendix B.2.5. The Johnson noise of the feedback resistor is somewhat

smaller than the voltage noise. By comparison, the input current noise is mostly negligible.

The current-to-voltage converter noise can then be compared with the intrinsic noise of the input signal. This intrinsic noise is the tunneling shot noise. The current-to-voltage converter noise should be comparable to or less than this.

The Johnson noise density is given by

$$e_r = \sqrt{4kTR_f} = 1.22 \times 10^{-6}\,\mathrm{V}/\sqrt{\mathrm{Hz}},$$

where $k = 1.38 \times 10^{-23}\,\mathrm{J/K}$ is Boltzmann's constant, $T = 300\,\mathrm{K}$ the absolute temperature and $R_f = 90\,\mathrm{M\Omega}$ the feedback resistor. This noise density must be integrated over all frequencies taking into account that the gain rolls off at the corner frequency ω_f of the feedback impedance. One gets

$$V_{nr} = e_r\sqrt{\frac{\pi}{2}f_f} = 129\,\mu\mathrm{V}.$$

Interestingly this value is not dependent on the feedback resistor, because the resistance term in the Johnson noise cancels with the resistance determining the feedback corner frequency, seen in[10]

$$\sqrt{4kTR_f} \times \sqrt{\frac{\pi}{2}\frac{1}{2\pi R_f C_f}} = \sqrt{\frac{kT}{C_f}}.$$

This is true for large resistances where the corner frequency, determined by the parasitic capacitance, is lower than the bandwidth of the following stages.

The shot noise density of the tunneling current is

$$i_n = \sqrt{2eI_t},$$

where $e = 1.6 \times 10^{-19}\,\mathrm{C}$ is the electron charge and I_t the dc part of the tunneling current. The shot noise depends on the actual value of tunneling current. The smallest expected value for the shot noise density is $6.3\,\mathrm{fA}/\sqrt{\mathrm{Hz}}$ with $I_t = 0.1\,\mathrm{nA}$ and the biggest is $63\,\mathrm{fA}/\sqrt{\mathrm{Hz}}$ with $I_t = 10\,\mathrm{nA}$. The resulting output noise voltages are $V_{sn} = 60\,\mu\mathrm{V}$ and $600\,\mu\mathrm{V}$. For the bandwidth we took the same $\sqrt{(\pi/2)f_f} = 105\sqrt{\mathrm{Hz}}$ as above for the feedback resistor noise. The input current noise of the amplifier is $0.22\,\mathrm{fA}/\sqrt{\mathrm{Hz}}$, i.e. much smaller than the smallest expected tunneling current shot noise and may therefore be neglected. The overall noise of the current-to-voltage converter, including shot noise, is summed up as

$$V_{nivc} = \sqrt{V_{na}^2 + V_{nr}^2 + V_{sn}^2}. \tag{5.19}$$

A graphical representation of the individual current-to-voltage converter noise contributions can be seen in figure 5.17. The s/n does increase with a smaller input capacitance. This is shown in figure 5.18. With only 5 pF as input capacitance the current-to-voltage converter noise performance is greatly improved. Ideally either the intrinsic noise of the source, the shot noise, or another unavoidable type of noise should be the major contributors to the overall noise. If

[10] In Simpson's book [171] this formula is part of the derivation of Johnson noise.

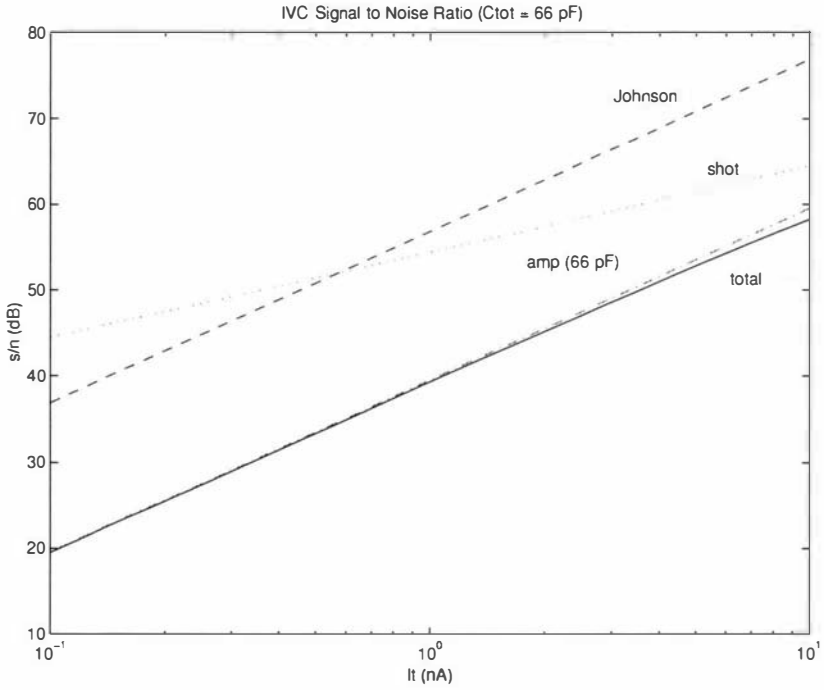


Figure 5.17: **Signal-to-Noise-Ratio of Current-to-Voltage Converter (66 pF):** The figure shows three individual signal-to-noise ratios (s/n) as a function of the tunneling current. One for the amplifier noise voltage, another for the feedback resistor Johnson noise and yet another for the tunneling current shot noise. The three s/n 'add' to give the total signal-to-noise ratio shown as the lower envelope. One sees immediately that the amplifier noise determines the s/n of the current-to-voltage converter. The s/n due to the amplifier voltage noise is hardly distinguishable from the total s/n and is mainly dependent on the input capacitance of the current-to-voltage converter, which is assumed to be $C_{\mathrm{tot}} = 66\,\mathrm{pF}$. Without the amplifier noise, the s/n due to feedback resistor Johnson noise would dominate for tunneling currents smaller than $0.6\,\mathrm{nA}$, while the tunneling shot noise would dominate at tunneling currents larger than $0.6\,\mathrm{nA}$. Values used in calculations were $V_{\mathrm{bias}} = 50\,\mathrm{mV}$, $R_{\mathrm{f}} = 90\,\mathrm{M\Omega}$ and $C_{\mathrm{f}} = 0.25\,\mathrm{pF}$. The s/n rises from $20\,\mathrm{dB}$ at small input currents ($0.1\,\mathrm{nA}$) to almost $60\,\mathrm{dB}$ at high input currents ($10\,\mathrm{nA}$). The shot noise s/n rises slower than the other s/n, since it grows with the input current, whereas the other noise contributions are either constant or do not vary considerably. The influence of the parameter bias voltage is not shown, since the s/n changes only marginally with bias voltage, noticeable only at high input currents.

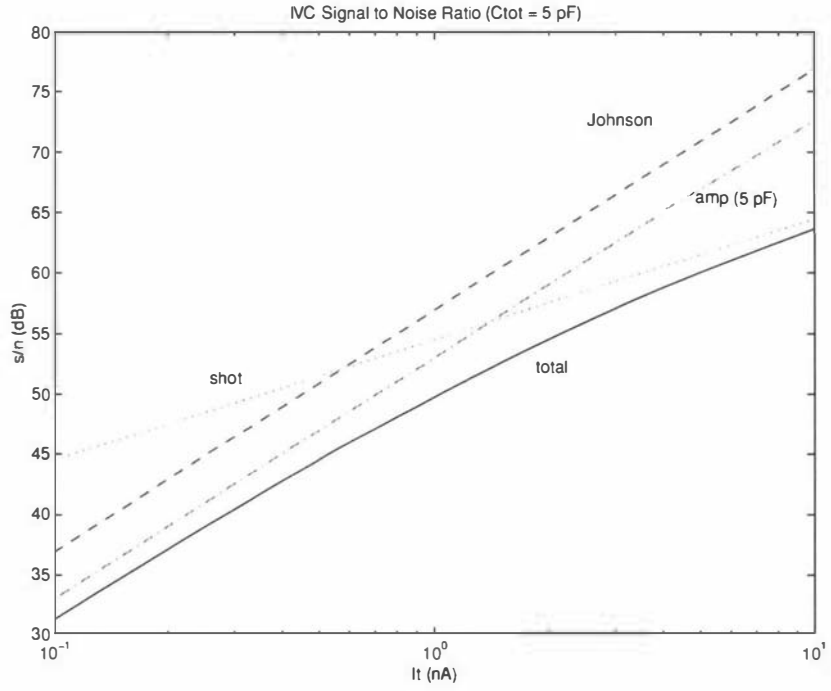


Figure 5.18: **Signal-to-Noise-Ratio of Current-to-Voltage Converter (5 pF):** The same curves as in figure 5.17 except that the amplifier noise voltage has been reduced by 10 dB due to an assumed smaller input capacitance of $C_{\rm in} = 5\,{\rm pF}$. Now the amplifier noise determines the s/n of the current-to-voltage converter only at smaller tunneling currents up to about $1.5\,{\rm nA}$. Above that the tunneling shot noise dominates. The current-to-voltage converter with $5\,{\rm pF}$ input capacitance comes closer to the ideal amplifier than the one used at the moment. The s/n rises from $31\,{\rm dB}$ at small input currents ($0.1\,{\rm nA}$) to $64\,{\rm dB}$ at high input currents ($10\,{\rm nA}$).

the problem of limited bandwidth is addressed, the signal-to-noise ratio of the feedback resistor can be improved by a factor of

$$20 \log 10^{\frac{1}{2}} \text{ dB} = 10 \text{ dB}$$

by using a 10 times greater feedback resistor. Although a larger resistance generates more Johnson noise, the current signal is turned into an even greater voltage. Judging from figure 5.18 this would not yet be a noticeable improvement unless the amplifier voltage noise can be reduced even further.

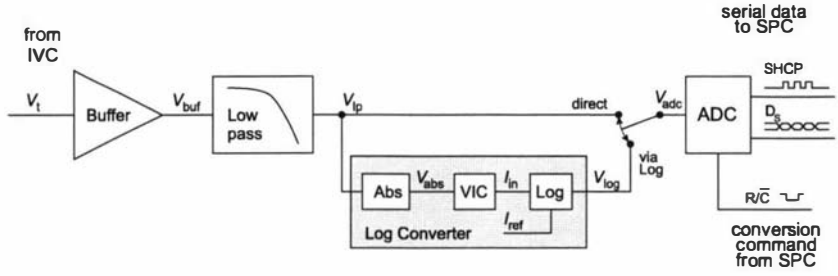


Figure 5.19: **ADC Module:** The ADC module is part of the input channel. The input voltage V_t from the current-to-voltage converter (IVC) is buffered, low-pass filtered and digitized by the analog-to-digital converter (ADC). The filtered voltage can also be logarithmically converted. The logarithmic converter (Log) is preceded by an absolute value converter (Abs) and a voltage-to-current converter (VIC). The serial data D_S are sent to the serial-to-parallel converter (SPC) in the DSP system interface, accompanied by a synchronization pulse train SHCP. The ADC receives conversion commands $R/\overline{C}$ from the SPC in regular intervals.

5.4.5 ADC module

The ADC module receives its input from the current-to-voltage converter, which is the very first stage of the input signal channel. The current-to-voltage converter is discussed in section 5.4.2. The ADC module digitizes the converted tunneling current and makes the digital word available to the DSP. The specifications for the ADC module are derived from those given for the input channel, as discussed in section 5.1.2. These were to keep the noise level below 1 pA and to provide a current range of 10 nA. With a current-to-voltage conversion factor of 10^9 V/A this translates to a noise level of less than 1 mV and a range of 10 V. The bandwidth should be appropriate for the planned sampling interval of 10 μs.

To obtain a digital number that is of use to the DSP, the input voltage passes through several processing stages: the voltage V_t is first buffered and then passed through an anti-aliasing filter. The signal is subsequently either converted by a logarithmic converter or fed directly into an analog-to-digital converter, where it is finally digitized. A diagram is shown in figure 5.19. The digitized value is then buffered and transmitted to the DSP system interface. Apart from the analog-to-digital converter and a logic buffer, the ADC module contains only analog components. Each of these sub-modules will be discussed in more detail in the following paragraphs.

Regardless of whether the direct, or the logarithmic path is selected, the input voltage V_{adc} to the analog-to-digital converter is turned into a number N_{adc}

$$N_{\mathrm{adc}} = \frac{V_{\mathrm{adc}}}{k},$$

where the resolution of the analog-to-digital converter is

$$k = \frac{20\,\mathrm{V}}{2^{16}} = 305\,\mu\mathrm{V}.$$

When not using the logarithmic converter, the overall formula for the ADC module is

$$N_{\mathrm{adc}} = \frac{G_{\mathrm{lp}} V_t}{k},$$

where $G_{\mathrm{lp}} = 1.95$ is the dc gain of the low-pass filter. The software represents the voltage (V_{t}), from the current-to-voltage converter in floating point format and uses

$$k_{\mathrm{adc}} = \frac{k}{G_{\mathrm{lp}}} = \frac{305\,\mu\mathrm{V}}{1.95} = 156\,\mu\mathrm{V}$$

to calculate the input voltage V_{adc} of the ADC module from the input number N_{adc}. The gain of the low-pass filter is therefore absorbed by the conversion factor, so that the software voltages agree with the voltages measured on a monitoring oscilloscope that receives the same signals as the ADC module. For the calculations which follow, voltages at the input of the analog-to-digital converter are also given as equivalent voltages at the input of the module. The factor k_{adc} is the effective resolution of the ADC module.

Power supply The ADC module has three power supply connections, $\pm V_{\mathrm{s1}}$ for the purely analog part and V_{s2}, from which the digital supply for the logic driver and the analog-to-digital converter is derived. To keep the amount of noise and interference that is coupled into the signal path as low as possible, the low-level and high-level ground returns are kept physically separate on the board. This is discussed in section 5.1.8. Only two Schottky diodes interconnect these grounds in case one of the grounds gets disconnected. The low-level ground is laid out as a star connection, which results in a minimum of resistively coupled noise [19, 135]. The high-level ground forms, as closely as possible, a ground plane. This way the ground impedance is minimal, which is important for digital signals [19, 100]. The main connection between the center of the star and module connector is done via the voltage regulator.

Analog-to-Digital Converter

The analog-to-digital converter is an ADS7809P from Burr Brown. Selection was based on dynamic range, conversion speed and appropriate digital interfacing. The ADS7809P is a sampling, 16-bit converter with 100 kHz maximum sampling rate, and serial data interface. Additional information on the ADS7809P can be found in the Burr Brown Application Bulletins AB-178, AB-084 and AB-95 [39].

Operating Principle The ADS7809P samples the input voltage with an built-in sample-and-hold amplifier. The input voltage is then digitized using successive approximation, which is carried out with the help of a switched capacitor network[11] explained below. The digital word is transferred to an internal register within the ADS7809P. An external read command initiates serial data transmission out of the register.

The switched capacitor network used for successive approximation is based upon charge transfer. The operating principle is explained in the handbook of Burr Brown [39]. Sedra and Smith [170] also explain a charge-redistribution converter in their book, but the capacitor network has a slightly different architecture to that of the ADS7809P. A schematic diagram of the input circuit of the ADS7809P is shown in figure 5.20. There is an array of 16 capacitors, one for each bit of the converter. One side of each capacitor is connected to a comparator. This connection can be grounded via a switch (S_g). The other

[11] Capacitor digital-to-analog converter or 'CDAC'

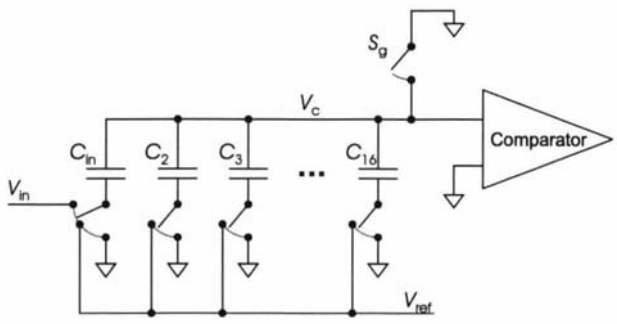


Figure 5.20: **Switched Capacitor Network:** The input circuit of the ADS7809P analog-to-digital converter uses capacitors instead of resistors to digitize the input voltage by successive approximation. There are 16 capacitors ($C_{\rm in}$, $C_2 \ldots C_{16}$), which can be connected to either ground or $V_{\rm ref}$. The size of the capacitors doubles from right to left, i.e. $C_{i-1} = 2 \times C_i$ for $i = 2, \ldots, 16$. On acquisition the first or input capacitor is connected to the input voltage, while all other capacitors are grounded and the comparator is grounded via the switch S_g. A charge that is proportional to the input voltage is sitting on the input capacitor. To convert the voltage, switch S_g is opened, which traps the charge on the input capacitor. The voltage at the comparator input is now $V_c = -V_{\rm in}/2$. Subsequently the capacitor switches are opened and closed in turn to achieve a re-distribution of charges that results in $V_c = 0$ (Diagram adopted from Burr Brown [39]).

side of each capacitor can be connected to either ground or a reference voltage ($V_{\rm ref}$). The first capacitor is special in that it doubles as the input capacitor ($C_1 = C_{\rm in}$) and can be connected to ground, the reference voltage ($V_{\rm ref}$), or the input voltage ($V_{\rm in}$). There are three phases of acquisition, the sample phase, the hold phase and the charge re-distribution phase.

In the sample phase the input capacitor is connected to the input voltage, while the comparator input is grounded. All other capacitors are connected to ground. The resulting charge on the input capacitor is

$$Q_{\rm in} = C_{\rm in} V_{\rm in}.$$

The next step is the hold phase, where the ground switch S_g is opened and the input switch is connected to ground like all other capacitors. As a result, the charge $Q_{\rm in}$ is trapped on the comparator capacitors. This way, a sample of the input voltage has been taken, since this charge is proportional to the input voltage. The voltage on the comparator input is then

$$V_c({\rm in}) = -\frac{Q_{\rm in}}{\sum_{\rm all} C_i},$$

where $\sum_{\rm all} C_i$ is the sum of all capacitors in the array. Note that the comparator voltage is negative, since during the sampling phase opposite charges were attracted on the capacitor plates of the input capacitor. All these capacitances are related to each other in a binary fashion as $C_2 = C_{\rm in}/2$, $C_3 = C_{\rm in}/4$, which can be summarized as

$$C_i = \frac{C_{\rm in}}{2^{i-1}} \quad \text{for} \quad i = 1, \ldots, 16\,.$$

The sum of all capacitors can now be calculated and is found to be

$$\sum_{\rm all} C_i = \left(1 + \frac{1}{2} + \frac{1}{4} + \ldots + \frac{1}{2^{15}}\right) C_{\rm in} \approx 2C_{\rm in}.$$

As explained in the book of Sedra and Smith [170], an extra terminating capacitor with a capacitance of $C_{\rm in}/2^{n-1}$ can be added for an n-bit converter to make the last approximation an equality. For a high-resolution converter the approximation is already very good even without the terminating capacitor. Note that the comparator input voltage is

$$V_{\rm c}({\rm in}) = -\frac{C_{\rm in}V_{\rm in}}{2C_{\rm in}} = -\frac{V_{\rm in}}{2}\,.$$

The last step of acquisition is the successive approximation itself. It is the charge re-distribution phase, where some of the capacitors are connected to $V_{\rm ref}$, while the remaining capacitors are connected to ground. The capacitors that are connected to $V_{\rm ref}$ attract the charges originally trapped on the comparator input connector. The aim is to balance the charge and zero the comparator input voltage to within ±1 LSB of the converter. During the successive approximation each capacitor switch is tested in turn. Connecting the first capacitor to the reference voltage adds a voltage increment of $V_{\rm ref}/2$ to the comparator input voltage ($V_{\rm c}$), since the comparator input sees the capacitor network as a capacitive divider with equal capacitance $C_{\rm in}$. The top half of the divider is given by the input capacitance $C_{\rm in}$, while the bottom half is given by the parallel capacitance of the remainder of the capacitor network, which adds up to the same capacitance. The other capacitors add smaller and smaller voltage increments, which are

$$V_{\rm c}({\rm after}) = V_{\rm c}({\rm before}) + \frac{V_{\rm ref}}{2^i} \quad \text{for} \quad i = 1, \ldots, 16\,.$$

The voltage on the comparator input generated by connecting $V_{\rm ref}$ to some of the capacitors is

$$V_{\rm c}({\rm ref}) = \frac{\sum_{\rm ref} Q_i}{\sum_{\rm all} C_i} = \frac{\sum_{\rm ref} C_i}{2C_{\rm in}} V_{\rm ref} = \sum_{\rm ref} 2^{-(i-1)}\frac{V_{\rm ref}}{2}$$

where the index 'ref' in $\sum_{\rm ref}$ refers to the capacitors that are connected to $V_{\rm ref}$. When the charges are balanced, the comparator input voltage is zero and

$$\frac{V_{\rm in}}{2} = \sum_{\rm ref} 2^{-(i-1)}\frac{V_{\rm ref}}{2}$$

from which the input voltage is found to be

$$V_{\rm in} = \sum_{\rm ref} 2^{-(i-1)} V_{\rm ref} \quad \text{for} \quad i = 1, \ldots, 16\,.$$

The reference voltage is known and the analog-to-digital converter transmits the equivalent digital word given by the bit-number of all the capacitors in the sum $\sum_{\rm ref}$.

Connections The ADS7809P has a resistor divider input network allowing different input ranges to be selected. The converter is used with an input range of ±10 V without extra components for gain and offset trim. Connections to the resistor divider input of the converter are made according to the data sheet [40]. The ADS7809P incorporates an internal voltage reference, which has been used in preference to an external reference.

To prevent switching transients being coupled via the power supply lines the analog supply was decoupled with a 100 μF electrolytic, a 10 μF electrolytic tantalum and a 0.1 μF ceramic capacitor. The digital supply voltage is also well decoupled.

Digital Interface and Timing The ADS7809P has a serial data output, which can either be synchronized to an internal clock or synchronized to an external clock fed into the ADS7809P. The current design uses the internal clock for reasons explained below. The internal clock runs at a frequency of 2.3 MHz and the 16-bit output data are accompanied by a pulse train of 16 pulses (SHCP). The period of the internal clock is 440 ns and the overall conversion time is $16 \times 440\,\text{ns} = 7\,\mu\text{s}$.

There are three serial links to the ADC module, the serial output data D_S, the pulse train SHCP and the conversion command $R/\overline{C}$. The serial data D_S are synchronized to SHCP. The ADC module receives the conversion command externally from the serial-to-parallel converter in a 12.8 μs interval. The chip select of the ADS7809P is constantly active. Digital output is in twos complement format. The two digital output lines are buffered with a 74LS125 driver.

Although an external system clock was initially used, the design was changed to use the internal clock, since it offers several major advantages. Using the internal clock of the ADS7809P simplifies the logic design of the interface between the analog-to-digital converter and the serial-to-parallel converter, resulting in fewer logic components and therefore an easier layout and less digital noise. On the serial-to-parallel converter side there is no need to generate a pulse train, since the data are already synchronized to the pulse train from the analog-to-digital converter. This pulse train can be fed directly into the shift register. There is also no need to synchronize the data to the system clock. On the ADC module no circuitry is needed to generate the chip select at the right time. With an external clock it was not possible to just leave the chip select low. The generation of the conversion command at certain intervals is done in the serial-to-parallel converter instead of the ADC module, which has the advantage of minimizing digital noise in the ADC module.

Signal Buffer

To ensure the least attenuation of the signal voltage coming from the transimpedance amplifier the input impedance of the ADC module is kept as high as possible. A non-inverting amplifier is placed at the input of the ADC module and acts as a signal buffer.

Again, the OP27GP has been used. According to the manufacturer's recommendation the follower is used with a 1 kΩ resistor in the feedback path to limit input currents under transient conditions. As with the output driver of the IVC module, the signal buffer introduces an offset error of 180 μV.

Low-Pass Filter

The next step in the input signal path is a low-pass filter. It serves as an anti-aliasing filter for the analog-to-digital converter. The cut-off frequency is $f_c = 50\,\text{kHz}$. The filter design is discussed in appendix B.1.5. The function of the low-pass filter is to prevent unwanted high-frequency signals from being

folded back into the useful signal band and being seen by the analog-to-digital converter as an extra signal (for more details see section 4.4). Since the current-to-voltage converter bandwidth at the moment is much lower than the Nyquist frequency of the analog-to-digital converter, limiting the input bandwidth of the converter primarily serves to reduce folded-back input noise. However, in the future, the bandwidth of the current-to-voltage converter amplifier might be increased and the ADC module must cater for this possibility. The ADC module is also designed to be more generally useful and should therefore have a relatively broad bandwidth. The conversion time is currently set to $t_{adc} = 12.8\,\mu s$, which is slightly more than the possible $10\,\mu s$. The Nyquist frequency for that conversion time is

$$f_N = \frac{1}{2t_{adc}} = 39.1\,\text{kHz}.$$

Although this is the case, the low-pass filter is still set to 50 kHz, following the original design idea. The slight increase in bandwidth will only very slightly increase the noise at the converter input, since the maximum signal frequency is much less than f_N. With $f_c = 50\,\text{kHz}$ the ADC module as such is laid out to accommodate the maximum sampling rate of the analog-to-digital converter. To increase the control bandwidth the conversion time should be decreased back to its guaranteed minimum of $10\,\mu s$.

Logarithmic Conversion

The logarithm of any physical quantity can only be taken if this quantity is positive and dimensionless. To perform the logarithmic conversion in the ADC module the actual logarithmic converter is preceded by a signal rectifier and a transconductance amplifier as seen in figure 5.19. The logarithmic converter outputs a voltage proportional to the ratio between its input current and a reference current. In appendix B.3 the circuits belonging to the logarithmic conversion channel are discussed in detail, while here only an overview is given.

The rectifier delivers the absolute value V_{abs} of the incoming voltage from the low-pass filter V_{lp},

$$V_{abs} = |V_{lp}|.$$

The transconductance amplifier (VIC) converts V_{abs} into a corresponding current (I_{in}):

$$I_{in} = A_{vic} V_{abs},$$

where $A_{vic} = 10^{-5}\,\text{A/V}$ is a constant. Finally the output of the logarithmic converter will be proportional to the logarithm of the ratio I_{in}/I_{ref},

$$V_{log} = -V_{log}^{0} \ln\left(\frac{I_{in}}{I_{ref}}\right),$$

with the constants $V_{log}^{0} = 2.94\,\text{V}$ and $I_{ref} = 3.35\,\mu\text{A}$. The reference current I_{ref} is generated by a current source. Using logarithmic conversion, the overall formula for the ADC module is

$$N_{adc} = -\frac{V_{log}^{0}}{k} \ln\left(\frac{A_{vic} G_{lp} V_t}{I_{ref}}\right).$$

Speed

A 4-pole Bessel low-pass filter is the main bandwidth limiting stage of the ADC module. The filter has a step rise time of 10 μs. The analog-to-digital converter needs a maximum of 2 μs to acquire the latest sampled datum and a maximum of 8 μs to convert this datum into a stream of digital serial output data. If the conversion times are taken as fixed, then it takes at least 20 μs for the ADC module to record the 90% level of an input voltage change. In the current system the conversion command is repeated only every 12.8 μs instead of the minimum 10 μs. The timing of the converter is such that it converts the previously acquired number first, then samples and acquires the next datum. This means that with the current system there is a 2.8 μs period between acquisition and conversion where the converter is idle. The fastest possible recording time is therefore 10 μs + 12.8 μs = 22.8 μs in the current system. The worst case would be that the acquisition cycle just ended before the voltage to be recorded was changed. In this case the change is recorded only after another sampling interval has passed, i.e. after $2 \times 12.8\,\mu$s = 25.6 μs in total. On average the time to record the 90% level of any voltage change is 24.2 μs, which is 2 sampling intervals for the program.

Reading from the converter can be synchronized by setting a hardware flag on the DSP system interface after a new datum has arrived. The load command for the analog-to-digital converter should then be started by the DSP and not the serial-to-parallel converter, although the timing circuitry may remain on the serial-to-parallel converter. This solution is similar to the one used with the parallel-data register. For an input signal that occurs at a random instant, the gain in speed is only 2.8 μs which is negligible. In a closed-loop system however, the program dead time can be minimized if the program knows exactly when a new datum has arrived.

Noise

The noise produced by the ADC module consists of noise contributed by the analog signal processing electronics and noise contributed by the analog-to-digital converter, which is of two types: quantization and transition noise. The quantization process can be interpreted as adding a certain amount of white noise to the incoming signal and is discussed in section 4.3.3. Transition noise is an extra noise that occurs in high-resolution analog-to-digital converters [39]. The transition noise is larger than the quantization noise. Hence, for a 16-bit analog-to-digital converter one has to look at the transition noise specification rather than the quantization noise, when deciding on a specific converter.

The noise from the remaining electronics (V_{nel}) in the ADC module is calculated by combining the individual noise contributions from each part of the module. The noise of the different stages was calculated using table B.1 in appendix B.1.1.

$$\begin{aligned} V_{\mathrm{nbo}} &= G_{\mathrm{lp2}} G_{\mathrm{lp1}} V_{\mathrm{nin}} = 1.8 \times 1.084 \times 1.2\,\mu\mathrm{V} = 2\,\mu\mathrm{V} \\ V_{\mathrm{n1o}} &= G_{\mathrm{lp2}} G_{\mathrm{lp1}} V_{\mathrm{nb}} = 1.8 \times 1.084 \times 15.7\,\mu\mathrm{V} = 31\,\mu\mathrm{V} \\ V_{\mathrm{n2o}} &= G_{\mathrm{lp2}} V_{\mathrm{n11}} = 1.084 \times 9.2\,\mu\mathrm{V} = 10\,\mu\mathrm{V}, \end{aligned}$$

where V_{nbo} is the output noise voltages of the buffer and V_{n1o} and V_{n2o} are the output noise voltages of the two low-pass filter stages. The buffer does not con-

tribute any significant noise and can be disregarded. The greatest contributor to noise is the first low-pass filter stage and this is due to the comparatively high resistances at the input to that stage and the gain of the second low-pass filter stage. If this resistor were reduced by a factor of ten the noise should decrease by a factor of $\sqrt{10}$. Disregarding the buffer, the resulting noise of the contributions listed above is

$$V_{\text{nel}} = \sqrt{31^2 + 10^2}\,\mu\text{V} = 33\,\mu\text{V}.$$

Reflected back to the input of the ADC module we get an equivalent input noise of

$$V_{\text{nel}} = \frac{1}{1.95} \times 33\,\mu\text{V} = 17\,\mu\text{V}.$$

In equation (4.8) the quantization noise is given as

$$V_{\text{q}} = \frac{k_{\text{adc}}}{\sqrt{12}} = \frac{156\,\mu\text{V}}{3.46} = 45\,\mu\text{V},$$

where k_{adc} is the resolution of the analog-to-digital converter. This value already corresponds to that found at the input of the ADC module. A typical value for the transition noise of the analog-to-digital converter is stated in the data sheet [40] as

$$V_{\text{tr}} = 1.3 \times k_{\text{adc}} = 1.3 \times 156\,\mu\text{V} = 203\,\mu\text{V},$$

with the same k_{adc} as above.

Combining these results, the ADC-module noise has an expected value of

$$\begin{aligned} V_{\text{nadc}} &= \sqrt{V_{\text{tr}}^2 + V_{\text{q}}^2 + V_{\text{nel}}^2} \\ &= \sqrt{203^2 + 44^2 + 17^2} = 208\,\mu\text{V}. \end{aligned}$$

The biggest contribution is the transition noise of the analog-to-digital converter. The quantization noise stays well below that level. Judging from how little the component noise adds to the transition noise of the analog-to-digital converter one could use less expensive components without significantly increasing the total noise.

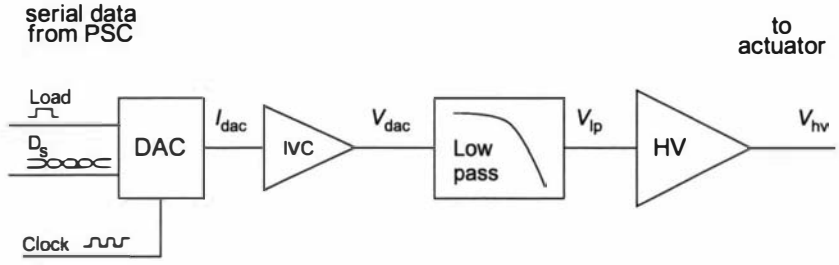


Figure 5.21: **HV Module:** The HV module consists of a digital-to-analog converter (DAC), current-to-voltage converter (IVC), low-pass filter and high-voltage amplifier (HV). The DAC receives serial data from a parallel-to-serial converter (PSC) in the DSP system interface. The low-pass filter is a reconstruction filter for the digitized output V_{dac} from the DAC. The output voltage V_{hv} leads to the probe head actuators.

5.4.6 HV module

The High-Voltage (HV) modules generate and condition analog control voltages for the piezoelectric actuators according to the design specifications in section 5.1.4, equation (5.3). The voltages have a range of ±190 V and should have a noise level less than 300 μV. The bandwidth of the high-voltage signals is 100 kHz. Each HV module receives digital signals from the DSP and delivers a high-voltage signal as its output to the probe head. There are in total four HV modules, each unit carrying two separate channels. All HV modules are identical and each channel can be divided into three functionally different parts, a digital-to-analog-converter plus transimpedance amplifier, a low pass filter and a high voltage operational amplifier. A diagram is shown in figure 5.21.

Digital-to-Analog Converter

Choosing a suitable digital-to-analog converter is crucial for the design of the HV module. The converter has to have a large dynamic range and a fast conversion rate, while accepting serial data as input. With the design goals of a 380 V total voltage range and a 300 μV resolution, the converter would need a dynamic range of better than 122 dB.

This is not easily achievable with a single digital-to-analog converter since a resolution of at least 20 bits is required for a dynamic range of 120 dB. Another option that can be used to achieve a wide range with good resolution is to use two digital-to-analog converters of lower resolution and a summing amplifier. The two converters can be used to produce signals with coarse and fine resolution which can then be combined by the summing amplifier. However, in this design we nevertheless opted to use a single high-resolution digital-to-analog converter, since high performance digital-to-analog converters are available and the design with a single digital-to-analog converter is simpler. A certain loss of resolution in the vertical tip movement was accepted since a reasonable horizontal movement had to be achieved with the same modules. The first calculations were based on the horizontal and vertical sensitivities from Binnig and Smith [29] and Besocke [25]. Since these sensitivities were about four times smaller than the actual horizontal sensitivity of the center actuator in this probe head, the available scan area of this microscope is larger than originally planned. Accordingly four times smaller supply voltages can be used in the future. A resolution

of 300 μV can be achieved while still having a reasonable scan area. The horizontal and vertical resolution is compared to other noise sources in the system in section 7.5.2.

For the HV module a 20-bit digital-to-analog converter was chosen. Judging from the data sheet [8] a dynamic range of 18 bits, i.e. 108 dB, can be achieved. An overall system signal-to-noise ratio of 108 dB is not unrealistic, but low noise components have to be used along with a very careful system layout and signal routing. Good power supply decoupling is also essential. The typical signal-to-noise ratio[12] of the AD1862N is stated as 113 dB, which makes a signal-to-noise ratio of 108 dB just possible.

The overall weight (k_{hv}) of the least significant bit is

$$k_{\mathrm{hv}} = \frac{380\,\mathrm{V}}{2^{20}} = \frac{380\,\mathrm{V}}{1.05 \times 10^{6}} = 360\,\mu\mathrm{V}. \tag{5.20}$$

The converter is an R-2R-ladder type AD1862N from Analog Devices [8]. Digital inputs to the AD1862N are the clock signal Clock, the control signal Load and the serial data D_s. The AD1862N has a maximum clock speed of 17 MHz and a serial data interface. The minimum conversion time is $20/(17\times10^{6}\,1/\mathrm{s}) = 1.18\,\mu$s, which is more than eight times faster than the input signal conversion. The actual conversion time, determined by the period of clock CLK, is $20 \times 200\,\mathrm{ns} = 4\,\mu$s. The analog output of the AD1862N is a current in the range of ±1 mA. The offset error of the AD1862N is $\pm5\,\mu$A, the gain error is 2%. Following recommendations of the AD1862N data sheet [8], two noise reduction capacitors have been connected to the digital-to-analog converter.

Transimpedance Amplifier

The digital-to-analog converter converts the incoming serial data (D_s) into a current (I_{dac}) corresponding to its numerical value.

$$I_{\mathrm{dac}} = k_{\mathrm{dac}} N_{\mathrm{hv}}.$$

where k_{dac} is the weight of the least significant bit. On the HV module an external operational amplifier turns the current I_{dac} into a corresponding voltage (V_{dac}) using the high-precision 3 kΩ on-chip feedback resistor of the AD1862N. The operational amplifier is wired in the transimpedance configuration. The output voltage is therefore

$$V_{\mathrm{dac}} = -R_{\mathrm{tr}} I_{\mathrm{dac}},$$

where R_{tr} the trans-resistance of the transimpedance amplifier. The maximum output voltage of the digital-to-analog converter plus external amplifier is ±3 V. The external amplifier is an OP27 low-noise operational amplifier. A bandwidth limiting capacitor of 470 pF is put across the amplifier feedback resistance, resulting in a cut-off frequency of 113 kHz.

Low Pass Filter

The low-pass filter is a two-pole Bessel filter with a cut-off frequency of 100 kHz, which removes spurious high-frequency components from V_{dac}. The filter is of

[12] The signal-to-noise ratio is defined in relation to a full scale sine wave within a bandwidth of 30 kHz.

the Sallen-Key type; the design is discussed in appendix B.1.5, while a circuit diagram is shown in figure B.4. An active filter adds amplifier noise to the output signals, so an alternative would be to use a passive low pass filter. It is also possible to use the high-frequency roll-off of the high-voltage amplifiers as the only low-pass filter, as was done by Clark *et al.* [46]. In this case the output signal (V_{hv}) will contain more digital high-frequency noise, because the filter has a gentler high frequency roll-off. The noise due to the low-pass filter actually used will be discussed shortly.

Since the filter has a dc gain (G_{lp}), its output voltage at low frequencies is

$$V_{lp} = G_{lp} V_{dac},$$

where $G_{lp} = 1.27$. With an input range of $\pm 3\,\mathrm{V}$ from the digital-to-analog converter, the output range of the low pass filter is $\pm 4\,\mathrm{V}$. The first-order errors for the low-pass amplifier are $2\,\mathrm{mV}$ due to bias current and $1.3 \times 250\,\mu\mathrm{V} = 330\,\mu\mathrm{V}$ due to offset voltage. The offset of the low-pass filter is amplified by the high-voltage amplifier to $2.3\,\mathrm{mV} \times 47 = 0.11\,\mathrm{V}$. The relative gain error is 0.34%.

High-Voltage Amplifiers The HV amplifiers used in this project are PA85s from Apex. They are commonly used to control piezoelectric actuators since they combine high output voltage and current with a relatively low noise, and are fast enough for this type of application [12]. A circuit diagram is shown in figure 5.22.

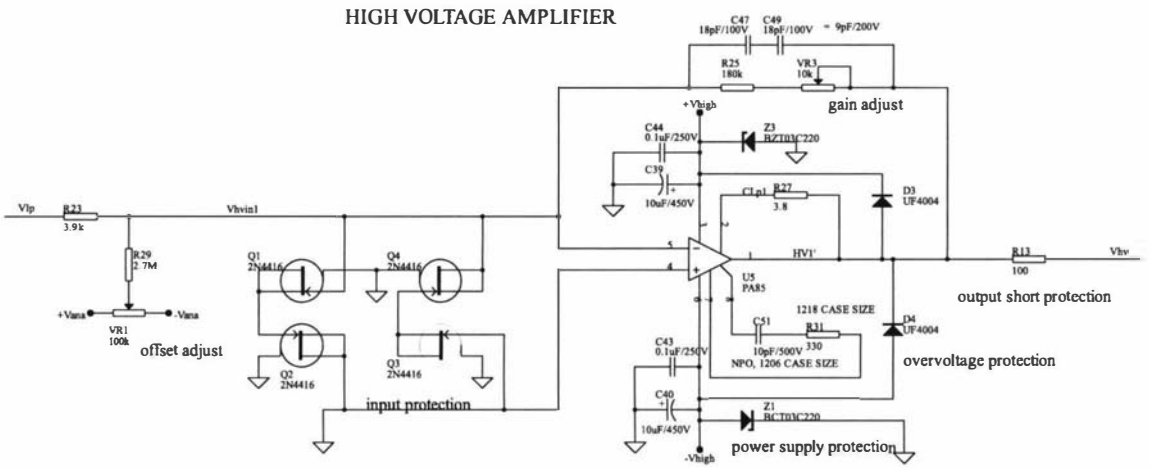


Figure 5.22: **High-Voltage Amplifier:** The high-voltage amplifier is an inverting amplifier based on the operational amplifier PA85 from Apex. The input is the low-pass filter voltage V_{lp}, the output the amplified high-voltage signal V_{hv}. The protection circuit, offset and gain trim as well as power supply decoupling is shown.

The input to each high-voltage amplifier is V_{lp} from the low-pass filter. The filtered voltage signal is amplified by the high-voltage amplifier to give

$$V_{hv} = -G_{hv} V_{lp}.$$

The amplifier is connected as an inverting amplifier with a gain of $G_{HV1} = 47$. The offset errors for the high-voltage amplifier due to bias current and offset voltage are

$$V_{hv,os} = (1 + \frac{R_{25} + VR_3}{R_{23}})V_{os} + (R_{25} + VR_3)I_b.$$

The bias current term is minute ($6\,\mu$V), since the PA85 is a FET input amplifier. The offset voltage term is $48 \times 2\,\text{mV} = 100\,\text{mV}$. The gain error is given by the resistor tolerances. The relative error for metal film resistors is 2% in the worst case. Gain and offset trim are provided as part of the HV module.

The size and ratio of the input and feedback resistances is discussed in appendix B.1.5. The resistances have been determined using a specially written computer program. It is important that the resistors do not generate too much noise and are capable of dissipating the heat generated when the output voltage is at a maximum, since the feedback resistor can then have the total high-voltage supply voltage across it. For a 1 W-resistor the resistance value has to be larger than $36\,\text{k}\Omega$ according to equation (B.7). With the current HV module, the feedback resistor is bypassed with two capacitors to provide a decreasing gain at frequencies higher than 100 kHz. The combined capacitance should be 9 pF and two capacitors are connected in series, so that the maximum voltage across each is not more than 100 V. In the future, the cut-of frequency could be set to a higher value, so that the low-pass filter remains the main filter in the output signal path and the step response is not slower than given by this low-pass filter.

High-Voltage Amplifier Protection

Although the high-voltage amplifiers are used in the same manner as ordinary operational amplifiers, care has to be taken to prevent damage to the amplifiers due to the high supply voltages. During the course of this project two amplifiers have been accidentally destroyed while being exposed to adverse conditions. Several components have therefore been added to the HV module to ensure that the high-voltage amplifiers are protected. The following list itemizes what has been done for protection along with other issues that need to be addressed for proper functioning according to the specifications in the data sheet:

- **Heatsinking:** Since each PA85 has a quiescent current of about 25 mA, a considerable amount of electrical energy is dissipated as heat within the PA85s:

  $$\dot{Q} = (+V_{high} - (-V_{high}))I_q = 380\,\text{V} \times 25\,\text{mA} = 9.5\,\text{W},$$

  where $\dot{Q}$ is the rate at which heat is dissipated, $\pm V_{high}$ is the supply voltage and I_q is the quiescent current.

  Each PA85 is mounted onto a heatsink to remove excess heat, thereby keeping the body temperature of the amplifier below 70°C. The chosen heatsink has a heat resistance of 2.4 K/w, which means a temperature rise of about 23 K above ambient temperature. The heat resistance value is

only valid under the conditions that the heatsink is mounted vertically and that air can flow easily to and away from the heatsink, which means that ventilation slits in the case are required. The actual temperatures were higher than indicated here, probably because the heat sinks were not entirely free standing.

- **Current Limiting:** A current limiting resistor is connected between the output of the PA85 and a dedicated current sensing pin of the PA85. The resistor value of 3.8 Ω has been chosen to limit the output current of any PA85 to a maximum of 200 mA.

- **Stability:** The PA85 features external frequency compensation: an R-C-network is connected to dedicated pins 7 and 8. In this circuit the highest recommended cut-off frequency has been chosen to allow a signal bandwidth of 100 kHz. Both the resistor and the capacitor are surface-mounted devices and are installed as close as possible to the respective pins.

- **Overvoltage Protection:** Mounted close to the PA85 are two overvoltage protection diodes (BZT03-C220), which are connected between ground and each high voltage supply line. The diodes shunt overvoltage spikes greater than 233 V and can absorb a total of 600 W × 100 μs = 0.06 J, which could occur due to a current of 2.5 A at 240 V with 100 μs duration.

- **Input Protection:** The inverting and non-inverting inputs of each PA85 are interconnected via four JFETs. Each JFET acts like a diode and so the equivalent circuit is 2 back-to-back diodes clipping input voltages to ±1.4 V. JFETs are used instead of diodes since they possess a smaller intrinsic capacitance and will therefore switch faster, i.e. hopefully before damage is done. The 2N4416 JFETs used have a maximum input capacitance 4 pF.

- **Output Protection:** Across each high voltage output and both of the power supply rails there are UF4004 diodes. These are fast switching diodes (50 ns), which can stand currents of up to 1 A. These are meant to absorb any electrostatic discharge to the amplifier outputs due to accidental mishandling of the amplifier board.

Noise

The noise specifications for the HV module are the most challenging of all the modules, so careful layout of the printed circuit board is especially important. This task is made harder by the limited amount of space on the board, mainly due to the size of the heatsinks for the high-voltage operational amplifiers. The HV module uses dual operational amplifiers and the low-pass filters have only two, instead of four, poles.

The actual output noise is built up from the intrinsic noise of each of the components, such as Johnson noise and amplifier noise, as well as externally injected noise, such as interference, ground loop noise or capacitively coupled noise. To calculate the component noise, appendix B.1.4 can be consulted first to decide about the gain distribution of the three relevant stages. In appendix B.1.4 it is stated that the least amount of extra noise is ensured when the earliest stage

has the maximum possible gain and the last stage has the least affordable gain. This has not been considered in the first design, but is recommended as a future design improvement. The validity of the argument has been tested on one HV module. The improved design is discussed below, the original design in the following paragraphs.

The component noise can be calculated according to table B.1 in appendix B.1.1. Taking those results we can determine the individual noise contributions of the HV module gain stages as

$$\begin{aligned} V_{\mathrm{nto}} &= G_{\mathrm{hv}} G_{\mathrm{lp}} V_{\mathrm{nt}} = 47 \times 1.27 \times 15\,\mu\mathrm{V} = 895\,\mu\mathrm{V} \\ V_{\mathrm{nho}} &= G_{\mathrm{hv}} V_{\mathrm{nl}} = 47 \times 3.15\,\mu\mathrm{V} = 148\,\mu\mathrm{V}, \end{aligned}$$

where V_{nto} is the output noise voltage due to the low-pass filter and V_{nho} is the output noise voltage due to the high-voltage amplifier. To find the total component noise of the HV module, the noise of the digital-to-analog converter is added to the above noise. The AD1862 digital-to-analog converter data sheet states a minimum signal-to-noise ratio of 110 dB, which already includes the amplifier noise of a current-to-voltage converter using the on-chip feedback resistor of $R_{\mathrm{f}} = 3\,\mathrm{k}\Omega$. In the data sheet the operational amplifier used was the NE5534[13]. If noise reducing capacitors are used with the converter, as is the case in the current design, the signal-to-noise ratio of the converter rises to 119 dB. This last value was used to find the DAC noise level at the output

$$\frac{190\,\mathrm{V}}{119\,\mathrm{dB}} = \frac{190\,\mathrm{V}}{891 \times 10^3} = 213\,\mu\mathrm{V}.$$

Since the signal-to-noise specification for the AD1862 is given for a full scale sine wave it will already incorporate the quantization noise. The quantization noise was discussed in section 4.3.3. For the HV module it would be

$$V_{\mathrm{q}} = \frac{k_{\mathrm{hv}}}{\sqrt{12}} = \frac{360\,\mu\mathrm{V}}{3.46} = 104\,\mu\mathrm{V},$$

where k_{hv} is the resolution of the HV module at the output. This noise is taken as being part of the AD1862 noise specification. The sum of all noise contributions listed above is

$$V_{\mathrm{n,tot}} = \sqrt{895^2 + 148^2 + 213^2}\,\mu\mathrm{V} = 932\,\mu\mathrm{V}.$$

Measured values are presented in table 7.3.

Suggested Design Improvement

The smallest possible gain for the high-voltage amplifier is

$$G_{\mathrm{hv}} = \frac{190\,\mathrm{V}}{12\,\mathrm{V}} = 15.8,$$

where 12 V is maximum output voltage of the OP-270 operational amplifier used in the low-pass filter. The same low-pass filter has a fixed gain of $G_{\mathrm{lp}} = 1.268$

[13]The NE5534 is from Texas Instruments. An equivalent part is the SE5534 from Signetics.

by design. Judging from that, the first stage, which is the current-to-voltage converter for the digital-to-analog converter, should have a trans-resistance of

$$R_{\mathrm{tr}} = \frac{12\,\mathrm{V}}{1.268} \times \frac{1}{1\,\mathrm{mA}} = \frac{9.5\,\mathrm{V}}{1\,\mathrm{mA}} = 9.5\,\mathrm{k}\Omega.$$

With these values the component noise has been calculated according to table B.1 in appendix B.1.1. Taking those results we can calculate the individual noise contributions of the HV module gain stages as

$$\begin{aligned} V_{\mathrm{ndo}} &= G_{\mathrm{hv}} G_{\mathrm{lp}} R_{\mathrm{tr}} I_{\mathrm{nd}} = 15.8 \times 1.268 \times 9.5\,\mathrm{k}\Omega \times 1.58\,\mathrm{nA} = 297\,\mu\mathrm{V} \\ V_{\mathrm{nto}} &= G_{\mathrm{hv}} G_{\mathrm{lp}} V_{\mathrm{nt}} = 15.8 \times 1.268 \times 15\,\mu\mathrm{V} = 301\,\mu\mathrm{V} \\ V_{\mathrm{nho}} &= G_{\mathrm{hv}} V_{\mathrm{nl}} = 15.8 \times 3.10\,\mu\mathrm{V} = 49\,\mu\mathrm{V}, \end{aligned}$$

where V_{ndo} is the output noise voltage due to the current-to-voltage converter, V_{nto} the output noise voltage due to the low-pass filter, and V_{nho} is the output noise voltage due to the high-voltage amplifier. Since current-to-voltage converter noise is already included in the converters signal-to-noise ratio, this current-to-voltage converter noise might be subtracted. It is smaller than the value above, because a smaller feedback resistor is indicated in the data sheet. The noise to be subtracted is $15.8 \times 1.268 \times 6\,\mu\mathrm{V} = 120\,\mu\mathrm{V}$ which leaves a noise level of $272\,\mu\mathrm{V}$ instead of $297\,\mu\mathrm{V}$. This noise has to be added to the noise of the digital-to-analog converter to find the total HV module component noise.

The contribution of the converter is the same as above, $213\,\mu\mathrm{V}$. The sum of all noise contributions listed above is therefore

$$V_{\mathrm{n,tot}} = \sqrt{272^2 + 301^2 + 49^2 + 213^2}\,\mu\mathrm{V} = 461\,\mu\mathrm{V}, \tag{5.21}$$

which is almost half the calculated noise of the previous configuration.

Passive Filter Option In case the active filter is omitted, the total noise reduces to

$$V_{\mathrm{n,tot}} = \sqrt{\frac{\pi}{2}} \times \sqrt{272^2 + 49^2 + 213^2}\,\mu\mathrm{V} = 437\,\mu\mathrm{V},$$

where the same high-voltage amplifier gain (G_{hv}) can be used, since the current-to-voltage converter will output voltages up to 12 V instead of 9 V. A correction factor of $\sqrt{(\pi/2)}$ was multiplied to account for the gentle roll-off of a simple RC filter, as derived in appendix B.1.2, equation (B.3). If a higher order passive is used, then the noise could be as low as $350\,\mu\mathrm{V}$, which comes closer to the desired noise specification. Clark *et al.* [46] use a commercial seven pole passive filter (APQ-25 from Toko America), but no comparable component was conveniently available during the design phase of this project.

Speed

The HV module has three analog signal processing stages, the current-to-voltage converters for the digital-to-analog converters, the low-pass filters and the high-voltage amplifiers. The 2-pole Bessel low-pass filter, designed for a 100 kHz bandwidth has a step rise time of $4\,\mu\mathrm{s}$. Since both other stages are also bandwidth limited a longer step rise time than that is expected. The step rise time

is discussed in appendix B.1.3. Two of these stages have a step rise time of approximately 8 μs, which is the value listed as t_{sr2} in table B.2 in appendix B.1.3. Combined together with the above value gives an overall step rise time greater than 12 μs. The actual response of the HV module was measured with an oscilloscope, where a time constant τ of $5.7 \pm 0.7\,\mu$s was found. The corresponding step rise time is

$$\ln 10\,\tau = 2.3 \times 5.7\,\mu\text{s} = 13\,\mu\text{s},$$

which is reasonably close to the predicted value, i.e. within the error of $\ln 10 \times 0.7\,\mu\text{s} = 1.6\,\mu$s. In a future design the bandwidth of the current-to-voltage converter and high-voltage amplifier should be shifted to higher frequencies to give a faster step rise time. An alternative is to omit the active low-pass filter altogether as mentioned above. The resulting step rise time should then be close to the 4 μs of the Bessel filter alone.

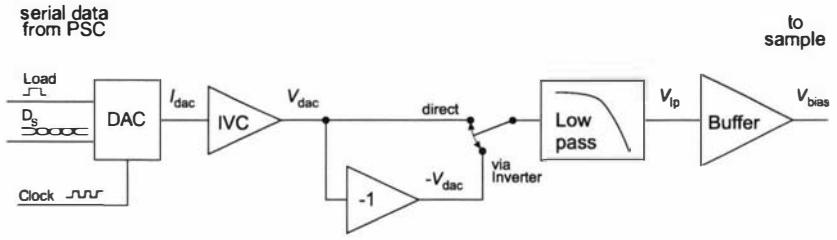


Figure 5.23: **Bias Module:** The Bias module is similar to the HV module in that it consists of a digital-to-analog converter (DAC), current-to-voltage converter (IVC), low-pass filter and output amplifier. The DAC is configured in unipolar mode. Negative bias voltages can be chosen by the inclusion of an optional inverter. Again, the low-pass filter is a reconstruction filter for the output V_{dac} of the DAC. The output amplifier is a driver for the output coaxial cable to the sample.

5.4.7 Bias Module

The bias voltage between tip and sample is generated by the Bias module. As discussed in section 5.1.2, the specified range of the bias voltage is $\pm 10\,\mathrm{V}$ with a resolution of a few millivolts and a bandwidth of 10 kHz. As with the high-voltage module, the Bias module consists of a digital-to-analog-converter, a transimpedance amplifier, a low pass filter and an output driver, which together produce an analog output voltage (V_{bias}) from a digital input number (N_{bias}). A diagram is shown in figure 5.23. The Bias module should output zero volts on reset or after the Bias module is switched on. For reasons of internal architecture the digital-to-analog converter is therefore used in unipolar mode. To allow bias voltage outputs of either polarity an optional inverter has been incorporated into the design, which can be switched into the signal path via a relay. The inverter is placed after the transimpedance amplifier. The overall formula for the Bias module is

$$V_{\mathrm{bias}} = k_{\mathrm{bias}} N_{\mathrm{bias}},$$

where k_{bias} is the weight of the least significant bit. Since the digital-to-analog converter is unipolar, k_{bias} will be

$$k_{\mathrm{bias}} = \frac{10\,\mathrm{V}}{2^{12}} = 2.4\,\mathrm{mV}.$$

Digital-to-Analog Converter

A 12-bit digital-to-analog converter, PM-7543 from Analog Devices, is used on the Bias module as a converter since it has sufficient resolution and speed. The PM-7543 is supplied with a 5 MHz clock signal. The conversion time is $12 \times 200\,\mathrm{ns} = 1.2\,\mu\mathrm{s}$. This is double the minimal conversion time. The output current of the converter settles within a minimum of $1\,\mu\mathrm{s}$.

A digital-to-analog converter is of an R-2R-ladder type and requires an external reference voltage V_{ref}. A +5 V voltage reference REF02 from Burr Brown is used. The transimpedance amplifier circuit is the same as that suggested in the PM-7543 data sheet [8]. The amplifier uses the on-chip feedback resistor of the PM-7543 and the reference voltage to output voltages from 0 to 5 V. The output of the digital-to-analog converter and transimpedance amplifier is

$$V_{\mathrm{dac}} = k_{\mathrm{dac}} N_{\mathrm{bias}},$$

with $k_{\mathrm{dac}} = 1.22\,\mathrm{mV}$.

Since the PM-7543 does not offer twos complement input code, it was decided to run the converter in unipolar, rather than in bipolar mode. The main reason was that in unipolar mode the converter will output zero after it has been powered up, while it would revert to minus full scale when used in bipolar mode. It is possible to invert the most significant bit on the digital side before the word gets sent to the converter, but that would not prevent the converter from outputting negative full scale when power was applied to the bias module before it was applied to the digital electronics.

Low-Pass Filter

The low-pass filter is a two-stage, four-pole Bessel filter with a cut-off frequency of 10 kHz. The filter design is discussed in appendix B.1.5. The output of the low-pass filter is

$$V_{\mathrm{lp}} = G_{\mathrm{lp2}} G_{\mathrm{lp1}} V_{\mathrm{dac}} = G_{\mathrm{lp}} V_{\mathrm{dac}},$$

with a gain of $G_{\mathrm{lp}} = G_{\mathrm{lp2}} G_{\mathrm{lp1}} = 1.8 \times 1.084 = 1.95$.

Output Driver

This discussion is similar to the one for the output driver of the IVC module in section 5.4.5. An OP27GP operational amplifier is used to drive the output of the Bias module. It is connected as a follower with a 1 kΩ resistor in the feedback path. The resistor limits input currents through the input protection diodes under transient conditions. The amplifier can drive its mainly capacitive loads, the coaxial cable and the tip-sample capacitance, without going into oscillation. The cable capacitance is $2.3\,\mathrm{m} \times 100\,\mathrm{pF/m} = 230\,\mathrm{pF}$. The tip-sample capacitance is much smaller, estimated to be at most 3 pF. The capacitive load limit for the OP27GP lies at 290 pF.

As with the output driver of the IVC unit, this driver also introduces an offset error of $180\,\mu\mathrm{V}$.

Noise Calculations

The component noise is calculated according to table B.1 in appendix B.1.1. The individual noise contributions of the Bias module gain stages are

$$\begin{aligned} V_{\mathrm{nbo}} &= G_{\mathrm{lp2}} G_{\mathrm{lp1}} V_{\mathrm{nin}} = 1.8 \times 1.084 \times 2\,\mu\mathrm{V} = 4\,\mu\mathrm{V} \\ V_{\mathrm{n1o}} &= G_{\mathrm{lp2}} G_{\mathrm{lp1}} V_{\mathrm{nb}} = 1.8 \times 1.084 \times 5.5\,\mu\mathrm{V} = 11\,\mu\mathrm{V} \\ V_{\mathrm{n2o}} &= G_{\mathrm{lp2}} V_{\mathrm{n1i}} = 1.8 \times 13.3\,\mu\mathrm{V} = 24\,\mu\mathrm{V}, \end{aligned}$$

The noise of the current-to-voltage converter as well as the noise of the inverter and output driver (not calculated here) is neglected, since it is so small. The sum of the component noise is

$$V_{\mathrm{nel}} = \sqrt{11^2 + 24^2}\,\mu\mathrm{V} = 26\,\mu\mathrm{V}.$$

The quantization noise is irrelevant, since in standard microscope operation the Bias module will supply a constant output voltage. Even in the case were a V-I characteristic curve is being recorded the quantization steps are known and can be taken into account by the software.

5.4.8 The Switchboard

A switchboard is linked between the HV modules and the probe head to allow reassignment of the HV module connections to the outer actuator electrodes.

As discussed in section 2.3.1, there are a total of four actuators in the probe head, with each actuator having four electrodes. One electrode pair ($\pm V_x$) produces movement in the x-direction, the other ($\pm V_y$) in the y-direction. One positioning control signal can move an actuator in the positive or negative direction depending on its polarity. The same signal can move one actuator in the positive direction and another one in the negative direction if the signal is applied to opposite electrodes. In general, two signals per actuator are needed. However, only three independent signals are needed for the control of the outer actuator. This means that 3 rather than 6 HV modules are needed, which is time- and cost effective.

The outer actuators sit at the vertices of an equilateral triangle. Assuming that all electrodes are aligned in the same direction, two different modes of operation are possible:

Circular Mode The outer actuators move in a corkscrew fashion. There are three independent control signals necessary to achieve motion in a corkscrew manner. The outer actuators, sitting on the perimeter of a circle, all have to move tangentially to this circle. For reasons of symmetry several of the control signals needed will have the same magnitude.

Lateral Mode All outer actuators move together in one direction. Only two pairs of control signals are needed: all outer actuators receive the same signals for their x-electrode pair and y-electrode pair respectively.

Switching between lateral or circular mode can be done with three (double) switches, as is explained in the following paragraphs.

The way the three outer actuators are driven is most conveniently given in vector notation. It is assumed that all actuators are aligned, so that their x-electrode pair points in the $\binom{1}{0}$ direction and their y-electrode pair points in the $\binom{0}{1}$ direction. In lateral mode all actuators are driven with the same signals

$$\mathbf{d}_1 = \mathbf{d}_2 = \mathbf{d}_3 = \begin{pmatrix} d_x \\ d_y \end{pmatrix},$$

where $\mathbf{d}_i$ with $i = 1,2,3$ are the drive signals in vector notation, and d_x and d_y are arbitrary. In circular mode the drive vectors must point in tangential directions, which are given as

$$\mathbf{d}_1 = \begin{pmatrix} -d \\ 0 \end{pmatrix}; \quad \mathbf{d}_2 = \begin{pmatrix} \frac{d}{2} \\ \frac{\sqrt{3}d}{2} \end{pmatrix}; \quad \mathbf{d}_3 = \begin{pmatrix} \frac{d}{2} \\ -\frac{\sqrt{3}d}{2} \end{pmatrix},$$

where d is again arbitrary. Three different signals, d, $d/2$ and $\sqrt{3}d/2$, are therefore needed for the circular mode.

Movement modes such as these have also been described by Wilms *et al.* [191]. Although in their probe head design the outer actuator electrodes are aligned tangentially to the circle, the result is the same. Wilms *et al.* have to use the lateral mode signals for the circular mode and vice versa to achieve the same movements. The switchboard connections are described in B.0.4.

Chapter 6

Software

6.1 Overview

In the system described here an extra layer between operator and hardware is provided by software. There are two main programs, one running on the PC and one on the DSP. The software organizes the simple tasks and controls all essential microscope variables to achieve

- a safe tip-sample approach (APPROACH),
- recording of topographical images (RASTER),
- recording of spectroscopic curves (CURVES),
- nanolithography and
- a check of system performance (STEP).

The names of the corresponding software routines are mentioned in parenthesis and can be found in figure 6.1. The microscope variables are shown as the top row in figure 6.1. Any of these variables, probe head position, tip position and bias voltage, can be determined and set by the microscope operator.

The major routines, such as APPROACH and RASTER, call upon subroutines as illustrated in the diagram in figure 6.1. All routines are discussed in detail in this section.

There are two important demands imposed on the DSP main program. Once the tip is in tunneling distance from the sample, the tip-sample distance needs to be controlled continuously. At the same time the operator must be able to influence any ongoing process, for example interrupt it, start a different operation or upload data. The tip-sample control has to be executed in real time, which is an additional constraint on the DSP program. In figure 6.1 the two concurrent processes, the tip-sample control and the communication with the PC, are indicated as shaded boxes.

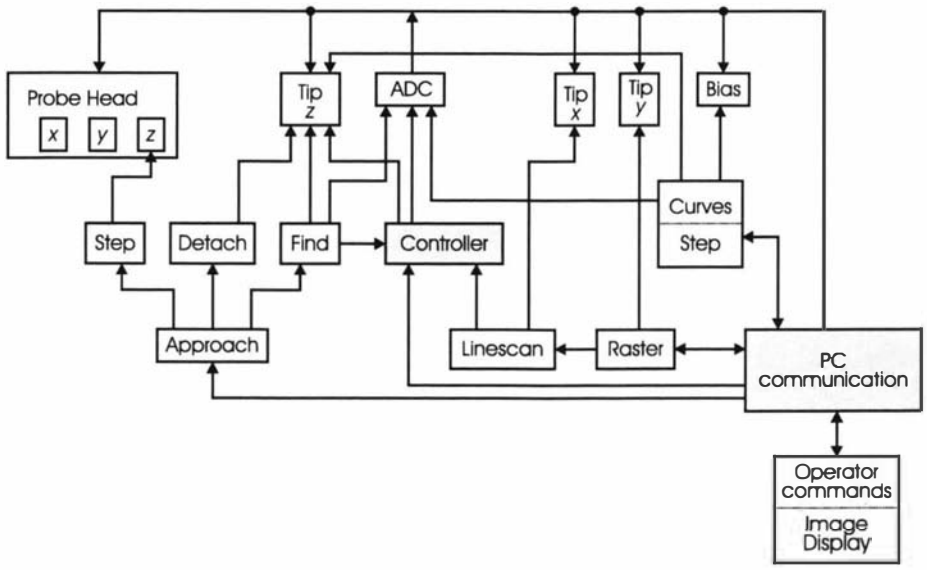


Figure 6.1: **Software:** The diagram shows all essential microscope variables and the most important software routines. The microscope variables are shown along the top of the diagram, while the remaining blocks present the routines. The arrows indicate that a routine is invoked by another. The two shaded boxes are tasks that often run in parallel with the remaining routines.

6.2 Programming the DSP

6.2.1 Programming

The program for the DSP is written in the high-level language C. The code is generated and cross compiled on the PC and downloaded to the DSP via an RS232 serial data link.

The cross compiler for the ADSP-21020 DSP as well as downloading software was provided by the manufacturer of the evaluation board [7]. On the DSP side, software to handle program downloading is stored in ROM.

The downloading software for the PC was written for the MS-DOS operating system. The software still functioned under the succeeding operating system Windows 3.1, but did not do so under Windows 95. A new piece of software had to be written and compiled for Windows 95 to interact with the software residing in the DSP. The machine language code for the ADSP-2111 processor, which handles the downloading of the software, was made available by the manufacturer. The exact functioning had to be inferred from this code and a corresponding program was written.

6.2.2 Processor

The ADSP-21020 is a floating point digital signal processor. The floating point format has many advantages, mentioned in section 5.3.1. The main ones are high precision and wide dynamic range. Within the program the process variable and the controller output are floating point variables. The process variable is converted from its integer representation as soon as it has been read from the hardware, while the controller output is converted from floating point format to

integer just before it is sent to the hardware.

6.3 DSP Main Program

6.3.1 Single Program Approach

The DSP program has to handle two regular processes concurrently. It has to communicate with the supervising program on the PC and it must interact with the tip-sample system via the probe electronics.

One way of writing the DSP software would be to establish a real-time operating system allowing multiple processes simultaneously [15, 24, 45]. Such an operating system can be more powerful, but is also more complex. Dr. Blair Hall, one of the supervisors of this project, wrote a real-time operating system for the PC and DSP, which is presented in appendix C.

In the here described project the DSP software was so far programmed using a single program approach, which is sufficient for simple operations [24]. The program uses polling to communicate with the PC and a real-time clock to time the interaction with the tip-sample system.

6.3.2 Polling

The process of polling is described as interrogating a device for readiness at regular intervals. Readiness is indicated by a hardware status line of some sort. Action is taken if the status line indicates that the device is ready. The first conditional branch in the flow chart of the main routine in figure 6.2 shows such a poll. The status line is a flag that the DSP can test. The PC sets the flag in the DSP system interface. Reading the PC command resets the flag.

The alternative to polling is the use of interrupts. However, the use of interrupts to attend to the probe electronics was ruled out, because of the programming overhead. Within a desired sampling time of $10\,\mu$s the DSP can execute

$$\frac{10\,\mu\text{s}}{30\,\text{ns}} = 333$$

instructions. Interrupts can be used by calling dedicated routines, so-called interrupt dispatchers, from the runtime C library. Using the normal interrupt dispatcher from the library requires approximately 125 cycles of interrupt overhead, which is not acceptable. In contrast, checking a flag status or reading the on-chip timer of the ADSP-21020 takes only a few cycles.

Attending to the probe electronics is the most important task of the DSP. This has to be done in short regular intervals as mentioned in section 4.4. Since there are no interrupts, this time interval may in principle vary. It is, however, more convenient to lay down a certain fixed time slot, the sampling time t_s, for the DSP program. Communicating with the PC is the second main task of the DSP. Polling is preferred over interrupts so that the tip-sample control is never interrupted, which could otherwise cause a tip crash. The polling intervals need not be as short as the sampling time.

6.3.3 Main Routine

Figure 6.2 shows the main routine running on the DSP. The program checks two

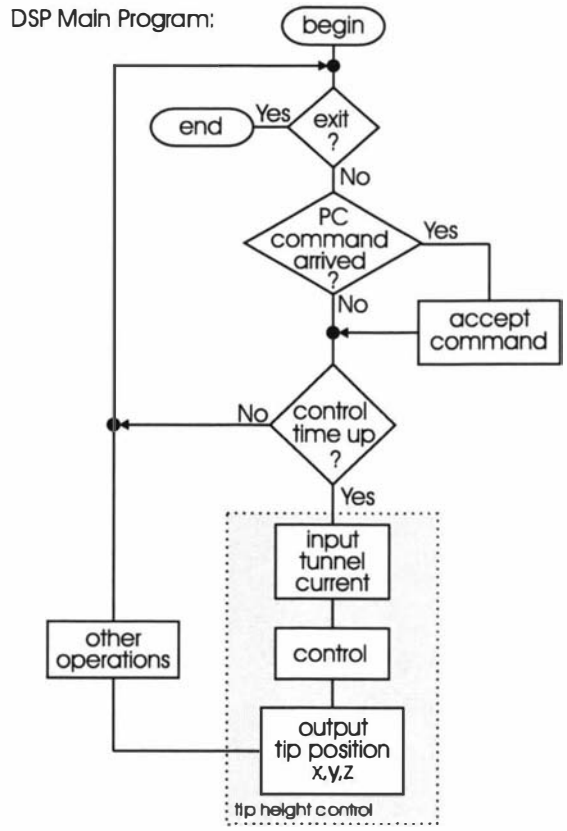


Figure 6.2: **Main Routine:** A single program handles all DSP operations. The main routine constantly supervises two processes, PC communication and tip positioning. The input channel from the PC is polled, while the input channel from the probe electronics is checked at constant intervals using a real time clock. The output channel is updated each time the input channel is.

conditions in its main loop, whether a PC command has arrived and whether the control time has run out. If neither is the case the DSP keeps on polling the PC, which provides for a rapid response to any operator input. The timing could also be changed so that the PC input is polled only every n^{th} cycle thereby reducing the programming load slightly. If $n = 78$, then the PC would be polled every millisecond, which should be enough for a human operator.

When it is time for control the DSP usually acts as a direct digital controller, which is discussed in more detail in section 6.9.2. As a controller, the DSP reads the tunneling current, computes the controller output word and writes this word out. Apart from tip-sample distance control this sequence can also be used to realize a topographical scan or recording of characteristic curves. It is important that output word computation and the actual output occur immediately after reading the tunneling current input. The output is a consequence of the input and an immediate output decreases the programming dead time.

The control voltage for the tip height (V_z) is calculated with an incremental algorithm of the form

$$V_z(\text{new}) = V_z(\text{old}) + \Delta V_z\,, \tag{6.1}$$

so that the output value changes only by a small amount. If feedback is active, the controller updates the increment ΔV_z. The advantage of using the incremental form of control is that the program controller can easily be replaced by other routines using the same position calculation principle. Other routines may take over from the controller by bypassing the controller and feeding their own position increment to the main routine. If the change is set to zero in software, the main routine holds the current position. The actuator position always changes smoothly, starting from its last held value. This has been used to implement the FIND and DETACH routine.

The position of these routines would be in the block labeled OTHER OPERATIONS. Apart from approaching the tip there are also operations such as recording a step response and the bookkeeping necessary to scan a sample and record characteristic curves.

6.3.4 Output to Probe Electronics

To communicate with the probe electronics, the DSP reads and writes to memory mapped registers on the DSP system interface. A single cycle of the DSP is too short for the interface to react. The DSP inserts a certain number of waitstates when communicating with the probe electronics. The memory location of the ports is specified in an architecture file located on the PC. The architecture file is linked with the C program on compilation. The number of waitstates is also determined in the architecture file. In the program the input and output channels are referred to by the name given in the architecture file.

Subroutines handle reading from and writing to the probe electronics. These subroutines convert between the floating point format used in the rest of the program and the integer format needed for the hardware. Reading from the DSP system interface requires a subroutine to set all higher bits equal to the most significant bit of the input word, since the input word coming from the ADC module contains 16 bits, but the data bus is 32 bits wide.

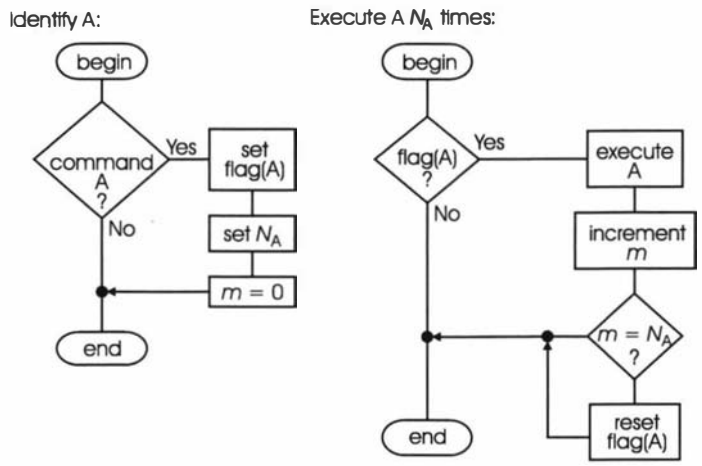


Figure 6.3: **Execution of a Task:** The flow chart in figure 6.3 to the right shows how a task A is repeated N_A times. The flow chart on the left explains how task A is set up just after the command to execute A has been received.

6.3.5 Data Transfer

The DSP uses polling to transfer data to the PC. If data are to be uploaded, the PC issues an appropriate command and the number of data that should be transferred. At this stage either a single byte or a block of floating point data can be transferred. As soon as the DSP detects the command it sends the data. The floating point data are broken into four bytes each, which are then transmitted. On the PC side groups of four bytes are then rearranged into floating point format. Care has to be taken, since the PC and DSP use different orders to store floating point data. The necessary hardware handshaking is described in section 5.3.5.

6.3.6 Tasks

The main DSP routine can execute several tasks, which involve either communication with the PC, controlling the tip-sample distance or both. The main routine regularly polls both the PC input and the tunneling current input. This is important since on the one hand a tip crash should be avoided and on the other hand the operator should be able to interrupt any DSP process. As a consequence of polling, many microscope operations are broken up in a manner that allows task execution as well as polling. Typically a task alternates with tip height control, for instance when scanning a line. In this case the routine LINESCAN moves the tip horizontally and the controller keeps a certain distance between tip and sample.

Figure 6.3 shows two flow charts, which describe how a task, labeled A, is executed N_A times and how it is initialized. The execution is organized with the help of flags. The DSP checks the flag that belongs to a task A once within each loop of the main routine.

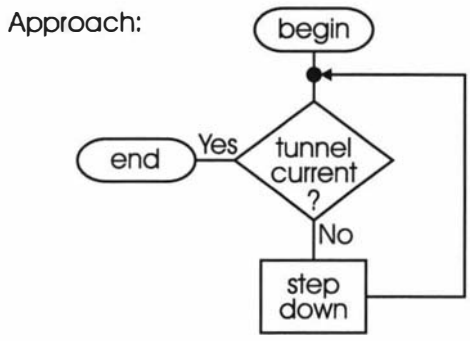


Figure 6.4: **Approach:** The routine APPROACH steps the probe head down towards the sample until tunneling current is encountered. Although not obvious this is a complex procedure. STEP is a routine called by APPROACH.

6.4 Approach

The first step when operating the microscope is to engage the tip so that it is within tunneling distance of the sample. Approaching the tip is done automatically as described in section 2.3.3.

The algorithm used during the approach is shown in figure 6.4. The probe head steps down the ramp until tunneling current is encountered. Although the routine APPROACH seems to be simple, the automatic approach is an involved process. Two different types of approach, probe head approach and center actuator approach, are alternated and the tip must not be allowed to crash. The APPROACH routine relies on the subroutine STEP, which in turn invokes DETACH and FIND.

6.4.1 Step

The STEP routine handles both the probe head approach and the center actuator approach. The flow chart for this process is shown in figure 6.5. The timing and the distances involved are described in detail in section 7.3. Here only the process itself is presented. The probe head steps down towards the sample with retracted tip. To withdraw the tip the routine DETACH is invoked. As mentioned in section 3.1.3 the probe head takes several small steps down the ramp instead of one large one. Taking too large steps agitates the probe head so that it does not stay on track, thus smaller steps make it easier to control the probe head. Finding tunneling current is left to the routine FIND.

6.4.2 Detach

As shown in figure 6.5 the tip has to be withdrawn before the probe head takes a step. The flow chart to detach the tip is shown in figure 6.6. The actuator is retracted by setting the height increment mentioned in equation (6.1) to a small negative value. The retraction is stopped by setting this increment to zero.

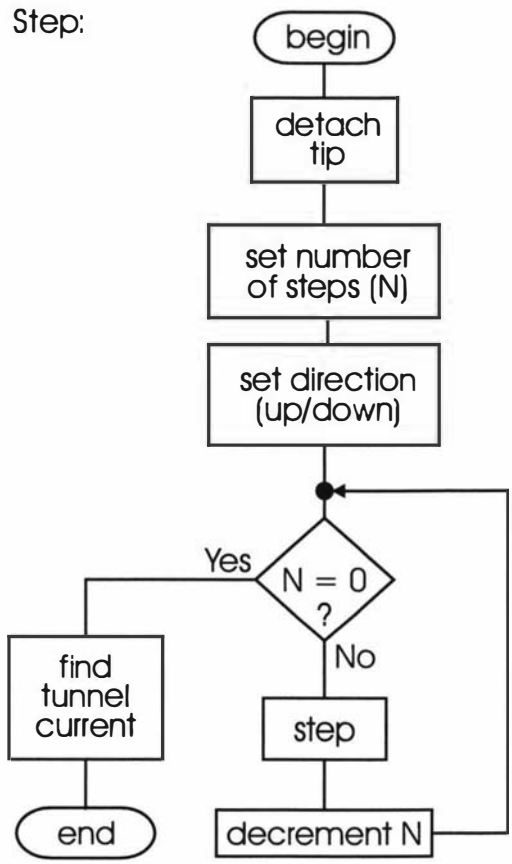


Figure 6.5: **Step:** The routine STEP lifts the tip and steps the probe head up or down the ramp. STEP is called by APPROACH, but the operator may also use it directly. DETACH and FIND are subroutines called by STEP.

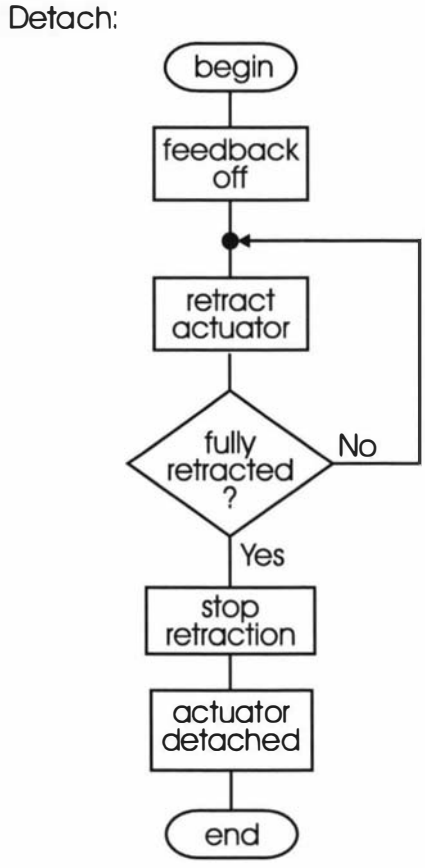


Figure 6.6: **Detach:** The routine DETACH replaces the feedback control and withdraws the tip to a maximum withdrawal length.

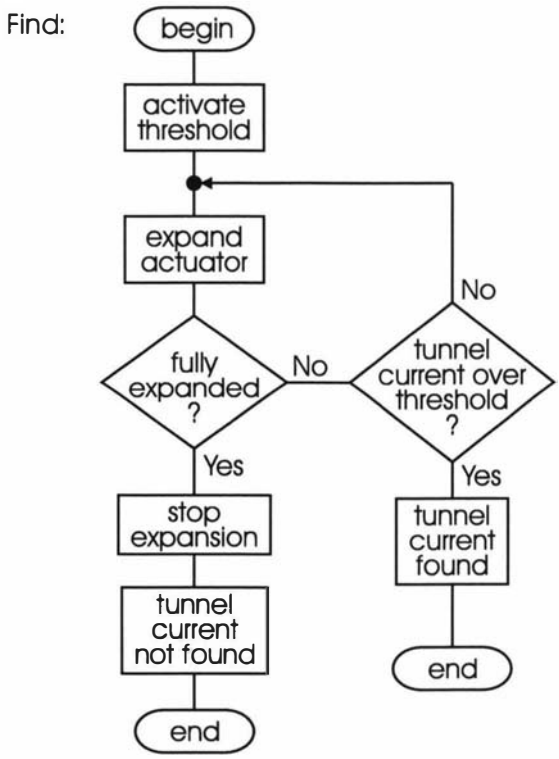


Figure 6.7: **Find:** FIND proceeds in a similar way to DETACH with the additional condition of having to prevent tip crash. FIND returns either that tunneling current has been encountered or that the actuator is at its maximum approach length.

6.4.3 Find

After the probe head has stepped down, the program searches for tunneling current. Figure 6.7 presents the appropriate flow chart. The routine FIND shown in figure 6.7 operates in a similar fashion to DETACH. An important difference is the condition under which the tip advance is terminated. In most cases during APPROACH, FIND will end after the center actuator is fully expanded. If the tip is close enough, then FIND will encounter tunneling current and indicate to the main routine that it is finished and feedback should be switched on. The timing and the possibility of tip crash using the approach method described by FIND and another method are discussed in section 7.3.

6.5 Sample Investigation

The microscope can scan the underlying surface topographically or record characteristic curves. A brief overview of these operations is given in section 2.3.3.

6.5.1 Raster Scan

The raster scan is presented as a flow chart in figure 6.8. A raster scan consists of a sequence of linescans. LINESCAN is a sub-routine invoked by RASTER. The

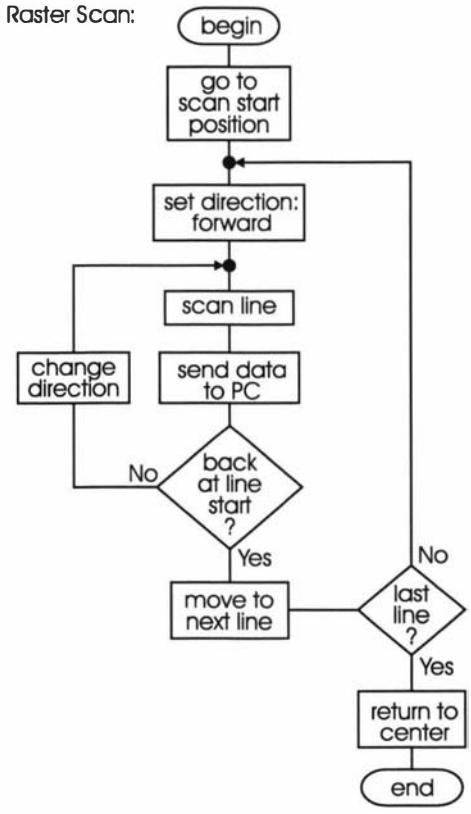


Figure 6.8: **Raster Scan:** The routine RASTER organizes several individual linescans. In both directions data are recorded and transmitted to the PC. RASTER invokes LINESCAN.

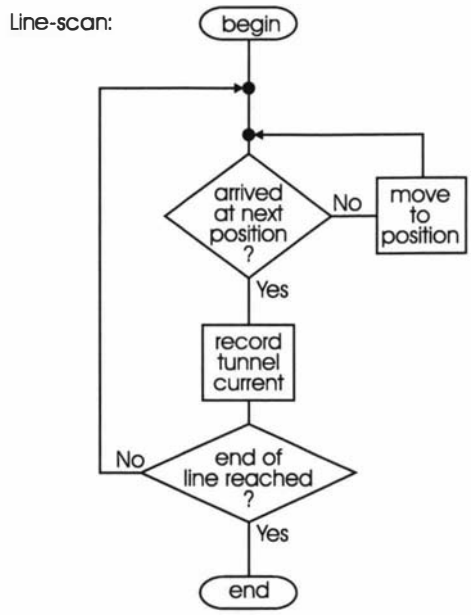


Figure 6.9: **Linescan:** The routine LINESCAN is used by RASTER, but can equally serve to record V-I curves. Two important issues are the method of data acquisition and point-to-point movement.

routine RASTER brings the tip first to its horizontal start position. The scan area is chosen so that the tip without horizontal offset is in the center of the scan area. The start position is at one corner of this area. RASTER records data in both directions of each linescan. The data are sent to the PC, the corresponding routine on the PC is discussed in section 6.7.2. After the raster scan is finished the tip will be in its original position. Moving the tip to the start position and back to the center are routines like LINESCAN but without data acquisition.

6.5.2 Linescan

A raster scan is two dimensional. It is made up from several one dimensional linescans. Before a line is scanned the bias voltage is set to a desired value and keep constant. The operator is able to set the horizontal increments of the linescan. A flow chart of a linescan is shown in figure 6.9. The routine LINESCAN needs more detailed explanation because of the way the tip is moved from one sample point to the next and because of the methods used to record data. Data acquisition is discussed first.

6.5.3 Acquisition Methods

At some point during a linescan data are collected. In figure 6.9 this point is indicated by the statement RECORD TUNNELING CURRENT. Scanning a line or recording a characteristic curve are very similar. For both modes one of two possible acquisition methods has to be chosen. The two methods are the mean value and the sampling method. Figure 6.10 shows a flow chart describing how data are acquired.

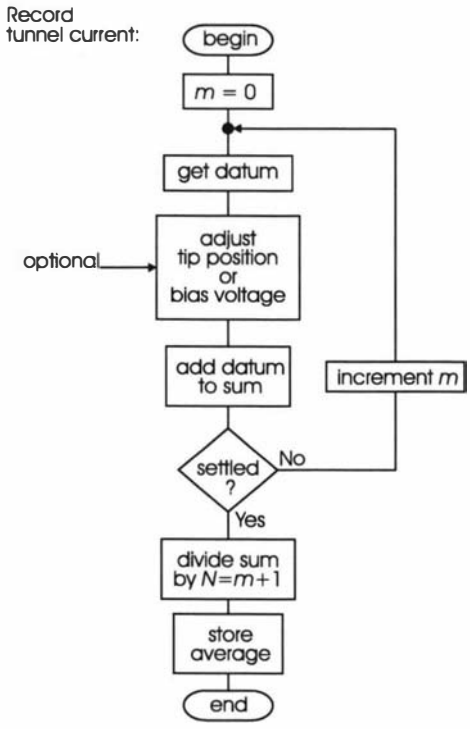


Figure 6.10: **Recording Tunneling Current:** The flow chart presents a routine that collects N data values and stores the average. One command is optional. If it is used, then the routine describes the mean value method of moving from point to point. Using the mean value method the routine changes either tip position or bias voltage slightly while collecting data. The average value is stored. The routine can also be used at one particular sample point, where bias voltage and tip height remain constant. In this case the routine describes only the data collection part in the linescan flow chart in figure 6.9.

Mean Value Method

Data can be acquired by collecting a certain number of data points while the variable under consideration (bias voltage or tip position) is changing. The data are then averaged and the mean value assigned to one point of the data set. Variable change and data acquisition are happening at the same time and the data are a collection of the past values. The mean value method is described by the flow chart in figure 6.10, where the optional statement is included.

Sampling Method

An alternative to the mean value method is to collect data at the sample points only. More than one datum can still be accumulated and averaged at one sample point, but the variable is changed between the collecting of data. This method, the sampling method, consists of two parts, the point-to-point movement and data collection at the sampling point. Ways in which movement from one sampling point to another can be achieved are described in the flow chart in figure 6.11. The data collection is described by the flow chart in figure 6.10. The optional statement is excluded unless the tip-sample distance has to be controlled, in which case acquisition and vertical tip adjustment may proceed simultaneously.

Reaction Time

After a change in bias voltage or tip position has occurred it will take a certain time before the data that are available to the DSP reflect the reaction to the change. This fact requires that the DSP waits for the reaction to the variable change, as it would otherwise collect incorrect data.

The reaction time of the bias voltage, the actuator and the current-to-voltage converter are different. They are discussed in section 7.2. The time the DSP needs to wait depends also on the way the change has been made. For instance if the actuator voltage changes step-wise, then ringing might occur and the tunneling current needs a certain time to settle. The actuator voltage should therefore be changed smoothly.

Moving between Sampling Positions

The tip can move from one sample point to the next using either the mean value or the sampling method. Using the sampling method, a great variety of possible paths lead from point to point. Figure 6.11 shows how the tip moves and lists several different conditions that lead to different trajectories. The path the tip would take using the different methods is illustrated in figure 6.12. The conditions and trajectories are labeled following Piner and Reifenberger [154]. For the incremental and Piner/Reifenberger (P/R) trajectories the program takes several smaller intermediate steps. The incremental method is getting as close as possible to the continuous motion of an analog system and is dubbed here the incremental method due to the small increments it has to take each sampling interval. For the incremental trajectory the tip height is adjusted after each small step. The P/R trajectory is distinct from the incremental method in that it adjusts the tip height only if the tunneling current is higher than a certain maximum.

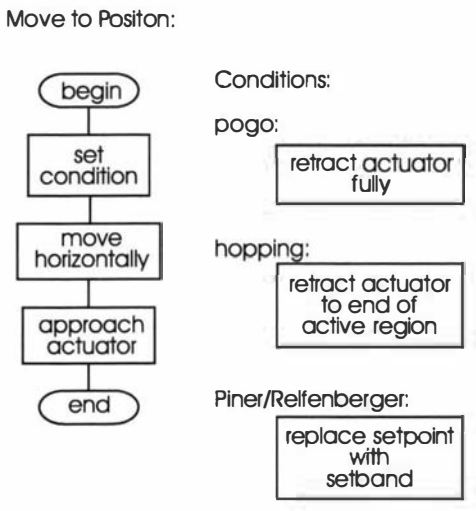


Figure 6.11: **Point to Point Movement:** The flow chart illustrates how the tip moves from one point to another when the sampling method of data acquisition is employed. To the right three different conditions are listed that each leads to a different trajectory, shown in figure 6.12. The condition labels are after Piner and Reifenberger [154].

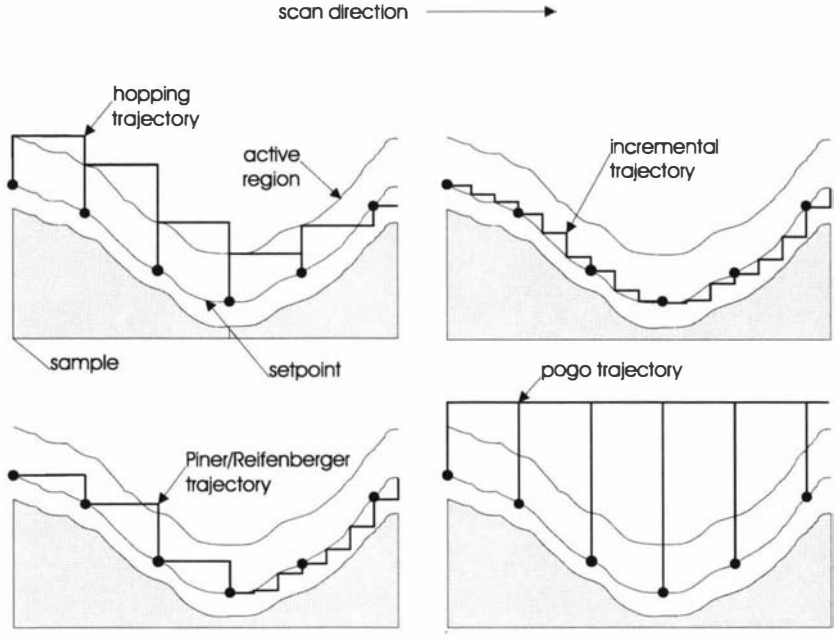


Figure 6.12: **Moving Trajectories:** Four panels show different ways of moving from point to point. Each panel shows the sample surface, the setpoint and the end of the active region. Sample points are indicated by dots. Scan direction for all panels is from left to right. The trajectories according to the different methods are drawn as thick lines. With the incremental trajectory, data can be recorded using either the mean value or sampling method. For the other three trajectories only the sampling method is possible. The conditions for these three methods are explained in figure 6.11.

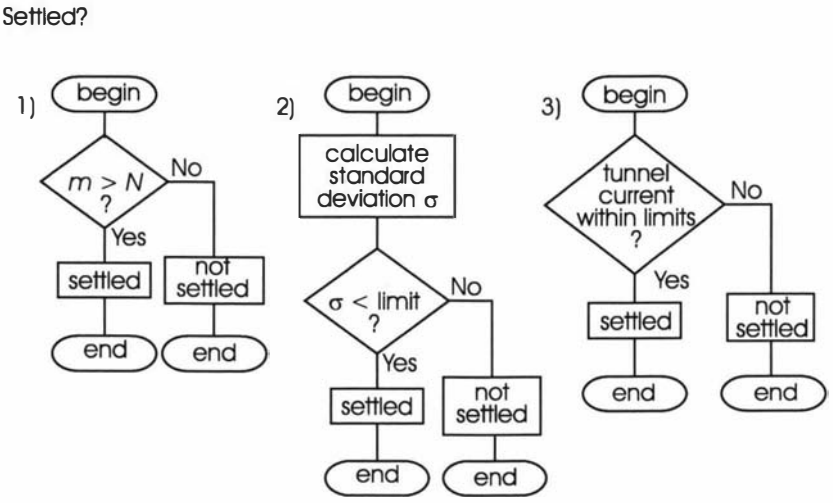


Figure 6.13: **Settling Options:** When averaging data at a sample point, the program waits until the data are settled. A few possible methods to establish stability are shown; calculating the standard deviation, waiting a certain number of cycles and making sure the current stays within a band of allowed values.

The interaction between tip and sample has a certain reach, which depends on the detection limit of the current-to-voltage converter, on the work functions of tip and sample and on the applied voltage. The interaction distance is indicated in figure 6.12 as the active region and is discussed in more detail in section 7.3.2. The hopping trajectory operates at the perimeter of the active region and is therefore furthest away from the sample, while still interacting with it. The pogo trajectory actually leaves the active region between sample points.

What all trajectories have in common is that at the sample point the program has to establish tunneling current stability. After the controller has adjusted the tip sample distance the program has to decide about when to move to the next sample point. There are several options of how to define whether the tunneling current has settled. Three different ones are shown in figure 6.13. The program may simply wait a certain time measured by a minimum number of cycles. This is option 1) in figure 6.13. The waiting time can also be made dependent on the data, for example by calculating the standard deviation or counting only data between an upper and lower limit. These are options 2) and 3) in figure 6.13. Choosing option 3) allows some data to be rejected, for instance the data values initially recorded after a position change. Option 2) measures the data quality directly. If the quality is consistent then option 1) is equally good as option 2) and needs less processing time. The standard deviation of the tunneling current has been measured under several conditions and it has been found that usually after 100 control-loop cycles the standard deviation did not change dramatically.

Using the sampling method with either the incremental or the P/R conditions causes the tip to move with variable horizontal speed across the sample. If the terrain is rough, each horizontal step will cause a large change in tunneling current. With the settling conditions 2) or 3) from figure 6.13, which depend on observing the quality of the tunneling current, a large change in current will take longer to reach the stability condition than a small change. Consequently

the tip will move faster over flat terrain than over rough terrain.

The mean value method uses separate horizontal and vertical motion of the tip when scanning a sample. Using the mean value method implies that stability of the tunneling current is not examined at each sample point, but that the tip moves with constant horizontal speed across the surface. During the linescan the tip-sample separation is controlled by the closed-loop control of the vertical tip motion. If the horizontal scan speed is not too fast, the trajectory will be equal to the incremental trajectory shown in the top right of figure 6.12. The DSP records data while moving across the sample surface. If too many data points are collected these can be reduced by averaging, on the other hand if more data points are required then the horizontal scan speed has to be lowered.

In section 7.4.2 the choice between the different methods is discussed with regards to tip safety, image acquisition speed and the nature of the sample.

6.5.4 Characteristic Curves

Characteristic Curves are briefly discussed in section 1.3.2. The tip remains at a set horizontal position, the bias voltage or height is changed and then the tunneling current is read. The process of recording characteristic curves is similar to taking a linescan. The flow chart of a linescan is shown in figure 6.9. Characteristic curves would usually be taken using a sampling method. Examples are shown in figures 8.1 and 8.2.

6.6 Step Response

Provision is made in the program to record the step response of the system. A flow chart is shown in figure 6.14. The step response routine cycles several times (N_{points}). After 8 cycles have passed, a step is added, which is kept for a time of N_{keep} cycles. The step can in principle be any microscope variable, tip height, bias voltage or setpoint. The step length is variable so that not only the step but also the impulse response can be measured. The operator can set the step length and height interactively. Tunneling current and the step are recorded in arrays. The values are added to a sum, since the step response routine can be repeated several times and an average taken. Averaging proceeds in a similar way to that depicted in figure 6.10. The step response can be recorded with or without feedback. If the open-loop response is to be measured, the tip height control block is skipped. After step response recording the feedback is switched back on automatically.

6.7 PC Program

6.7.1 PC Main Routine

The main PC routine is also written in C. A flow chart of the main routine is shown in figure 6.15. The routine serves mainly as a terminal so that the operator can control the microscope. The raster scan is automated, since it involves complex interaction between PC and DSP. This is only a first step towards a more complete automation that is possible with this system.

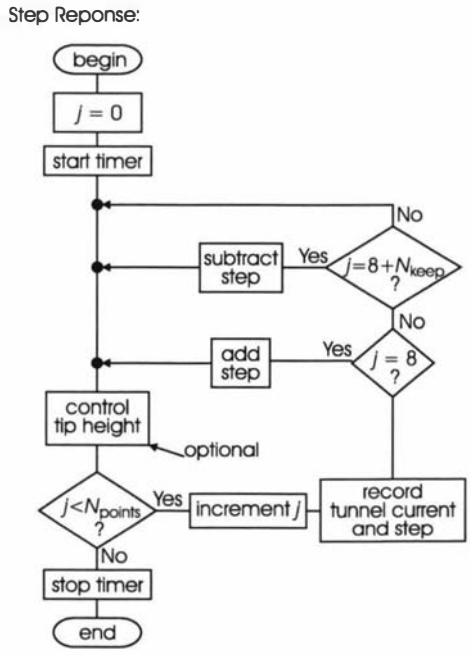


Figure 6.14: **Step Response:** The program records the response of the system by adding a step at $j = 8$ and keeping the step for a number of cycles (N_{keep}). The step can be a change in tip height, bias voltage or setpoint. By choosing $N_{\text{keep}} = 1$ the impulse response of the system can be measured. The variable N_{points} determines the length of the record. The step response routine is also used to time the DSP program.

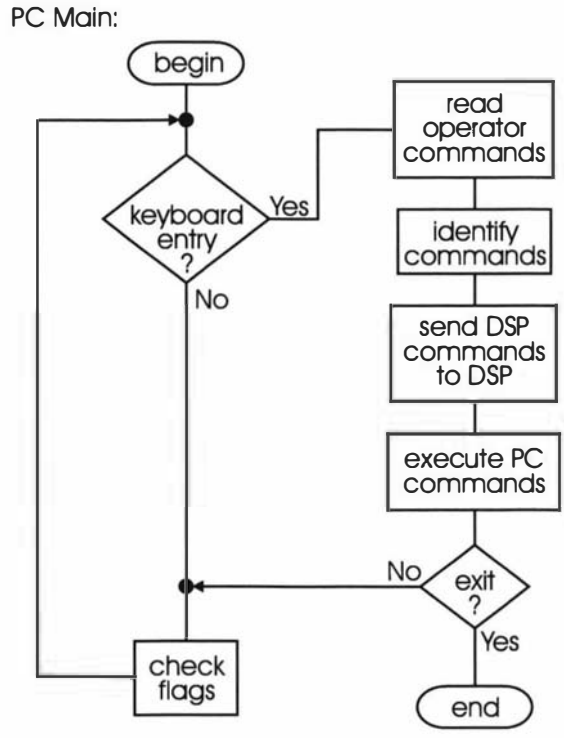


Figure 6.15: **PC Main Routine:** The main task of the PC main routine is to allow the microscope operator to change parameters and issue commands to the DSP. An operator command may be directed at the PC or the DSP. PC MAIN passes all DSP commands on to the DSP. Taking a raster scan calls for interaction between PC and DSP, which has been automated. To find out whether the DSP has completed a task, such as a linescan, PC MAIN checks a corresponding flag that has been set at the beginning of issuing this task. The operation to check flags in PC MAIN is currently used only for the raster scan.

During every execution loop, PC MAIN checks whether the operator issued a command. If a new command has been detected, PC MAIN decides on the recipient of the command and takes appropriate action. The recipient can be the DSP, the PC itself, or both. The program is ended by giving out an exit command, which is also sent to the DSP.

If a parameter, such as the bias voltage, is to be changed, the operator presses the appropriate key on the keyboard. PC MAIN prompts for the new value, reads it and sends it to the DSP in floating point format. The new value is read as a string and converted thereafter. If a conversion error occurs the old value is kept. Whether the value lies within the possible limits should also be checked, but has not been implemented so far.

6.7.2 Raster Scan

The DSP scans the surface in a raster fashion. The PC receives the scan command from the operator and passes it on the DSP. Depending on the desired size of the image a lot of data have to be collected and stored by the DSP. The DSP can hold a small image in on-board memory, however for a large image this memory would not be enough. Therefore both processors are programmed to transmit data after each linescan. The PC supervises the overall operation. The routine RASTER issues a linescan command to the DSP, waits until the scan has finished and finally reads and stores the linescan data. The PC main routine uses the same way of identifying a task issued by the operator as the DSP. A flow chart is presented in figure 6.3 to the left. The same flow chart shows how a task A is executed N_A times. For the raster scan this single task is replaced by two alternating tasks A and B as shown in figure 6.16, where issuing the linescan command is task A and reading and storing the data is task B. The main routine PC MAIN executes both tasks N_A times. Each task switches over to the other one after completion, which makes sure that the tasks alternate. For the raster scan, giving out the linescan command to the DSP is the task labeled A. Reception and storage of the linescan data is task B.

6.8 Image Display

Recorded images are currently displayed with a freely available program, SPIP written by Jan F. Jørgensen [108]. The STM group leader, Dr. Craig Eccles, is working on a piece of windows based image processing software to use under the Windows95 operating system. This new piece of software will also replace the currently used operator interface and will provide a platform to use the microscope leaving the basic tasks, such as data transfer, number of data points needed, tuning the controller to name only a few, to the software instead of the operator. Porter [157] reports a data acquisition system written in visual basic. He uses the Windows Application Program Interface to solve the problem that Windows95 is not a real-time operating system and directly interfacing to hardware is very complicated. Nakakura *et al.* [141] present a detailed discussion of imaging and control software based on Windows95 multimedia functions and virtual device drivers. Hoeven *et al.* [99] use a multiprocessor system for data acquisition, display and analysis based on Unix.

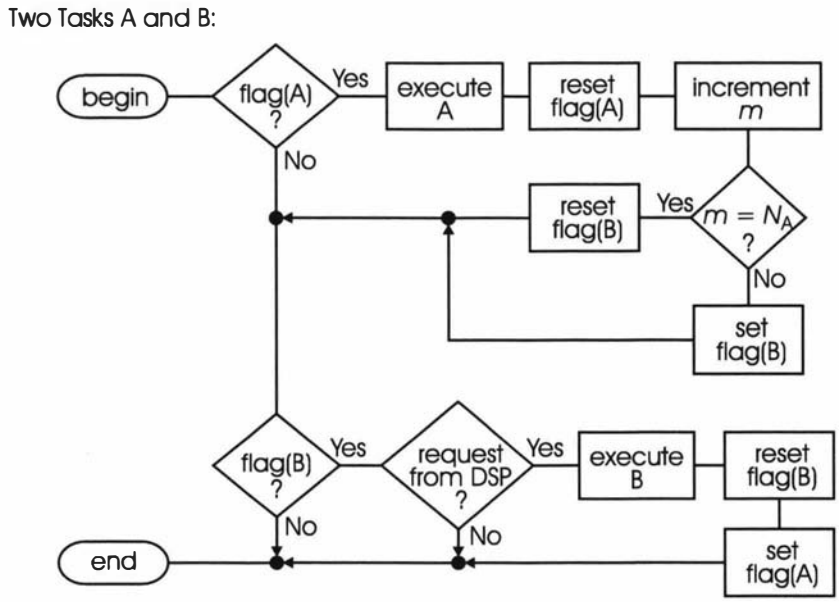


Figure 6.16: **Alternating Two Tasks:** The flow chart shows how one task (A) invokes a second task (B) and then waits until B terminates. In this case B itself waits for a response from the DSP program. Both tasks are executed N_A times. In this manner more than two tasks can be alternated as well. The two tasks A and B can also run simultaneously.

6.9 Control Algorithm

In section 4.1 the microscope as a control system is discussed. A control algorithm implemented in software holds the tip-sample distance constant using closed-loop control as mentioned in section 6.3.1. In this section the type of controller used is discussed first and then the discrete time version of this controller is subsequently presented.

6.9.1 PID Control

In this project the system uses a comparatively simple but very widespread control algorithm, the so-called PID controller [14, 24, 69, 187]. PID stands for the proportional, integral and derivative terms used in this controller. The input to the controller is the difference between the input signal and desired signal. This is called the error signal. The task of the controller is to make the error signal as small as possible.

A proportional controller outputs a control signal ($c(t)$) proportional to the error signal ($e(t)$)

$$c(t) = Pe(t)\,, \tag{6.2}$$

where P is the proportionality factor of the controller, which is just its forward gain. The magnitude of the forward gain determines how close the error signal will be to zero. If the forward gain is high, the error signal is close to zero. Note that the simple proportional controller can not achieve a zero error signal, because with a zero error signal, the control signal would be also zero, since the control signal output is proportional to the error signal.

To eliminate this offset an integrator can be added to the controller. The integrator contribution is

$$c(t) = I \int e(t)\,\mathrm{d}t\,, \tag{6.3}$$

with an integrator constant I. As long as there is an error, the integrator accumulates it and drives the control signals in such a way that the integrated error signal is zero. When the tunneling current is at the desired level, the error signal is zero and no correcting action is taken. Another way of understanding the integrator is to examine its gain and the effect of the gain on closed-loop accuracy given in equation (4.4). The integrator has a gain of $1/\omega$, i.e. an infinite gain at dc, which means that the integrator has a zero dc error.

A differentiator term puts out control signals proportional to the change in error signals,

$$c(t) = D\frac{\mathrm{d}e(t)}{\mathrm{d}t}\,, \tag{6.4}$$

with an differentiator constant D. This can be used to speed up the response of the controller, but also makes the system responsive to noise.

6.9.2 Discretization

The analog controller has to be put into a discrete value form before it can be implemented in software. The software reads a new input value, computes the error term and then puts out a new controller value. Usually this happens

once each sampling interval. At sample interval n the controller calculates the controller output c_n.

The control algorithm has to be a causal operator, which means that the current controller output c_n cannot be calculated from future values. Instead only present and past values are used for reconstruction. If the operator did not work on the most recent data then a certain amount of data could be stored and presented to the operator as future data. In this case an acausal operator for reconstruction could be used. This is rarely done, since the involved time delays are detrimental to the frequency response of the system and therefore usually not acceptable. Instead, a causal reconstruction method is used for real-time control, such as the zero order hold, and the controller uses the present value of the error (e_n) and in principle all past values of both controller, $c_{n-1}, c_{n-2}, \cdots$ and error, $e_{n-1}, e_{n-2}, \ldots$. If the controller uses past values of its own output it is called a recursive or autoregressive operator. If it uses only present and past values of the error then it is said to be of the moving average type.

In this system the software reads a process variable that is at least one sampling interval old, which means that the calculated error at sample interval n is actually e_{n-1} and not e_n. Suggestions of how to deal with this situation are discussed below. For now it will be assumed that e_n is available.

In addition to being causal, the control operator should only use a small number of past values. Using only low order algorithms for control brings three advantages with it, increased stability, shorter processing time and a better handling of transients [45]. The fewer the number of past values, the faster the controller can react to a change in the process variable. According to Chesmond [45] the phase margin of a controller using few values is better than the phase margin of a controller using many values. The greater the number of past values used effectively increases the sampling time, which causes a larger time delay and thereby an increasing phase shift with frequency. Despite the fact that only few values should be used for the controller, a greater number of values can be used for data averaging and subsequent storage at one sample point above the surface.

Discrete PID Controller

The proportional output, equation (6.2), is translated into a discrete form by

$$c_n = Pe_n.$$

Because it does not involve past values of the controller, the proportional operator is classified as a moving average type operator.

For the integral output the integration has to be approximated by a discrete expression. One of the simplest ways to approximate integration is to replace the integration by summation using first order finite differences such as

$$y_n = t_s \sum_{i=0}^{n} x_i,$$

where t_s is the sampling interval and x_n and y_n represent the input and output values of the operator in the sampling interval n. This summation is also a moving average operator, but can be expressed in recursive form as

$$y_n = y_{n-1} + t_s x_n, \tag{6.5}$$

which makes it an autoregressive operator. As it stands equation (6.5) is the rectangular rule of integration [45]. The output of an integral controller corresponding to equation (6.3), is then

$$c_n = c_{n-1} + It_s e_n.$$

Integration should not be formulated as a moving average operator, since this would involve a great number of past values, which is undesirable.

One type of derivative operator is

$$y_n = \frac{1}{t_s}(x_n - x_{n-1}).$$

This way of approximating a differential is called the backwards difference operator. It is of the moving average type. The output of a purely derivative controller like equation (6.4) would be

$$c_n = \frac{D}{t_s}(e_n - e_{n-1}).$$

In summary, the overall PID operator is given by two equations

$$s_n = s_{n-1} + e_n, \tag{6.6}$$

$$c_n = Pe_n + It_s s_n + \frac{D}{t_s}(e_n - e_{n-1}), \tag{6.7}$$

where equation (6.6) is the autoregressive integral operator, which operates on the sum (s_n) of the error terms. The PID operator in equation (6.7) is known as the position form of PID operator, since c_n is proportional to the actuator position.

The equations 6.6 and 6.7 can be re-cast into velocity form, which means that during each sampling interval only the increment (Δc_n) in actuator position is calculated, i.e.

$$\Delta c_n = c_n - c_{n-1}.$$

To find the difference the previous actuator position is found by delaying equation (6.7) by one and then subtracting the previous from the current value. This gives

$$\Delta c_n = P(e_n - e_{n-1}) + It_s(s_n - s_{n-1}) + \frac{D}{t_s}(e_n - 2e_{n-1} + e_{n-2}). \tag{6.8}$$

The difference in the sum of the errors can be found from equation (6.6) and substituted into equation (6.8) to find

$$\Delta c_n = P(e_n - e_{n-1}) + It_s e_n + \frac{D}{t_s}(e_n - 2e_{n-1} + e_{n-2}). \tag{6.9}$$

This PID operator with $D = 0$ is the one used in the current program.

The velocity form of PID operator is required so that other routines can easily replace the controller in the main routine as discussed in section 6.3.1. The controller uses the velocity operator and delivers only the change in actuator position Δc_n to the main routine.

Equation (6.9) can be ordered by the error terms as

$$\Delta c_n = (P + It_s + \frac{D}{t_s})e_n - (P + 2\frac{D}{t_s})e_{n-1} + \frac{D}{t_s}e_{n-2}. \tag{6.10}$$

Generally speaking the resulting operator is of the form

$$c_n = F_1 c_{n-1} + S_3 e_n + S_2 e_{n-1} + S_1 e_{n-2}, \tag{6.11}$$

which is an autoregressive moving average operator. The F_1 constant is the single coefficient of the autoregressive scheme and S_1, S_2 and S_3 are the coefficients of the moving average operator. Note, that the moving average operator can be made even shorter, depending on the P, I and D terms. By dropping the derivative action, the operator is made reasonably short, requiring only values from the previous interval. If the controller only operates with the I term then the velocity form of the PID operator is as short as possible with only one moving average operator coefficient. The PID operator in velocity form is by design not meant to be used without the integral action, which is acceptable since an I term alone will give stable control with the current set-up, but a P term alone will not.

With regards to the D term, Åström and Hägglund [14] state that the backwards difference operator for the derivative term is stable and gives good results for all derivative factors D. Often, however, the derivative term is manipulated to be less influential at high frequencies or it is completely omitted, because it enhances fast changes thereby increasing high frequency noise.

Possible Improvements to the PID Controller

Numerical integration is not as problematic as differentiation. For increased integration accuracy, approximations of higher order than the rectangular integration rule may be used. Higher-order integration rules can be found from the Newton-Cotes formulae [35]. The first three integration rules following the Newton-Cotes formulae are known by the names trapezoid rule, Simpson's rule and Newton's 3/8 rule. The trapezoid rule is

$$c_n = c_{n-1} + \frac{t_s}{2}(e_n + e_{n-1}),$$

and Simpson's rule is

$$c_n = c_{n-2} + \frac{t_s}{3}(e_n + 4e_{n-1} + e_{n-2}). \tag{6.12}$$

The trapezoid rule is able to integrate straight lines exactly, while Simpson' rule integrates polynomials up to third order exactly [87]. To use Simpson's rule in a control algorithm, the penultimate controller output c_{n-2} in equation (6.12) can be replaced by the previous controller output c_{n-1} [45].

Using the trapezoid rule instead of the rectangular rule results in a PID operator of the form

$$\Delta c_n = (P + \frac{1}{2}It_s + \frac{D}{t_s})e_n - (P - \frac{1}{2}It_s + 2\frac{D}{t_s})e_{n-1} + \frac{D}{t_s}e_{n-2}. \tag{6.13}$$

This operator is an autoregressive moving average operator of the same type as in equation (6.11), only with different coefficients. Note, that if it is decided to

use a P term, i.e. having $P \neq 0$, then the PI operator based on the trapezoid rule has the same length as the PI operator described by equation (6.10).

The same calculation using Simpson's rule gives an autoregressive moving average operator with three moving average coefficients and takes the following form:

$$\Delta c_n = (P + \frac{1}{3}It_s + \frac{D}{t_s})e_n - (P - \frac{4}{3}It_s + 2\frac{D}{t_s})e_{n-1} + (\frac{1}{3}It_s + \frac{D}{t_s})e_{n-2}. \quad (6.14)$$

6.9.3 Predictor Type Controller

According to Åström and Hägglund [14] PID control is generally sufficient if the controlled process is of second order, which is the case for the mass-spring system formed by the tip holder and the center actuator. Only systems of higher order and systems that exhibit dominant time delay will benefit substantially from using a more sophisticated control than PID.

As discussed in section 4.4, in this system, sampling proceeds at the fastest possible rate so as to achieve a high enough bandwidth for the closed-loop controller. As a consequence only the previous error term (e_{n-1}) and not the current error term (e_n) is available for calculations, which results in a time delay of one sampling interval. The time constant of the controlled process is currently given by the current-to-voltage converter. With a faster converter, it would be given by the lowest vertical resonance frequency of the probe head assembly. Both time constants are around

$$\tau = \frac{1}{2\pi \times 10\,\mathrm{kHz}} = 16\,\mu\mathrm{s},$$

which means that the time delay of one sampling interval, currently 12.8 μs, is comparable but not dominant. The time constants of this system are further discussed in section 7.2.

Åström and Hägglund point out that PI control alone is sufficient if the process is of first order or if the demand on the controller is not too high. Most SPM research groups report the use of a PI or only I type of controller [5, 88, 99, 78, 117, 141, 145, 164, 191]. This is true also for the direct digital controllers [20, 134, 149, 167, 192, 194]. In this project a PI controller as in equation (6.7) without D term was established ignoring the single time delay. However, the time delay can be taken into account as discussed in the following paragraphs.

If the PID operator is calculated from the last two available error terms, then the operator in equation (6.11) is of the following kind

$$c_n = F_1 c_{n-1} + S_2 e_{n-1} + S_1 e_{n-2}, \quad (6.15)$$

where the moving average operator is equal to a three-term moving average

$$c_n = F_1 c_{n-1} + S_3 e_n + S_2 e_{n-1} + S_1 e_{n-2},$$

with S_3 set to zero.

The easiest way to approach this situation is to ignore the delay when calculating the controller output. The operator in equation (6.15) is equal to the one in equation (6.11) delayed by one sampling interval. This is equivalent to

multiplication by the complex number Z in the Z transform picture of the operator. The frequency response remains the same in magnitude, but suffers an extra phase shift given by [110]

$$Z = e^{-j\omega t_s}.$$

Another way to approach the problem would be to use a modified operator. The operator in equation (6.15) is of the predictor type. A predictor is an operator that uses a previous input sample together with the last output sample to calculate the current output [45]. A simple predictor type of integrating operator is the Euler algorithm

$$y_n = y_{n-1} + t_s x_{n-1},$$

which is simply the rectangular rule delayed by one sampling interval and therefore equal to just ignoring the delay. The shortest possible control algorithm is the I only operator from equation (6.10), but based on the Euler rule. It is

$$c_n = c_{n-1} + I t_s e_{n-1}.$$

A predictor type integration rule of second order is the Adams rule

$$y_n = y_{n-1} + \frac{t_s}{2}(3x_{n-1} - x_{n-2}).$$

This rule tries to predict the current value by extrapolating from a straight line [45]. Using the Adams rule together with the delayed proportional and derivative terms will not result in a longer operator than the PID operator in equation (6.13) delayed by one interval. The operator should perform better, since prediction is incorporated. The resulting operator is given by

$$\Delta c_n = (P + \frac{3}{2} I t_s + \frac{D}{t_s}) e_{n-1} - (P + \frac{1}{2} I t_s + 2\frac{D}{t_s}) e_{n-2} + \frac{D}{t_s} e_{n-3}. \qquad (6.16)$$

The operator can also be expressed as [187]

$$\Delta c_n = K \left[(1 + \frac{3}{2}\frac{t_s}{t_i} + \frac{t_d}{t_s}) e_{n-1} - (1 + \frac{1}{2}\frac{t_s}{t_i} + 2\frac{t_d}{t_s}) e_{n-2} + \frac{t_d}{t_s} e_{n-3} \right], \qquad (6.17)$$

where two time constants t_i and t_d have been introduced. The constant K and the time constants are related to the constants P, I and D by

$$K = P, \quad \frac{1}{t_i} = \frac{I}{K}, \quad t_d = \frac{D}{K}.$$

Re-organizing did not change the control law in principle, but makes it easier to interpret. The forward gain of the controller is given by K. The integrator time constant t_i is the time it takes for the integrator term to grow as large as the proportional term. The derivative time constant t_d can be interpreted as the time taken by the controller to predict future values from the currently known slope.

The derivative term need not be used. If it is, then the controller can be modified to prevent differentiating a setpoint change [24]. While the operator is still in the form of equation (6.9), the error term

$$e_n = s_n - u_n$$

is replaced by

$$e_n = s - u_n,$$

where s is now a constant setpoint. Differences in error terms now turn into

$$e_n - e_{n-1} = s - u_n - (s - u_{n-1}) = u_{n-1} - u_n$$

without involving the change in setpoint. Only the integral term still contains the setpoint, as can be seen for example with the Adams rule

$$\frac{1}{2}(3(s - u_{n-1}) - (s - u_{n-2})) = s - \frac{3}{2}u_{n-1} + \frac{1}{2}u_{n-2}.$$

Having the setpoint only in the integrator and not in the proportional and derivative terms allows for a smooth setpoint change. Equation (6.17) is written as

$$\begin{aligned} \Delta c_n &= K[(u_{n-2} - u_{n-1}) + \frac{t_s}{t_i}(s - \frac{3}{2}u_{n-1} + \frac{1}{2}u_{n-2}) + \\ &\quad + \frac{t_d}{t_s}(2u_{n-2} - u_{n-1} - u_{n-3})], \end{aligned} \tag{6.18}$$

In the future it might be interesting to explore the controller of equation (6.18).

6.9.4 Software Control

There are other advantages which software control can provide. Some examples are discussed in the following paragraphs.

The software can easily incorporate non-linearities, such as the tunneling current, which depends exponentially on the tip-sample distance. This relationship can be linearized in software using the log function. The logarithm is available as a subroutine optimized for the ADSP-21020 DSP. The necessity for including dedicated hardware to perform this function can then be removed.

The software can also deal with actuator saturation. The piezoelectric actuator can only extend to a certain length determined by the maximum voltage delivered by the HV modules. When the actuator reaches this extension, the error term, being non-zero, begins to grow due to the integrator action. This situation is known as integrator wind-up. In case the input comes close to the setpoint again the high value of the integrated controller output has to be slowly decreased. This is not acceptable for the center actuator because such a situation would most likely lead to a tip crash. If the tip does not crash then integrator wind-up will result in an overshoot. To prevent wind-up the software limits the controller output to a maximum value after which the integrator output is held constant.

The software can also adapt control parameters according to a given control situation. For example, while the tip is approaching, the bias voltage and setpoint are set so that there is a large tunneling gap. After tunneling current has been found, the desired setpoint can then be set automatically.

The literature also discusses the transfer between automatic and manual control. This is not an issue in this project, since manual control is impossible.

6.9.5 Parameter Tuning

The three independent parameters of the PID controller need to be set to reasonable values. A high forward gain is highly desirable to achieve accurate closed-loop control. However, the gain cannot be freely chosen due to the possibility of instability. If the forward gain is too high, the controller will act violently to small changes and may even oscillate. In the worst case the tip will crash into the sample.

There are three ways to tune the parameters of the controller, by trial and error, from a mathematical model or semi-empirically [187].

It is impossible to record an open-loop response of the system without being able to control the tip-sample distance in the first place [149]. When using the control system the first time reasonable P and I control terms had to be guessed.

After closed-loop control was established, the response of the system to a change in setpoint could easily be studied by changing the setpoint after a certain time and recording the resulting reaction of the controller and the tunneling current. Inspecting the step response was used as a monitoring tool for manual trial and error tuning. The P and I terms used with the program were initially tuned in closed-loop by bringing the system just into oscillation and then reducing both parameters. The step response was checked on overshoot and the parameters fine tuned.

Apart from the trial and error method others are possible, but have not been pursued in detail yet. Parameters for PID control can be found using the semi-empirical Ziegler-Nichols procedures [15, 17, 24, 124, 188]. There are two methods that Ziegler and Nichols recommend, a closed-loop and an open-loop method. The open-loop method relates the tuning parameters to the inverse of the dead time of the step response. While trying to measure the actual dead time it was found that it was small compared to the measured time lag, but that the uncertainty in the dead time was high due to noise. The calculated tuning parameters are much higher than the ones found by trial and error, but the uncertainty is high, too.

The closed-loop method involves finding the oscillation period of the closed loop with proportional control only. This can be done if a position form of the PID control algorithm is used instead of the velocity form, since in the velocity mode the integrator cannot be switched off.

The open loop step response could be used in conjunction with a mathematical model of the overall process to find better controller parameters. A mathematical model is in principle available from the equations for the tunneling current, the piezoelectric actuator, the current-to-voltage converter and the rest of the electronics.

Two research groups reported the self-optimization of digital PID controllers with an evolutionary algorithm. This algorithm needs no detailed knowledge of the controlled process. Scheuring *et al.* [167] managed to improve the algorithm previously published by Baselt *et al.* [20]. Paillard *et al.* [149] discuss parameter tuning of a computer controlled STM in detail. They derived the open-loop behavior of the system from the closed-loop behavior. With the knowledge of the open-loop behavior they could tune the controller by simulation. The use of simulation together with a hill-climbing algorithm to find good control parameters is also suggested in the book of Leigh [124].

6.9.6 Outlook

Because the control algorithm is in software it can easily be changed. The control parameters can be optimized depending on the given situation. The algorithm can also be adapted to other types of scanning probe microscopy. Other types of controllers, which could be more complicated but faster, can be implemented. Usually this requires an in-depth knowledge of the complete system. In figure 4.5 an appropriate box labeled 'System Information' is indicated. If a model of the tip-sample system is available, the outcome of the calculated actions the controller wants to take can be anticipated and taken into account.

Chapter 7

System

7.1 Introduction

This section looks at the behavior of the scanning probe microscopy system as a whole. In particular, the speed and noise performance of the system and the tip-sample approach are examined in detail.

7.2 System Speed

An important consideration for feedback control is the time it takes for the system to respond to a change. This change might be a modification of the setpoint by the user, or a tip-sample distance variation caused by external disturbances, or simply changes in the surface height. When the tip-sample distance alters, the change is registered by the controller after the tunneling current is converted to a voltage, low-pass filtered and converted into a digital word. The computed reaction of the software controller is represented by another digital word, which is converted to an analog voltage, low-pass filtered and applied to the center actuator. In other words the time lag of the analog signal processing, as well as the dead time of the system have to be considered when assessing the system speed.

7.2.1 System Dead Time

In a feedback system, gain and bandwidth are closely related. If the system response time consists only of an analog time lag then a higher controller gain results in a faster response. A digitally induced time delay (dead time[1]), however, cannot be overcome by negative feedback. The effect of dead time was discussed in section 4.4. Apart from the PID controller discussed in section 6.9.3, one possible way of taking dead time into account is to incorporate a Smith predictor into the controller [15, 45, 186].

Sampling Time In the specifications defined in section 4.1 a sampling time of 10 μs was planned. The actual sampling time used was 12.8 μs, which is equal to the conversion time of the analog-to-digital converter, t_{adc}.

[1] Dead time is sometimes defined as pure time delay to distinguish it from time lag.

Dead Time In section 4.4, equation (4.9), the overall dead time of a sampled-data system is given as

$$t_{\mathrm{d}} = t_{\mathrm{adc}} + t_{\mathrm{dac}} + t_{\mathrm{program}} ,$$

where t_{adc} is mentioned above, t_{dac} is the conversion time of the digital-to-analog converter and t_{program} is the program time. The conversion time for the digital-to-analog converter is $t_{\mathrm{dac}} = 4\,\mu\mathrm{s}$.

The program time consists of processing time and synchronization time. Processing time is the time to read and write the input and output channels and compute the controller action and takes only a fraction ($< 1\,\mu\mathrm{s}$) of the conversion time of the analog-to-digital converter t_{adc}.

A drawback with the current system is an extra dead time introduced due to a lack of synchronization. The time between the instant a datum arrives in the input register and the instant the program reads from the register cannot be predicted, since the conversion command for the analog-to-digital converter is generated independently of the software. As discussed in section 5.4.5, the detection of an event at a random instant in time will not be sped up by synchronization, but the programming dead time t_{program} will, since the synchronization time vanishes. If the program waits a time t_{adc} then a new datum will definitely have arrived. Let the program inputs be denoted v_n. As a consequence of dead time the program will, in the best case, read sample v_{n-1} in the sampling interval n. In the worst case it will read input sample v_{n-2}. Hence the maximum dead time in the current system is $t_{\mathrm{d,max}} = 2 \times t_{\mathrm{adc}} + t_{\mathrm{dac}} = 29.6\,\mu\mathrm{s}$. The minimum dead time in the current system is $t_{\mathrm{d,min}} = t_{\mathrm{adc}} + t_{\mathrm{dac}} = 16.8\,\mu\mathrm{s}$. Since the DSP real time clock is not synchronized with the DSP system clock, they will drift and the actual dead time will beat between those extremes. Measured in sampling times this is between 1.3 and 2.3 intervals, which gives an average of 1.8.

The analog-to-digital converter read operation can be synchronized and shortened as mentioned in section 5.4.5. The maximum dead time would then be slightly more than $t_{\mathrm{md}} = 10.4\,\mu\mathrm{s} + 4\,\mu\mathrm{s} = 14.4\,\mu\mathrm{s}$ depending on the processing time. The sampling time could be chosen as $t_{\mathrm{s}} = t_{\mathrm{md}} = 10.4\,\mu\mathrm{s}$, where the program reads input sample v_{n-1} reliably in interval n.

7.2.2 Measured Response Time

The response time of the system was measured by programming an output voltage step and recording the change in input voltage. The response time was measured with the HV module connected directly to the ADC module, as well as with the usual set-up. The tip was however not in tunneling mode. The direct measurement shows that after 3 sampling periods a voltage change is recorded. This delay was not unexpected.

The time lag of the HV module is given in section 5.4.6 and is repeated in table 7.1. It takes $13\,\mu\mathrm{s}$ for the output of the HV module to rise to 90% of its final value. Together with the conversion time $t_{\mathrm{dac}} = 4\,\mu\mathrm{s}$ this is $17\,\mu\mathrm{s}$. The time lag of the ADC module is also listed again in table 7.1. The low-pass filter of the ADC module adds a minimum of $10\,\mu\mathrm{s}$, which is its step rise time. The fact that the input is not a step anymore is disregarded.

Let the sampling intervals be counted by the index n. The output of the HV module is started at the beginning of interval $n = 0$. The voltage step arrives

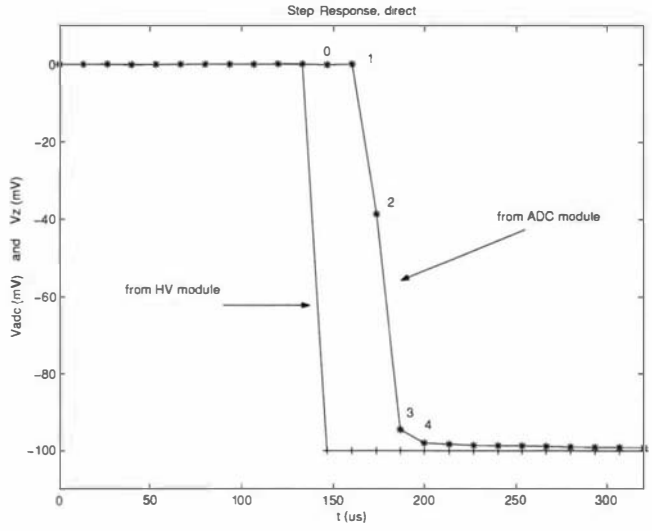


Figure 7.1: **Direct Step Response:** The output of the HV module was connected directly to the input of the ADC module. The sampling time was 12.8 μs. The output was changed at interval $n = 0$. The program registered the full step at interval $n = 4$, but already at $n = 2$ an initial response can be seen.

at the analog-to-digital converter after $(17 + 10)\,\mu\text{s} = 27\,\mu\text{s}$, which is just before sampling interval $n = 3$. The program can see v_3 at best in interval $n = 4$. It will definitely see it in interval $n = 5$. The earliest possible change can be seen in interval $n = 2$ for minimum dead time and $n = 3$ for maximum dead time. Judging from a maximum dead time of 2.3 sampling intervals, the chances are 70% that a change is seen already after 2 sampling intervals, which is the case in figure 7.1. The result of the second experiment is shown in figure 7.2. The input voltages are due to capacitive coupling of the step voltage applied to the center actuator. A change in voltage is first registered at interval $n = 4$, which gives a response time of $4 \times 12.8\,\mu\text{s} = 51.2\,\mu\text{s}$. Because of capacitive coupling the time it takes the program to read the input voltages is shorter than it would be with a system of essentially low-pass filter behavior. The time period of the observed oscillation is about $T = 205\,\mu\text{s}$, which corresponds to a frequency of $f = 4.9\,\text{kHz}$. This is in accordance with the peak frequency of the capacitive coupled current-to-voltage converter of $f = (5 \pm 0.5)\,\text{kHz}$ in figure 5.14. It can be concluded that the step response in this set-up is dominated by the current-to-voltage converter.

Ramp Response Time When the tip is approaching the sample, the program has to detect sample proximity early enough to prevent a tip crash. The tunneling current will rise exponentially when the tip comes into tunneling distance. In the following analysis, the rise in current is approximated as a linear ramp. The question of how much time is needed for a RC low-pass filter to respond to a ramp input is answered in section B.1.3 as being the time constant τ. As an estimate, the low-pass filters of the ADC and HV module are also assumed to follow an input ramp with a delay τ. The time constants of the modules are $\tau_{\text{hv}} = 5.7\,\mu\text{s}$ and $\tau_{\text{adc}} = 4.3\,\mu\text{s}$.

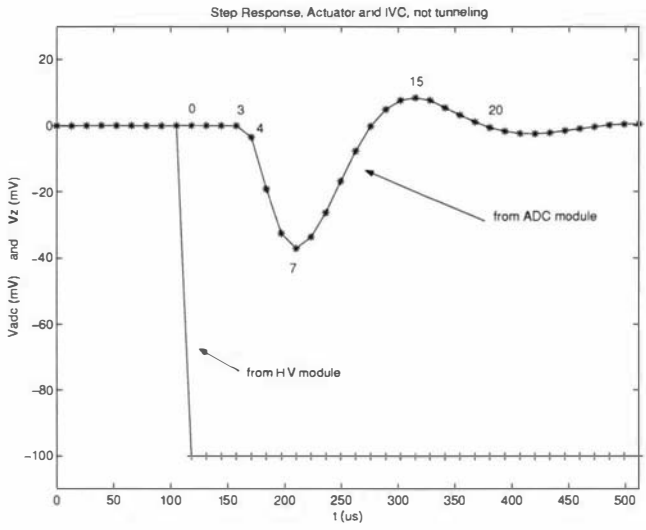


Figure 7.2: **Step Response via IVC:** A step response similar to the one in figure 7.1 was recorded, having the same sampling time. The output of the HV module was connected to the center actuator with the tip far away from the sample. An input current was measured that was coupled capacitively from the actuator electrodes to the current-to-voltage converter. The first reaction to the change in output voltage is seen in interval $n = 4$. The period of the observed oscillation is approximately $T = 16 \times 12.8\,\mu\text{s} = 205\,\mu\text{s}$.

Time constants for the current-to-voltage converter and the center actuator are also estimated. Both are second-order systems having a peak at a resonance frequency f_{res}. As discussed in appendix B.1.3, a time constant can be derived from the resonance frequency as

$$\tau = \frac{1}{2\pi f_{\text{res}}}.$$

This results in

$$\tau_{\text{ivc}} = \frac{1}{2\pi\, 5\,\text{kHz}} = 32\,\mu\text{s}$$

for the current-to-voltage converter and

$$\tau_{\text{ca}} = \frac{1}{2\pi\, 44\,\text{kHz}} = 3.6\,\mu\text{s}$$

for the center actuator. These values are listed in table 7.1.

The sum of the time constants of HV module, current-to-voltage converter, center actuator and ADC module results in a time of

$$t_{\text{ramp}} = (5.7 + 32 + 3.6 + 4.3)\,\mu\text{s} = 46\,\mu\text{s}.$$

If the start of the voltage output is again at the beginning of interval $n = 0$, then t_{ramp} falls into interval $n = 3$, that is before interval $n = 4$ and can therefore be registered by the program only in interval $n = 5$ or in the worst case $n = 6$. The most straight forward way to speed this process up would be to use a faster current-to-voltage converter.

	τ (μs)
Center Actuator, longitudinal	3.6
Center Actuator, lateral	23
Probe Head, longitudinal	14
Probe Head, lateral	66
IVC Module	32
ADC Module	4.3
HV Module	5.7
Bias Module	22

Table 7.1: **Time Constants:** The table shows an overview of all important system time constants, the mechanical ones in the upper half, the electrical ones in the bottom half.

7.2.3 Modeled Bandwidth

The bandwidth of the complete system was also modeled using the results from sections 3.3.5, about probe head vibrations, 5.4.2, about the current-to-voltage converter, 6.9.2, about the digital controller, 5.4.5, about the ADC module and 5.4.6, about the HV module. The resulting Bode plot is shown in figures 7.3 and 7.4. A few assumptions had to be made, which are listed in the caption together with other model parameters. The gain is shown in figure 7.3. The phase is shown in figure 7.4. To find the response of the digital controller for the model, the Z transform of the digital control law from equation (6.9) was taken. For an arbitrary sample value x_n and its predecessor x_{n-1}, taking the Z transform can be described as [110]

$$x_{n-1} = Zx_n. \tag{7.1}$$

To find the spectrum of an operator in the Z domain the Z transform is evaluated on the unit circle [110], which means Z is replaced by $e^{-j\omega t_s}$.

The controller from equation (6.9) is here repeated as a PI controller with $D = 0$:

$$c_n = c_{n-1} + Pe_n - Pe_{n-1} + It_s e_n.$$

Applying equation (7.1) on the PI controller gives

$$(1 - Z)c_n = ((P + It_s) - PZ)\, e_n,$$

from which the Z transform of the controller is found to be

$$\frac{c_n}{e_n} = \frac{(P + It_s) - PZ}{1 - Z}. \tag{7.2}$$

To find the frequency response,

$$Z = e^{-j\omega t_s}$$

is inserted in equation (7.2). In the model, an extra delay of one sampling interval is taken into account by multiplying the error term e_n with Z according to equation (7.1). This results in an extra factor Z in equation (7.2). In the model the P term was set to zero and an I term of $I = 6250$ was used, which were the conditions often used in the experiments.

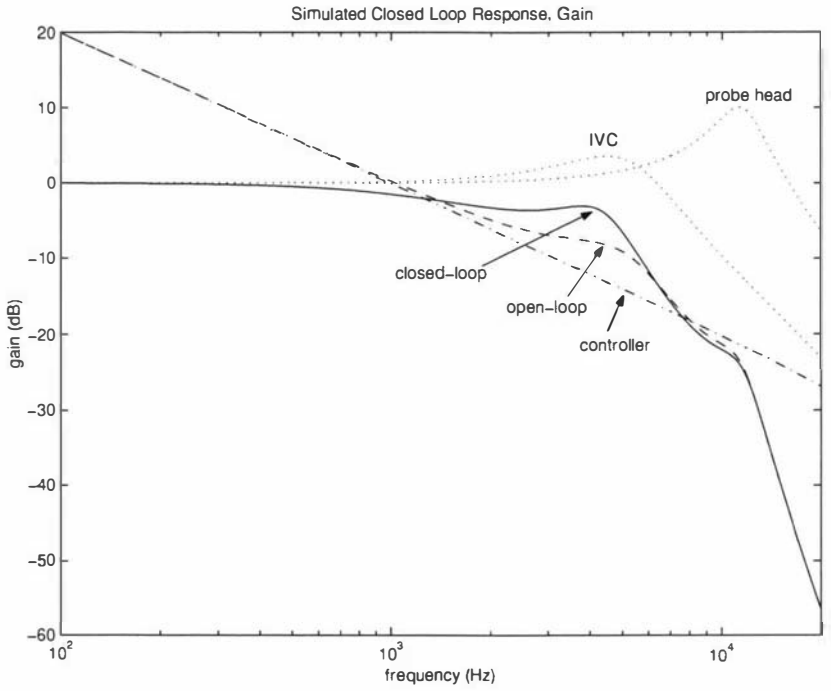


Figure 7.3: **Simulated System Response, Gain:** The magnitude of the simulated response of the system is shown. The probe head response is modeled as being a simple spring mass system resonating at $11.5\,\mathrm{kHz}$, which is the lowest vertical frequency of the probe head. The δ value describing the resonance width was taken to be $\delta = 11.5 \times 10^3\,\mathrm{s}^{-1}$. The frequency response of the current-to-voltage converter is taken from the model derived in section 5.4.2, with $f = 5.2\,\mathrm{kHz}$ and $\delta = 11.5 \times 10^3\,\mathrm{s}^{-1}$. The controller was the digital PID controller from equation (6.9) with $I = 6250$ and $P = D = 0$. The frequency response of the controller was found from the Z transform by inserting $Z = \exp(-j\omega t_s)$ as described by Karl [110]. The anti-aliasing filter at $50\,\mathrm{kHz}$, the reconstruction filter at $100\,\mathrm{kHz}$ and the zero order hold response from equation (4.11) were also taken into account, although clearly their influence is negligible. The open-loop gain is the loop gain of the control loop, i.e. it is the product of all transfer functions.

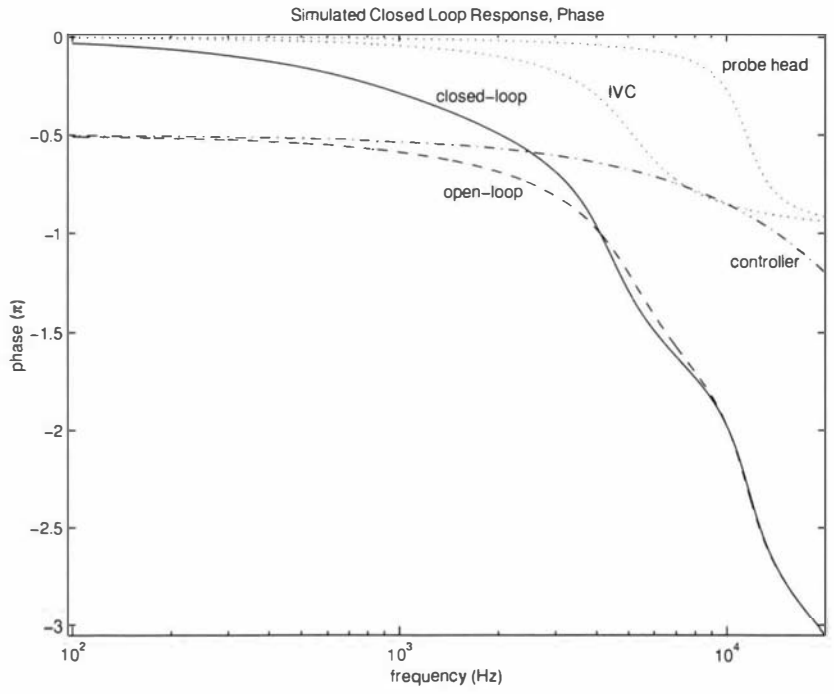


Figure 7.4: **Simulated System Response, Phase:** The phase of the simulated response of the system is shown. The conditions are equal to the ones in figure 7.3.

In this model the open loop gain crosses 0 dB at about 1.1 kHz and has a phase of $-180°$ ($-\pi$ in the graph) at about 4.5 kHz. From these values a gain margin of 8 dB and a phase margin of 67° can be deduced. This fits well with the recommendations mentioned in section 4.2.2 of either 8 dB for the gain margin or 30° for the phase margin. As can be seen in figures 7.3 and 7.4, in this system neither the probe head nor the digital system is the bandwidth limiting factor, but the current-to-voltage converter is. The model correctly predicts the maximum bandwidth of the system, since it will start to oscillate in practice if the I term is increased slightly beyond the value used in the model.

7.3 Step Size of Tip-Sample Approach

7.3.1 Automatic Approach

The automatic approach algorithm of the tip to the sample is discussed in section 6.4. A flow chart is shown in figure 6.4. For convenience the steps involved are repeated here.

The tip is approached by repeating the following steps:

1. retracting the tip fully,

2. approaching the microscope by stepping down the ramp,

3. extending the tip towards the sample and

(a) either activating the feedback control if tunneling current has been found or

(b) if no signal is found, start over.

Another method of tip approach that had initially been tried is to step the probe head down the ramp with closed-loop control active, until a small tunneling current is encountered. The probe head takes small steps, so as to give the closed-loop control a chance to react.

The first of these approach methods has two advantages, the tip is less likely to crash and it is faster. The decreased crash risk is a consequence of approaching the probe head with the center actuator retracted as explained below.

The sample holder ramp is optically smooth, but not smooth on the nanometer scale. The legs of the probe head encounter hills and valleys several nanometers in height while moving on the ramp. If the tip is not retracted while the probe head moves, the chances are high that the tip will impact the sample before it can be retracted. A second issue is the abruptness of the stick-slip movement of the probe head, which excites the resonant vibrations of the probe head. If the amplitude of the resulting vibration is large enough, the tip will make contact with the sample.

7.3.2 Approach Speed

In the following sections the approach speed for both methods is calculated, starting with the first method.

The step size of the coarse approach mechanism must be smaller than the maximal extension of the tip, which is $\pm 0.64\,\mu\text{m} = 1.3\,\mu\text{m}$. The time it takes for one complete approach step is the sum of the times to retract and extend the tip and the time to step down the ramp.

Probe Head Approach

To avoid taking overly large steps as the probe head moves down the ramp, a number of smaller steps are taken. In section 3.1.3 the step size of the probe head is discussed. A typical small step would result in a downward movement of 19 nm. Choosing $1.0\,\mu\text{m}$ as the maximum vertical probe head movement is less than the maximal extension and leaves a certain margin of movement for the feedback control. In this case the probe head will take

$$\frac{1.0\,\mu\text{m}}{19\,\text{nm}} = 53$$

such steps. The stick-phase of each typical step lasts 2 ms, which results in a time of $53 \times 2\,\text{ms} = 110\,\text{ms}$ for all steps. The microscope needs some time for the tunneling current to settle down after each set of steps, because every step of the microscope results in a surge in tunneling current caused by electro-mechanical coupling during the abrupt slip phase of the microscope movement. After all the steps have been taken, a 15 ms delay is added to allow the tunneling current time to settle down. The total time is therefore $(110 + 15)\,\text{ms} = 125\,\text{ms}$.

Center Actuator Approach

The center actuator carries the tip. If the tip advances too rapidly towards the sample it will crash[2], since the controller is unable to respond fast enough to stop it. The approach speed finally chosen has to be consistent with the bandwidth of input and output channels, in particular that of the current-to-voltage converter.

An additional problem is that changing the voltage on the outer electrodes of the tip-carrying actuator gives rise to a spurious tunneling current via the parasitic capacitance between the electrodes and the wire that is connected to the tip. If not accounted for, this capacitive coupling is also a factor which limits the tip approach speed. By considering both effects, a maximum tip approach speed can be calculated.

Approach Algorithm

The general approach algorithm used to control the center actuator is described in section 6.4, figure 6.7 as the routine FIND.

When advancing the tip towards the sample the response time of the microscopy system to a ramp input has to be taken into account. This is discussed in section 7.2, where a response time of $t_{\text{ramp}} = 46\,\mu$s was determined. The fact that the threshold has been passed can be detected by the controller only after 5 or 6 sampling intervals. The fact that a change in the control variable is not instantly detectable is a well known problem in control applications [15, 124].

There are at least two methods that can be used to advance the tip and also take the response time into account. The software can wait for the complete response time after each tip advancement or the software can advance the tip by a small amount each cycle. In general, for a digitally controlled system, it is better to take many small steps than a few big steps [57]. The reason for doing this is to avoid excitation of unwanted actuator resonances and prevent strong capacitive coupling. Changing the output voltage at larger time intervals has the same effect as using a zero order hold with lower sampling frequency, which the reconstruction filter is not designed for. This is discussed in section 4.4.

The way in which the tip is moved towards the sample in fine-approach mode, is to first acquire a new datum and then move the tip. Following this, a decision is made whether the tip is close enough or not. Both actions happen within one sampling interval, but the decision is based on a datum that reflects the previous tip position from $n = 5$ or $n = 6$ sampling intervals earlier. Hence the tip must not move too far within each step. A reasonable step size is $\Delta s/(n+2)$, where Δs is the sensing distance of the current-to-voltage converter, which is a fraction of the active region of tip-sample interaction. The active region is mentioned in section 6.5 and is the reach of interaction due to the tunneling effect, which depends on the tip, the sample and the bias voltage. The sensing distance depends on the tunneling effect as well, but also on the dynamic range of the current-to-voltage converter. An estimate of their magnitudes is given below.

[2] A heavy tip-crash can be recognized from the behavior of the controller. After the crash the controller tries to stabilize the tunneling current. Since in most cases the tip would have made contact with the sample, the center actuator would have to be retracted. In observed crashes the tip had to be withdrawn over a longer distance than the depth on impact, which is most likely due to attached sample material bridging the gap between withdrawing tip and sample.

After the control software detects that the tip has reached the active region, it will stop the advance and start closed-loop feedback control. At that point in time the tip is already $n + 1$ steps deep within the sensing distance. With the chosen step size this still leaves a single step as a safety margin before a possible impact. For increased safety a smaller step size can be programmed.

In the discussion so far, the active region is considered to be rigid. This is not the case in the real tip-sample system, but the rigidity assumption leads to a more conservative and therefore safer approach speed estimate. As mentioned in section 2.1.2 there are forces between tip and sample, which are attractive at larger distances and repulsive on closer encounters [44]. These repulsive forces will reduce the likelihood of a tip crash, especially for a soft sample, such as graphite, because of strong elastic deformations [174]. Any adsorbates or oxide layers on the tip apex will cause both attractive and repulsive forces to be noticeable at larger distances [62]. The repulsive force between such a layer and the sample should prevent the tip from reaching the sample. However, even if there is a gentle collision between this layer and the sample the impact might be desirable, since it could remove part of the layer or pick up some sample material as mentioned in section 3.2.

Active Region The active region mentioned above depends on the tunneling effect. It is not easy to give an exact expression for the active region, although qualitative statements are possible and a rough estimate can be made. Starting with equation (2.6) a qualitative argument for the active region is given by

$$s = s_0 \ln(\Lambda R_s), \tag{7.3}$$

where R_s is the resistance of the tunneling junction as in section 5.4.2. The higher the tunneling resistance R_s, the larger the tip-sample distance, which is the case for the lowest possible tunneling current and the highest bias voltage. However, equation (7.3) is not valid for high bias voltages, since then the thin-barrier tunneling becomes Fowler-Nordheim tunneling. Tip-sample distances of 100 nm can be reached in Fowler-Nordheim mode [84]. Fowler-Nordheim tunneling begins at voltages that are approximately equal to the work function of tip and sample [166], i.e. at about 4 V. Before a stable tip-sample distance is reached, the tip might come close enough to the sample for high currents to temporarily flow. It is a matter of knowing whether the underlying sample can stand these high voltages and possibly high tunneling currents as to whether or not the tip should be approached in Fowler-Nordheim mode. In any case the tunneling resistance should be set to a high value when approaching the sample. After tunneling current has been encountered the tunneling resistance can be changed to a more appropriate value.

If a sensitive sample is anticipated, the bias voltage should be high but remain below levels where sample damage might occur. A reasonable choice is 1 V, although for a more sensitive sample a lower value might be advisable. In this region, equation (7.3) will be a good approximation and can be used to estimate the sensing distance. The sensing distance (Δs) depends on the current-to-voltage converter. For the feedback to operate, the tunneling current has to stay between the maximum input current (I_{max}) of the current-to-voltage converter and the noise level (I_{min}). To find the dynamic range of the converter, the noise level was measured and is listed in table 7.2. If a threshold level of

three times the rms noise level is taken for $I_{\min}$ then the dynamic range is

$$\left(\frac{I_{\max}}{I_{\min}}\right) = \left(\frac{5\,\mathrm{V}}{2.0\,\mathrm{mV}}\right) = 68\,\mathrm{dB}.$$

Given the tunneling equation (2.6) one arrives at:

$$\begin{aligned} \Delta s &= s_{\max} - s_{\min} = s_0 \ln\left(\frac{I_{\max}}{I_{\min}}\right) \\ &= 7.8 \times s_0 \end{aligned}$$

The sensing distance depends on the barrier height between tip and sample. The constant s_0 is related to the barrier height and was estimated in section 2.1.1 as $s_0 = 0.05\,\mathrm{nm}$. With this estimate the sensing distance is 0.4 nm. For the small bias voltages that were assumed, the sensing distance does not depend on the bias voltage. However, the position of the tip when tunneling at the setpoint does depend on the bias voltage and should be chosen as large as feasible to give the approach mechanism enough leeway.

With a lower limit of the tunneling current at 20 pA and a bias voltage of 1 V, the resulting tunneling resistance is 50 GΩ.

Approach Speed With a value of $\Delta s = 0.4\,\mathrm{nm}$, a response time of $n = 6$ sampling intervals and a sampling interval of $12.8\,\mu\mathrm{s}$ a fine-approach speed of

$$\frac{0.4\,\mathrm{nm}}{(6+2) \times 12.8\,\mu\mathrm{s}} = 3.9\,{}^{\mu\mathrm{m}}/_{\mathrm{s}}$$

is found for the tip. Hence, totally extending the center actuator will take

$$\frac{1.3\,\mu\mathrm{m}}{3.9\,{}^{\mu\mathrm{m}}/_{\mathrm{s}}} = 330\,\mathrm{ms}.$$

Note that the tip may be retracted at a higher speed. In the current program this is done within 20 ms.

Parasitic Capacitance

A voltage change of $4800\,{}^{\mathrm{V}}/_{\mathrm{s}}$ on the outer electrodes causes 110 pA of spurious tunneling current to flow, from which a parasitic capacitance of 20 fF can be inferred. If a value of 20 pA is acceptable, the approaching speed can be up to $1000\,{}^{\mathrm{V}}/_{\mathrm{s}}$. With a vertical sensitivity of $3.4\,{}^{\mathrm{nm}}/_{\mathrm{V}}$ the center actuator may approach the sample at a speed of

$$3.4\,{}^{\mathrm{nm}}/_{\mathrm{V}} \times 1000\,{}^{\mathrm{V}}/_{\mathrm{s}} = 3.4\,{}^{\mu\mathrm{m}}/_{\mathrm{s}},$$

which is about a factor of 2 slower than the above mentioned speed.

The capacitively coupled current adds a certain offset to the input current. This offset has to be taken into account when determining the tunneling current threshold value. The tip-approach speed should be kept below a certain maximum. The threshold is set according to the maximum expected offset. In the program the actuator is approached at a constant speed. The coupled current offset is constant and the offset can be either subtracted from the threshold

value or from the measured input current. Advantage can be taken of the known polarity of the coupled current and the expected tunneling current.

The way the center actuator is poled, a negative voltage is needed to approach the tip towards the sample. The resulting coupling current is therefore always negative. The threshold value and the expected tunneling current depend on the sign of the bias voltage, which can be of either polarity. It is advantageous to use positive bias voltages.

A positive bias voltage results in positive currents opposite to the capacitively coupled currents. The threshold value can remain where it is or be set closer to zero. If it remains in place then the controller has less time to react to the detected tunneling current, since the tip has to advance closer to the sample to increase the tunneling current by the amount of coupled current. If the threshold is set closer to zero, then the threshold is triggered easily by noise before the approach started. A solution to this problem is to start the approach and then change the threshold.

With a negative bias voltage the coupling currents have the same sign as the expected tunneling currents. The controller cannot distinguish coupling and tunneling current. The threshold has to be set higher, which decreases the dynamic range of the amplifier and diminishes the active region.

The approach can always be done with positive bias voltages. The sign of the bias voltage can be changed after tunneling current has been found.

7.3.3 Total Approach Time

Assuming that the capacitively coupled current is corrected for and using the tip approach speed of $3.9\,\mu m/s$, the time to approach and retract the tip is $(330 + 20)$ ms $= 350$ ms. Together with the time of 125 ms to screw the probe head down a step, the total time for a single approach step is found to be $(350 + 125)$ ms $= 475$ ms. This amounts to a lowering speed of $1.0\,\mu m/475$ ms $= 2.1\,\mu m/s$ per step. Hence, the time to cover the 0.4 mm high ramp is

$$\frac{400\,\mu m}{2.1\,\mu m/s} = 190\,s \approx 3\,min.$$

In practice the approach time is longer, typically 10 min. The calculation has not taken into account the fact that every now and then the threshold is triggered by noise spikes. In such an event the program stops the approach, switches feedback on and waits a certain time. If no tunneling current is detect after this time the approach is resumed.

7.3.4 Approach Speed of Second Method

In this method the probe head walks slowly down the ramp. The maximum distance that the coarse approach has to cover is 0.4 mm, which corresponds to 500×10^3 steps of 0.8 nm. Moving the microscope agitates the tip, so that after each step the tunneling current has to settle down again. For a small step of 6 V this time is about 0.5 ms, so that the total approach time for method two is $500 \times 10^3 \times 0.5$ ms $= 250$ s $= 4$ min. The lowering speed would be $1.6\,\mu ms$, slightly slower than the method discussed above.

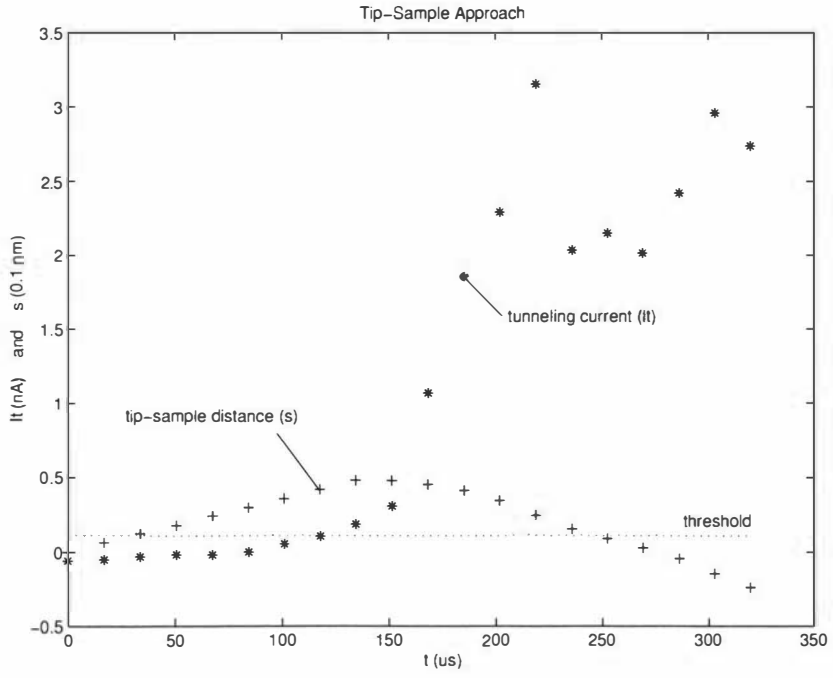


Figure 7.5: **Tunneling Current on Approach:** The observed tunneling current (I_t), the threshold value and the tip-sample distance (s) is shown while the FIND encounters tunneling current. This graph is a detailed version of figure 7.6. The distance s has been calculated from the vertical actuator sensitivity and was set arbitrarily to zero at time zero. The threshold value is 0.11 nA. Note the capacitively coupled current offset at times below about 100 μs. The sampling interval was 16 μs.

7.3.5 Recorded Tip-Sample Approach

A tip-sample approach using the first method was recorded and is shown in figure 7.6. Figure 7.5 shows the details of how the routine FIND switches from tip advance to feedback control. The tunneling current passes the threshold at about 120 μs, the tip advance is stopped one interval later. The controller reacts only slowly to the detected tunneling current. It takes 4 ms for the controller to stabilize the tunneling current. This is shown in figure 7.6. During this time the tunneling current is contained within 5.5 nA. On closer inspection it is found that it takes 0.7 ms to stop the tunneling current from rising further and instead reaching a level 4.5 nA. The controller keeps on retracting the center actuator over the whole period shown. This is due to creep of the actuator, a problem discussed and shown in figure 3.2. Creep lasts about 40 s, much longer than the 16 ms displayed here.

The time delay between detecting tunneling current and feedback being switched on results in a current overshoot. At the same time the tunneling current gets an additional boost due to capacitively coupled current, since the direction of actuator advancement is changing. The reaction of the controller depends on the control parameters, which are set low so that the closed-loop control does not oscillate. It takes a while until the error signal has accumulated enough for the controller to retract the tip. Finally, and this should be the strongest effect, the controller also has to compensate for creep.

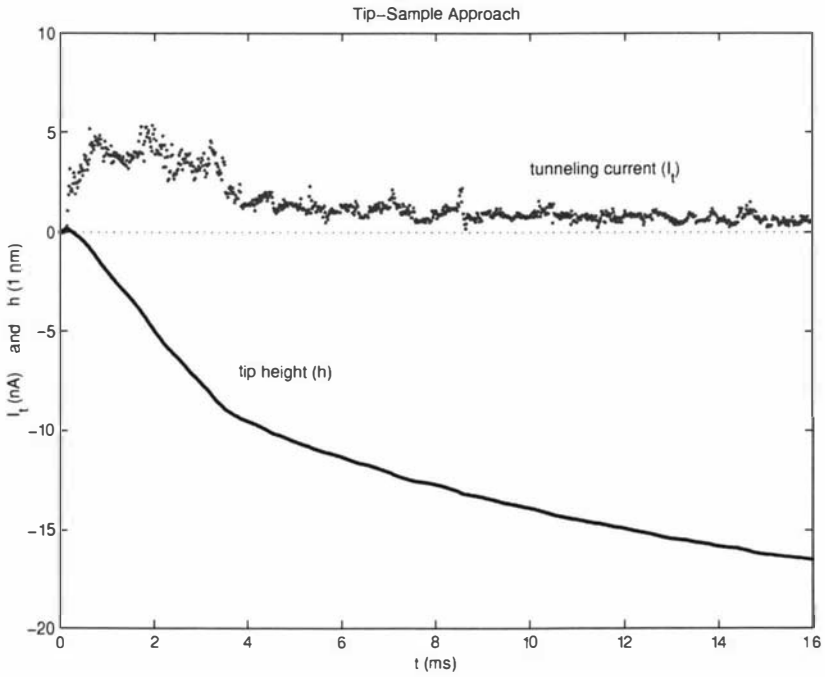


Figure 7.6: **Tunneling Current on Approach (Overview):** The approach shown in figure 7.5 is presented in full. Everything else is equal to figure 7.5. Up to a time of 4 ms the tunneling current is larger than the setpoint. The actuator control voltage retracts the tip strongly during this time. At times greater than 4 ms the retraction is less strong. The actuator is still retracted due to creep as was shown in figure 3.2. The tunneling current stays below 5.5 nA within the first period of 4 ms and settles below 1 nA after that.

All phenomena, capacitive coupling, time delay and creep are known before hand. Therefore the reaction time can be decreased by some preemptive action on the part of the program.

Creep can last for minutes, depending on the actuator expansion. If no action is taken, the operator has to wait until the tip-sample distance is stable. Creep can be reduced by moving the probe head toward the sample to a point at which the center actuator is relaxed. In this case the controller need only wait a short time.

To reduce the reaction time the program could also employ one of the following strategies. For a certain amount of time the actuator can be retracted continuously regardless of the error term. Another strategy is to manipulate the feedback control. The integrator term can be initialized to a certain value other than zero. Alternatively the feedback parameters can be temporarily set to a higher value than usual. In general, the controller can be programmed to behave differently to current under- and overshoot, since only current overshoot threatens the tip with crash.

None of these above mentioned suggestions have been implemented yet.

7.4 Scan Methods

7.4.1 Scan Speed

When a sample is to be imaged, an appropriate scan method, scan area and horizontal resolution has to be chosen.

The choice of scan speed has a strong influence on the image quality. There will be an optimum scan speed for the microscopy system as explained in the following paragraph.

At very small scan speeds the image will contain a certain amount of low frequency noise, coming from thermal drift, low-frequency building vibrations and $1/f$ noise from the electronics and the tunneling process. As the scan speed is increased, the useful information from the sample is shifted to higher frequencies and the image quality improves. If the scan speed is so high that the controller cannot provide the vertical tip movement needed to track the surface contours, the image information will only be available in the tunneling current image and the risk of tip crash is increased. At even higher scan speeds tip crash becomes increasingly likely. The highest scan speed is either given by this safety consideration, by the highest vertical resonance of the probe head or by the highest acquisition speed of the input channel of the electronics. Whichever method is used, the controller should always be adequately tuned so as to have the highest possible bandwidth.

Safe Step Size

On rough terrain the tip is more likely to crash. The scan speed in this system is given by the horizontal step size taken per sampling interval. The safe horizontal step size depends on the setpoint and the slope of the neighboring terrain. From a geometrical argument the horizontal step Δh should be

$$\Delta h < \frac{s}{\tan \beta}, \tag{7.4}$$

where s is the vertical tip-sample distance and β is the slope. An illustration can be seen in figure 7.7.

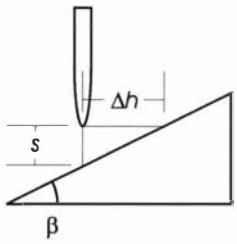


Figure 7.7: **Slope:** The tip is a distance s away from the sample. With a sample slope of β the maximum horizontal distance is Δh.

The further the tip is away from the sample and the shallower the slope, the larger the horizontal step size may be. It has to be borne in mind that the tip-sample distance has an influence on the image quality. Generally speaking, the closer the tip to the sample, the higher the contrast and resolution. The horizontal step size has to be adequate with regards to repetitive features of the surface, usually given by the lattice of the examined crystal. If not enough data points in a scan are used, then spatial aliasing may occur, as pointed out in a somewhat elaborate discussion by Paggel and Förster [148]. Spatial aliasing would produce so-called Moiré patterns.

An upper limit is imposed on the horizontal speed, since the vertical tip motion has to keep up with the horizontal motion. Several scan methods are discussed in section 6.5. With the sampling methods, horizontal and vertical tip motion are coupled, while with the mean value method both motions are independent.

With the mean value method the scan speed must be based on an estimate of the tip-sample distance and the maximum steepness of the terrain. While the tip is moving horizontally, the DSP receives a signal modulated by the surface features. The highest occurring frequency of this signal should stay within the bandwidth of the control loop if the surface is scanned in constant-current mode. In constant-height mode that frequency has to stay within the bandwidth of the input channel of the DSP. To estimate the highest horizontal tip speed in constant-current mode, a surface feature length of not less than 0.1 nm is assumed and the control bandwidth of 1 kHz from section 7.2.3 is taken. This results in a maximum speed of

$$V_{x,\max} = \frac{0.1\,\text{nm}}{1\,\text{ms}} = 0.1\,\mu\text{m/s}.$$

Most sampling methods establish a settled tunneling current at each sample point before moving on. Thus, the actual horizontal speed depends on the roughness of the terrain encountered. If the terrain rises or falls steeply, the speed is slow, whereas in flatter regions of the sample the speed is increased. Additionally, the scan speed depends not only on the terrain, but also on the amount of noise or disturbances in the system. The fastest possible speed will be as fast as the one for the mean value method.

Piner and Reifenberger [154] used another way of deciding whether tunneling current has settled, which was to count how often the measured tunneling current passes the desired current value, i.e. how often it over- and undershoots. This is due to their way of controlling the current, which is basically a type of on/off control. Owing to their on/off control, Piner and Reifenberger could not use the mean value method. Moving the tip horizontally and establishing settled tunneling current had to be done sequentially, which is the reason why they found the incremental method slower than the P/R method. In this project a digital PID controller is used, which is explained in section 6.9.2. With this controller, the tip height can be adjusted at the same time as the tip

is moved horizontally, since both processes happen essentially in parallel. The P/R method will therefore not be much faster than the incremental method, although there is a possibility that it is faster because the settling condition will be easier to fulfill. In any case it will be safer on downhill slopes.

Any time the center actuator is moved horizontally the movement should be as smooth as possible so as to avoid ringing, which is discussed in section 3.3.7. The maximum rate of horizontal acceleration of the tip is related to the lowest lateral resonance frequency of the tip-carrying assembly, which is comparatively low, 2.3 kHz. When using constant horizontal speed, the tip should be accelerated horizontally to the desired scan speed before the scan begins. At the end of the scan the tip should be gently decelerated and then accelerated in the opposite direction instead of moving the tip with a conventional triangular waveform.

7.4.2 Choosing an Acquisition Method

There are several issues that must be faced when assessing the different acquisition methods, these are, tip safety, speed, regularity. The size of the desired scan area and the nature of the sample also play a role. In general, tip safety and high speed are mutually exclusive. If the safety of the tip and sample is not an issue then a fast method can be chosen. This is usually the case when the scan area is small and the terrain is flat, in which case the tip can be scanned in constant-height mode. If the terrain is more or less flat but contains several large scale features the modified constant-height mode can be used, in which the controller is active but has low gain.

The mean value method has the advantage that the horizontal speed is constant, which allows for thermal drift correction. This was the reason why Hoster [102] abandoned a scan method using variable horizontal speed published by previous members of his research group [94, 95].

However, using an appropriate sampling method, the scan speed can be made to adapt to the terrain. Therefore the sampling methods can be safer. Additionally, since image quality and scan speed are coupled, better image quality should be easier to obtain using a sampling method rather than using the mean value method, which relies heavier on the operator to determine a proper scan speed.

Especially at close distances a safer scan method is advisable, since the tip is then more endangered.

Apart from Piner and Reifenberger [154] themselves, the P/R method with adaptable speed has been used by Wong and Welland [194], Robinson *et al.* [163] and Munoz *et al.* [139] and is also favored by Rong *et al.* [164]. Scheuring *et al.* [167] also used an adaptive speed method and state that due to the change in speed, real features of the surface may be distinguished from low frequency noise.

The hopping method might be a good choice for soft samples, since the tip is brought to the sample from above. In this case the risk of ploughing through a sensitive sample is reduced [163]. The hopping method is similar to the pogo method, but should be faster, because the tip remains 'in touch' with the sample. The hopping method has the advantage that it operates at the greatest possible tip-sample distance that still provides tunneling current as explained by equation (7.4). Hopping can therefore use larger horizontal

steps, increasing the scan speed, or it can be used as a safe scan method. Kim *et al.* [114] use the hopping method in scanning near-field optical microscopy. They point out that hopping eliminates horizontal hysteresis of the actuator.

The pogo method is rather slow under most circumstances, since it approaches the tip from a great distance, which has to be done slowly to avoid a tip crash. However, contradictory to expectation the pogo method may be faster than all other methods when a large scan area should be scanned with only a few sample points. This might be the case if only an overview is desired or if the tilt of the sample is to be inspected before a more time consuming high resolution scan is used. The pogo method is by far the safest method, since the tip is only stepped horizontally when it is far away from the sample. The pogo method will be useful when the system is used as a manipulation tool on nanometer scale. The tip can be programmed to approach selected places of the sample, then change the surface locally, retract and advance to the next point. The manipulated points need not lie close together. An image taken with the pogo method is shown in figure 8.6.

Contrast In general, the closer the tip the higher the image contrast. For metallic samples the distance should be comparatively small. Hallmark *et al.* [83] report a tunneling resistance as low as 10 MΩ to image gold in air, while Olesen *et al.* [144] report a resistance of 1 MΩ to image Ni, Pt and Au in ultra-high vacuum. Biological samples on the other hand should be imaged with the tip at a greater distance. Sáenz and García [166] suggest a method where they use their microscope in the near field emission regime with a tunneling resistance of 40 TΩ.

7.5 System Noise

The desired noise performance of the system is discussed in section 5.1.2. The findings are summarized in equation (5.4). An overall signal-to-noise ratio of 40 dB can be expected for a tunneling current of 1 nA. This section discusses the actual measured noise and the resulting signal-to-noise ratio.

7.5.1 Input Signal Noise

The major source of noise on the input signal channel is that contributed by the current-to-voltage converter and everything should be done to minimize this noise. To get the total signal-to-noise ratio of the digitized signal one has to add the contributions of the remaining analog electronics of the input channel as well as the contributions due to the analog-to-digital converter, which are quantization and transition noise. These have been calculated in section 5.4.5.

The overall noise for both the current-to-voltage converter and the ADC module is calculated as

$$V_{\mathrm{n,in}} = \sqrt{V_{\mathrm{n,adc}}^2 + V_{\mathrm{n,ivc}}^2}. \tag{7.5}$$

The noise calculations are graphically summarized in two figures, figure 7.8 and figure 7.9. Figure 7.8 shows the individual noise contributions of the current-to-voltage converter, the ADC module and the total noise for two different input

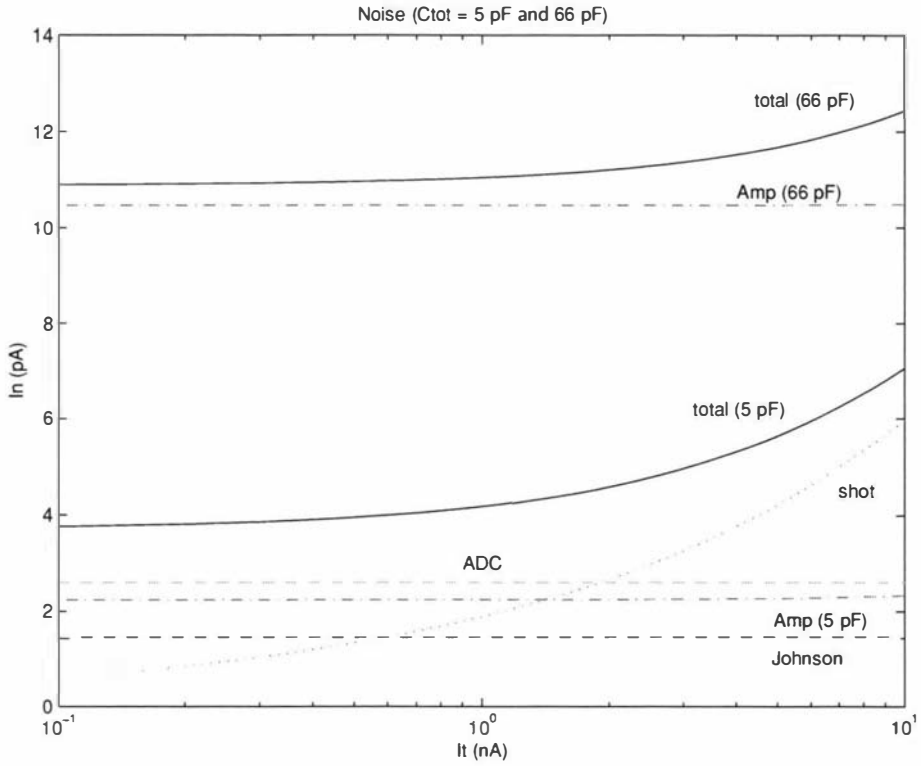


Figure 7.8: **Calculated Noise of Input Channel:** The individual noise contributions of the current-to-voltage converter, the noise of the ADC module and the total noise are shown as a function of tunneling current I_t. The calculations have been done for two different input capacitances; 66 pF and 5 pF. A larger input capacitance greatly increases the amplifier voltage noise of the current-to-voltage converter. For $C_{in} = 66\,\mathrm{pF}$ this noise dominates the noise of the input channel. The Johnson noise of the feedback resistor contributes least to the total noise. With the smaller input capacitance the amplifier voltage noise (Amp(5 pF)) is comparable to the noise of the ADC module. Shot noise only plays a significant role when the input capacitance is small. In this case it dominates at high tunneling currents ($I_t > 2\,\mathrm{nA}$), while the ADC module is the greatest noise source for tunneling currents lower than 2 nA. Values used for the calculations were $V_{bias} = 50\,\mathrm{mV}$, $R_f = 90\,\mathrm{M\Omega}$ and $C_f = 0.25\,\mathrm{pF}$.

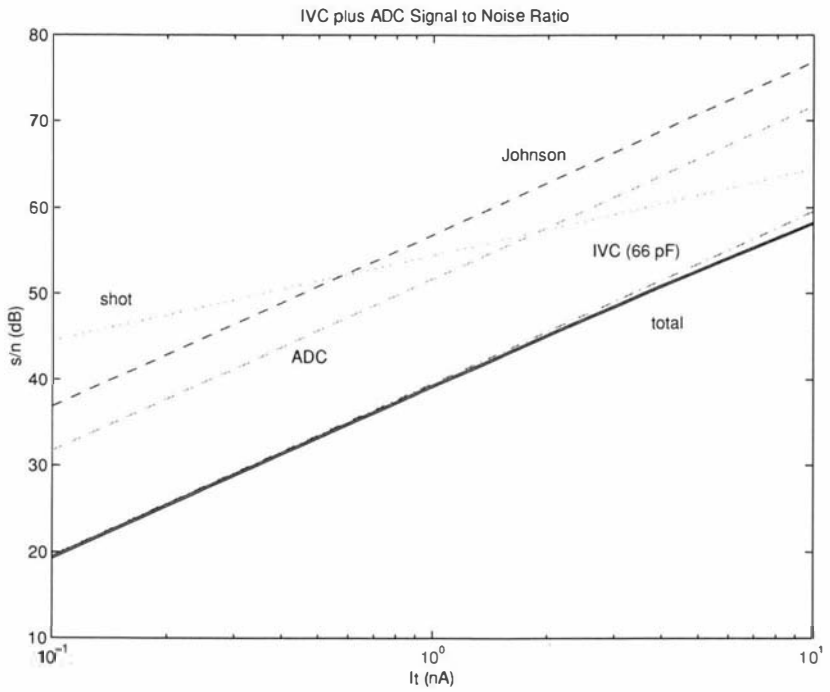


Figure 7.9: **Signal-to-Noise Ratio of Current-To-Voltage Converter and ADC Module:** The figure shows the signal-to-noise ratios (s/n) due to ADC module noise, current-to-voltage converter noise and the total s/n as a function of tunneling current. The top graph is the s/n due to the ADC module, which has hardly any impact on the total s/n, since the noise of the current-to-voltage converter is dominant. The s/n rises from 19 dB at small input currents (0.1 nA) to 58 dB at high input currents (10 nA). Values used in the calculations were $V_{\text{bias}} = 50\,\text{mV}$, $R_{\text{f}} = 90\,\text{M}\Omega$, $C_{\text{f}} = 0.25\,\text{pF}$ and $C_{\text{tot}} = 66\,\text{pF}$.

capacitances $C_{\text{in}} = 66\,\text{pF}$ and 5 pF. The graphs in figure 7.8 demonstrate how strongly the input capacitance of the current-to-voltage converter influences the total noise of the input channel [112]. Figure 7.9 shows the total signal-to-noise ratio split for the ADC module and current-to-voltage converter. As can be seen in the graph, the signal-to-noise ratio of the ADC module is much larger than the signal-to-noise ratio of the currently used current-to-voltage converter, although it is not as large as it could be. Since the feedback resistor has been decreased from 1 GΩ to 90 MΩ for bandwidth reasons, the largest useful signal is also smaller by a factor of 11. The best s/n for the ADC module alone is now only 71 dB instead of the possible 92 dB for an eleven times larger signal. But even though the s/n of the ADC module is smaller than planned, it is still more than adequate in the current situation. This is also true for an input capacitance of 5 pF. Figure 7.9 shows how the amplifier noise is the dominant noise contribution and that the signal-to-noise ratio is better at higher tunneling currents.

The input signal noise has been measured under different conditions, which include having the current-to-voltage converter connected to the probe head, having the amplifier not connected to the probe head and replacing the tunneling gap with a resistor. The rms noise observed under these different conditions is listed in table 7.2. Additionally listed are the input capacitances measured

Condition			C_{in} (pF)	V_n (μV)		I_n (pA)	
ADC	IVC	Remark		meas.	calc.	meas.	calc.
√	×	ADC open circuit	n.a.	210	210	2.3	2.3
√	√	input cable desoldered	9	410	370	4.6	4.1
√	√	with input cable	34	550	650	6.1	7.2
√	√	conn. to probe head	50	740	820	8.2	9.1
√	√	probe head and R_s= 29 MΩ	50	690	970	7.7	11

Table 7.2: **Measured and Calculated Input Channel Noise:** The table lists the measured and calculated output rms noise voltages and the corresponding input noise currents for the current-to-voltage converter (IVC). The noise voltage includes amplifier voltage noise, feedback resistor Johnson noise and noise from the ADC module. The measured values are standard deviations of data collected with the ADC module. Each data set contained 1024 data points, sampled at a sampling interval of 15 μs (12.8 μs for the o.c. measurement). The noise value for the ADC module is calculated in section 5.4.5. The calculated data for the IVC are arrived at by using formula 7.5 under the condition of no input current, i.e. zero shot noise ($V_{sn} = 0$) and infinite input resistance ($R_s \to \infty$). The input capacitance used to calculate the amplifier voltage noise (V_{na}) is listed in the fourth column. The measured noise voltages are reflected to the input of the IVC using $R_f = 90$ MΩ.

with an *LCR*-meter and the calculated noise using these capacitances. For the ADC module the calculated and measured values agree well. Both the measurements and the calculations also show the increase in noise with increasing input capacitance, although the calculation does not agree exactly with the measurement. The disagreement is partly due to the uncertainty in the capacitance measurement. The calculated noise can be larger than the measured noise, since the noise densities for the operational amplifier used for the current-to-voltage converter are the maximum values given by the manufacturer. In the last column, the drop in measured voltage noise is unexpected. However, it might be explained by assuming a certain amount of interference that couples capacitively to the current-to-voltage converter input. This interference would be worse for higher impedances.

7.5.2 HV and Bias Module Noise

The noise of the output channels, the HV modules and the Bias module, were also measured with the ADC module. The measurements are listed in table 7.3, where they are compared to the calculated rms noise values from sections 5.4.6 and 5.4.7. With the knowledge of the ADC module noise the measured noise values were corrected. The noise performance of the Bias module can not be measured adequately with the ADC module. The lower measured noise voltage of 180 μV instead of 210 μV reflects that connecting the Bias module to the ADC module presents a lower source resistance to the input buffer of the ADC module and therefore less noise.

While the calculated noise is approximately halved using the improved version of the HV module, the measured noise of the original HV module is lower than expected.

The measured noise for the HV module can now be related to the desired relative resolution ($\Delta s/s_0$), which is mentioned in section 5.1.4 as equation (5.2)

Module	V_n (μV) meas.	corr.	calc.	Δs_n (pm)	$\frac{\Delta s_n}{s_0}$ (dB)	Δh (pm)
HV	670	640	940	4.4	-27	35
HVn	540	500	470	3.4	-29	27
Bias	180	0	30		-64	

Table 7.3: **Measured Noise of HV and Bias Module:** The table lists the measured and calculated rms noise voltages for the Bias and HV module. The measured values are standard deviations of data collected with the ADC module. Each data set contained 1024 data points, sampled at a sampling interval of 15 μs for the Bias measurement and at 12.8 μs for the HV measurements. The noise value for the HV module is calculated in section 5.4.6 and the one for the Bias module in section 5.4.7. The noise voltages are corrected using 210 μV for the noise of the ADC module. For the HV modules the equivalent actuator movement Δs_n is also given, calculated from the corrected noise voltages, the sensitivity $r_z = 3.4$ nm/V and taking a factor of $\sqrt{4} = 2$ for the four actuator electrodes into account. The second to last column gives the resolution of the signal in dB. The vertical actuator movement Δs_n is compared to $s_0 = 0.1$ nm. In the same column a bias voltage level of $V_{bias} = 50$ mV is compared to the calculated noise in column 4. The last column gives the horizontal actuator movement (Δh) calculated from $r_h = 27$ nm/V and also multiplying by $\sqrt{4} = 2$. The actual Δh will be smaller, since x and y electrodes are orthogonal.

and here repeated as

$$\frac{\Delta s}{s_0} = \frac{1\,\text{pm}}{100\,\text{pm}} = 10 \times 10^{-3} \equiv -40\,\text{dB}\,.$$

The vertical actuator displacement Δs is caused by thermal drift, vibrations and electronic noise of the HV modules. In this section only the electronic noise is discussed. The displacement Δs_n due to the HV module noise ($V_{n,hv}$) was calculated and is listed in table 7.3. If only one HV signal is used to drive the center actuator then the displacement Δs_n would be given by

$$\Delta s_n = r_z V_{n,hv}.$$

A fact that has been overlooked in the design stage is that four signals are driving the center actuator, which increases the tip-sample distance noise by a factor of $\sqrt{4} = 2$. The resulting Δs_n with the currently used HV modules is 4.6 pm instead of 1 pm. Using the improved HV module the value is reduced to 3.7 pm, which is slightly better.

The horizontal uncertainty in actuator movement (Δh) was also calculated and is listed as the last column of table 7.3. With 40 pm, the value for Δh is also larger than the planned horizontal resolution of 10 pm.

The coarser than planned resolution can be improved by reducing the output voltage level from the HV module. Since the noise from the HV module is proportional to the gain of the high voltage amplifiers, the desired resolution can be achieved by using about a quarter of the currently used gain. The available scan area would still be adequate (1.25 μm $\times$ 1.25 μm) and the tip approach only marginally slower. As an extra benefit, a power-supply with lower voltage can be used.

7.5.3 System Signal-to-Noise Ratio

The findings from sections 7.5.1 and 7.5.2 are summarized in a graph in figure 7.10. Measurements from the system while tunneling are indicated in the

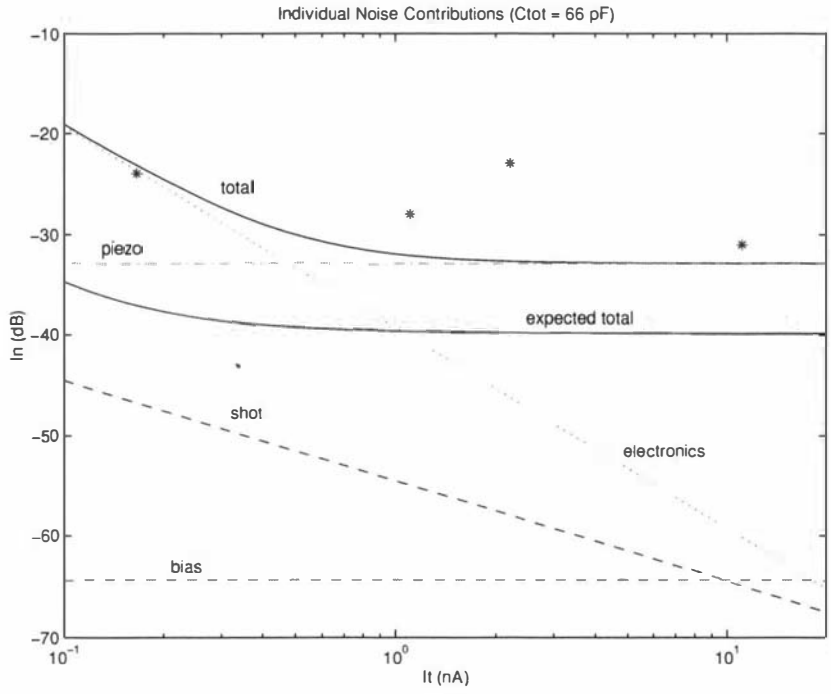


Figure 7.10: **Total Noise of the System (I_n/I_t):** The total noise of the system is shown, as well as all relative individual contributions. The graphs are calculations based on individual measurements of the noise of the modules presented in figure 7.9, table 7.2 and table 7.3. At small tunneling currents the system noise is dominated by the electronics of the input channel, at larger tunneling currents by the actuator control voltages. The intrinsic noise of the tunneling process, the shot noise, is only part of the expected total noise. The system noise does not reach the expected noise, but is only about 7 dB above it. The stars indicate measurement points, which agree reasonable well with the model, where vibrations were not taken into account. The relative noise of the Bias module is adequately small.

graph as stars. Remembering that the real-life measurement must be noisier than the model, since vibrations were not accounted for in the model, the values are in reasonable accordance with the model. The expected total noise from equation (5.4) from section 5.1.6 is also plotted. At a tunneling current of 1 nA, the system noise is about 7 dB above the desired value of 40 dB. At tunneling currents below 0.5 nA the input channel electronics contribute most of the noise, at higher currents the HV modules dominate. This can easily be remedied if the maximum HV output is reduced as suggested in the last section. The next step would be to build a better current-to-voltage converter, starting with a decrease in the input capacitance.

Chapter 8

Experiments

8.1 Characteristic Curves

8.1.1 I-s Curve

The variation of the tunneling current with tip-sample distance was recorded on a graphite sample. The resulting I-s curve is shown in figure 8.1. To find the apparent barrier height (Φ) the tunneling conductance G_t was calculated and the apparent barrier height calculated from

$$\Phi_{ap} = \left(\frac{1}{2} \frac{\hbar}{\sqrt{2m}} \frac{\partial \ln G_t}{\partial s} \right)^2 ,$$

which is equation (2.9) from section 2.1.1. The slope of the graph in figure 8.2 is constant at higher s values. From the slope of the graph an apparent barrier height of about $\Phi = 0.8$ eV was found. A low apparent barrier height is expected in air. The result can be explained by a contamination of the tip or sample. Coombs and Pethica [48] propose that the reason for an anomalously low barrier height is actual tip sample contact via an oxide or dust particle. In section 3.2 the oxide layer on the tungsten tip due to the etching process is mentioned. Additionally, Tiedje *et al.* [183] found material from the graphite sample on their tips. In conclusion, the curve in figure 8.1 does not describe the tungsten-graphite system alone. However the curves do demonstrate the ease with which I-s curves can be recorded with this instrument.

8.1.2 I-V Curve

Several local I-V curves have been recorded on graphite as shown in figure 8.3. The curves show the proportionality between bias voltage and tunneling current as expected for small bias voltages. This is discussed in section 2.1.1 and expressed in equation (2.5). With closer proximity to the sample as indicated by the higher current setpoints, the slopes become steeper, which is also predicted by equation (2.5). Since tunneling happened in air, adsorbates on tip and sample and an oxide layer on the tip will influence the measured curve. As with the I-s curve, figure 8.3 is shown to demonstrate the flexibility of this microscope.

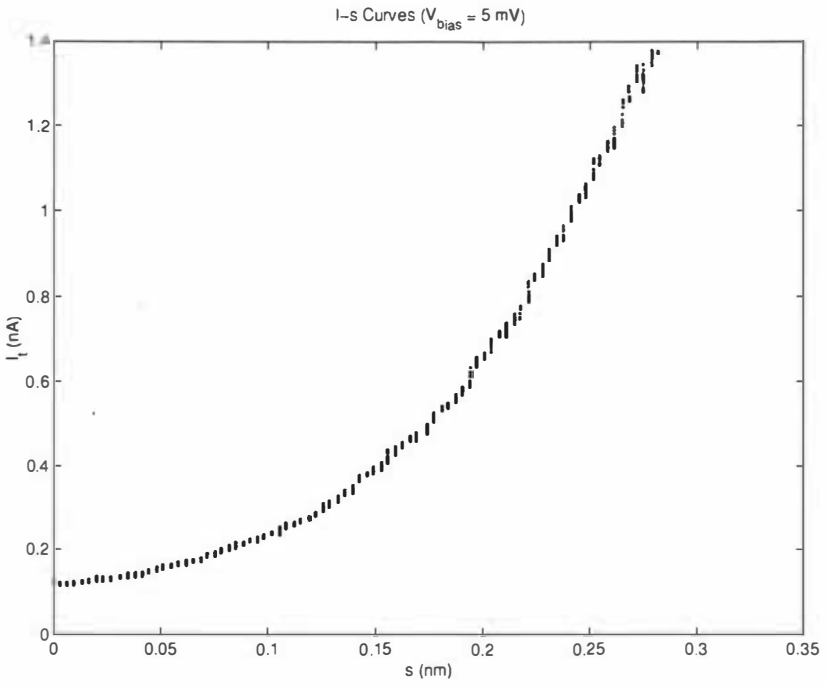


Figure 8.1: *I-s* **Curve:** The tunneling current rises exponentially with increasing proximity to the sample. The distance moved towards the sample (s) has been calculated from the vertical sensitivity of the actuator. The zero point on the s-axis is arbitrary. At each sample point 12 data values were recorded, all of which are displayed. The bias voltage was set to $5\,\mathrm{mV}$.

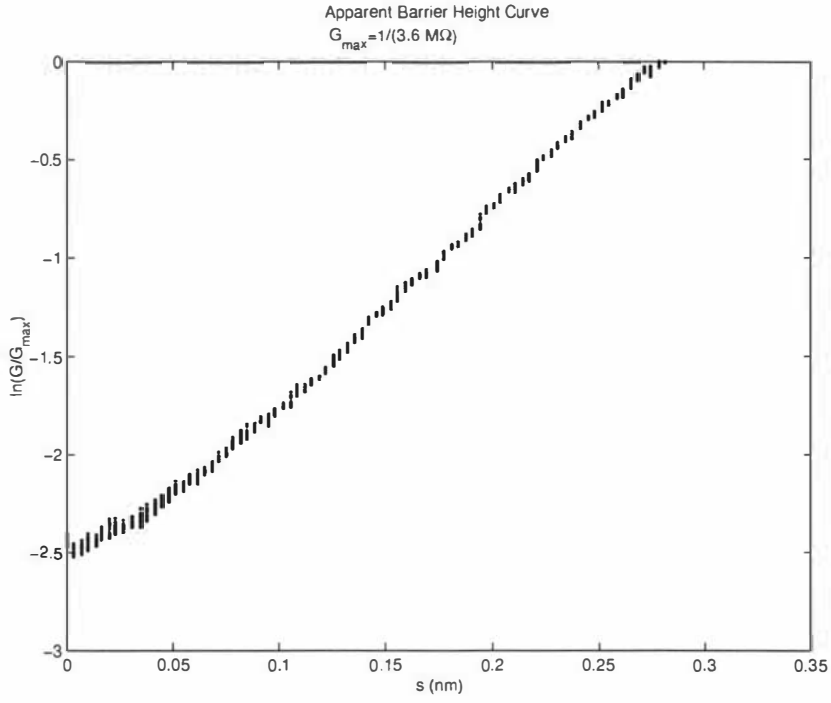


Figure 8.2: **Apparent Barrier Height Curve:** The tunneling current data from figure 8.1 have been displayed as the logarithm of the tunneling conductance G_t. The conductance has been normalized with the maximum conductance of $G_{t,max} = 1/(3.6\,\mathrm{M\Omega})$. The minimum conductance was $G_{t,max} = 1/(45\,\mathrm{M\Omega})$. From the slope the apparent barrier height Φ can be calculated and $\Phi = 0.8\,\mathrm{eV}$ has been found.

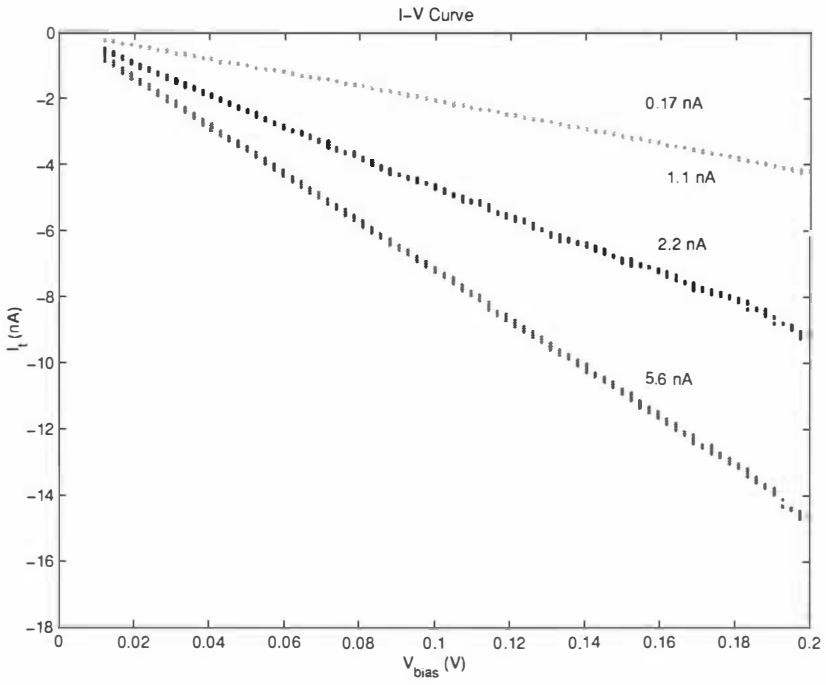


Figure 8.3: I-V Curves: At different tip-sample distances, the bias voltage was varied between 0 and 200 mV. The distance to the sample can not be given absolutely, but is indicated by the current setpoint. Each sample point contains 12 data values.

8.2 Raster Scan Images

Experiments were performed to verify the operation of the microscope in imaging mode. Many of these were done during the development of the instrument and do not represent the image quality currently available. Note that atomic resolution is very difficult to obtain and has thus far only been produced on a handful of occasions.

All images shown here were displayed using the SPIP program from Jørgensen [108]. The images are displayed as one might intuitively guess. If the data file is thought of as a $n \times n$ matrix, then the top left corner of the image corresponds to element $(1, 1)$ of the matrix. A linescan is displayed as lines from left to right. The top of the images corresponds to the direction of the $+y$ electrode of the center actuator, while the right side of the images corresponds to the $+x$ electrode of the center actuator.

8.3 TEM Grid

8.3.1 Incremental Method

Figures 8.4 and 8.5 were recorded using the incremental scan method described in section 6.5. They show a comparison between this microscope and a scanning electron microscope[1]. The image in figure 8.4 was generated with this

[1]Microscope photos courtesy of Doug Hopcroft from HortResearch.

microscope, while the image in figure 8.5 was made with the scanning electron microscope (SEM). The object is a grating replica, normally used to calibrate transmission electron microscopes. The grating was sputtered with a thin layer of gold to enhance the contrast for the SEM. This also made it a more interesting sample to look at, because of the nanometer sized gold blobs that formed on the surface. The grating replica has a spacing of 460 nm, which is confirmed by the SEM image in figure 8.5. The images show that this microscope has higher resolution and better contrast than the SEM. However, the image in figure 8.4 also shows distortions and needed 40 minutes recording time, while the SEM image was taken in real time. The SEM is at its limit of resolution with the image of figure 8.4.

8.3.2 Pogo Method

The pogo method is explained in section 6.5. Using the pogo method the image shown in figure 8.6 was recorded. The bottom right and top left corner of the image in figure 8.6 show the limits of the actuator range. These are the regions where the actuator cannot extend any further downwards (bottom right corner) or upwards (top left corner). Although the total vertical extension of the center actuator is $1.2\,\mu$m, this is only true along the actuator axis and not if the actuator is also horizontally extended. In fact, when the actuator has moved to its maximum horizontal position then no vertical movement is possible. The limits of the actuator movement are in the shape of a diamond, as shown in figure 8.7. Only if all actuator electrodes move the actuator with equal strength, are the resulting tip movements orthogonal to each other and normal or perpendicular to the sample. This is not the case for the actuator of the current set-up. As discussed in section 3.1.2, the $-x$ and $+y$ electrodes are likely to be stronger, since they show a larger double piezoelectric effect. The effect was about a factor of 2 stronger for both these electrodes relative to the remaining two electrodes. The consequent warping of the actuator reach is shown in figure 8.7. Although the program wants to move the tip upwards, the resulting tip movement is actually an oblique upwards and sidewards movement. The points in figure 8.7 that are indicated as 'left' and 'right' can be the utmost excursion points for either of the two electrode pairs and especially for the combination of the two, forming the diagonal from the top left corner to the bottom right corner. Points A and B in figure 8.7 show where the actuator loses touch with the sample, seen in figure 8.6. The difference in electrode strength can be taken into account by the software and a calibration has been undertaken.

8.4 Graphite

While the image in figure 8.6 shows the maximum scan area, an example of atomic resolution of highly oriented pyrolytic graphite is shown in 8.8. Highly oriented pyrolytic graphite is one of the easiest samples to look at with a scanning probe microscope, because a clean surface with large atomically flat areas is easy to prepare [58]. Graphite is layered, consisting of many planes, where the atoms within one plane are ordered like a honeycomb [13]. Several layers of graphite can be lifted off a bulk sample with adhesive tape, since the inter-plane bonds of graphite are weak compared to the bonds within a plane. The planes

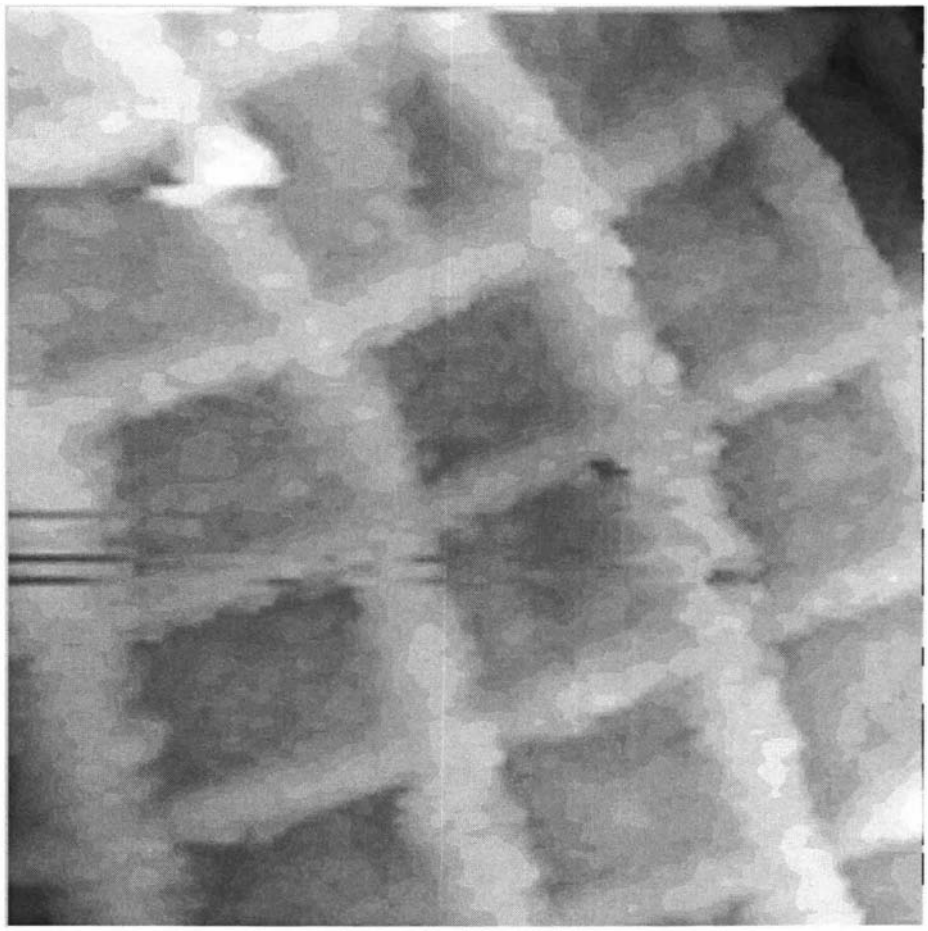


Figure 8.4: **TEM Grating:** This gold sputtered TEM grating was recorded in constant current mode with the incremental method explained in section 6.5. The maximum applied voltages for the horizontal deflection were equal in both directions, 46 V. The grid spacing is known to be 460 nm. Horizontal sensitivities of roughly $r_x = 28\,\mathrm{nm/V}$ and $r_y = 40\,\mathrm{nm/V}$ can be inferred. The slope of the sample has been neglected, but the error will be comparatively small. The height of the lines is 10 nm using the calculated vertical sensitivity of $r_z = 3.4\,\mathrm{nm/V}$. The total displayed z-range is 60 nm.

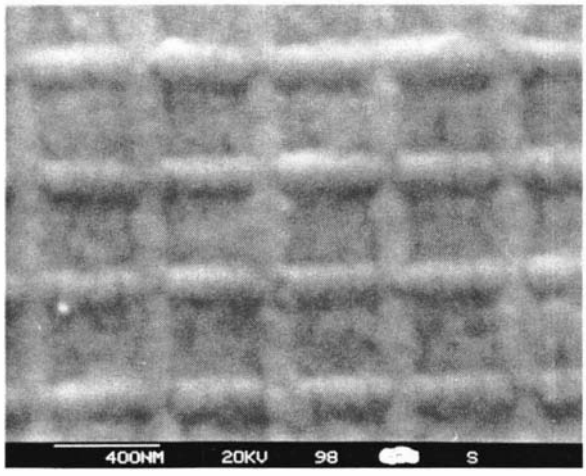


Figure 8.5: **TEM Grating, SEM image:** The gold sputtered TEM grating of figure 8.4 is shown as observed by an SEM. The indicated 400 nm scale confirms the grating spacing of 460 nm (Image from Doug Hopcroft, HortResearch).

of graphite are stacked in ABAB order. Within one layer of graphite there are two different sites for the carbon atoms, α and β, shown in figure 8.9. The sites differ in that one atom (α) has a neighbor in the layer below, while the other one (β) does not. The distance between two equal sites is 0.246 nm.

8.4.1 Interpretation

The interpretation of the graphite image is not straightforward, since the image shows threefold symmetry instead of the sixfold symmetry of the honeycomb structure of the carbon atoms. The reason for this is that in the case of graphite the microscope is not so much imaging the topography of the sample, but rather the local density of states at the Fermi level. The sites α and β differ with respect to the density of states. The carbon atoms at the β site, without neighbors in the lower layer, have a higher local density of states at the Fermi level and usually dominate the measured image [58], which is why only a threefold symmetry is seen. Leggett [123] cites a calculation from Tsukada *et al.* who confirm this result. They additionally find that the tunneling current can also be greater on the other sites, depending on the tip condition. In all cases the center of the carbon honeycomb is the site where the lowest current will be measured.

To achieve the greatest contrast when imaging graphite, the bias voltage should be close to zero, since the electronic corrugation is greatest at the Fermi level of graphite, while it diminishes with increasing voltages [27]. In principle, the observed corrugation decreases exponentially with the distance to the sample [44], which means that the contrast is also enhanced when the tip is close to the surface, i.e. at high tunneling currents [58]. The tip cannot be brought too close to the sample however, since at very close distances the effective barrier height Φ becomes small and with it the contrast.

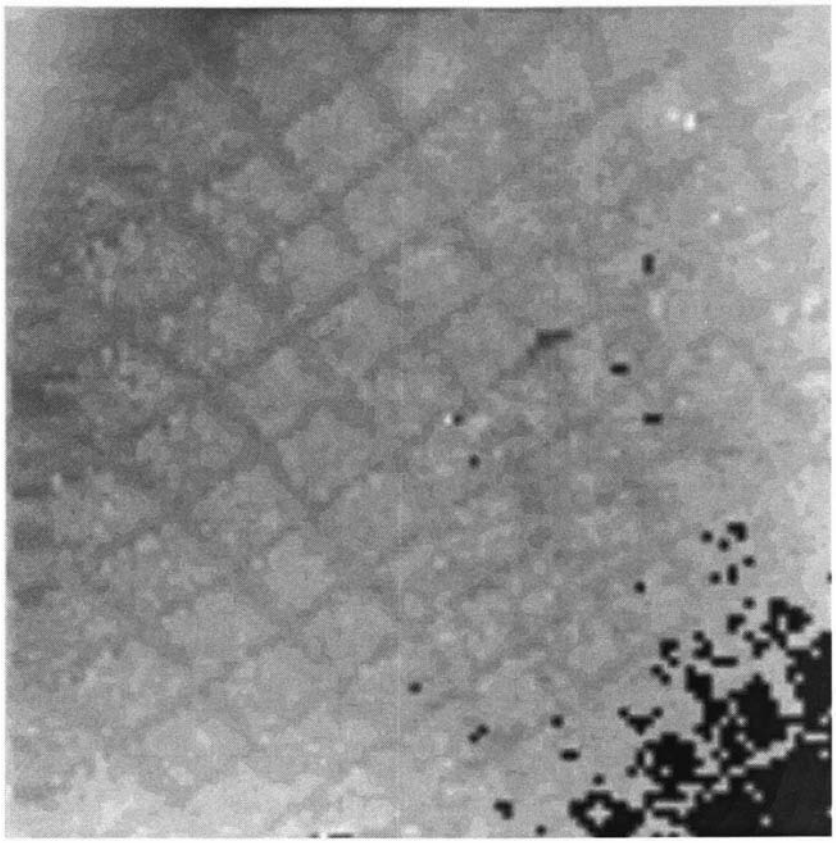


Figure 8.6: **Grid Imaged with Pogo Method:** This image shows the gold sputtered TEM grid of figures 8.4 and 8.5. The resolution is comparatively low, because the pogo method, explained in section 6.5, was used. The range is the maximum possible for the microscope. From the known grid spacing of 460 nm a side length of roughly 4.3 μm can be calculated. There is however a large slope in the original data, especially along the diagonal from top left to bottom right. The slope distorts in the upper left part of the image. In the corners of the diagonal the center actuator loses touch with the sample as explained in figure 8.7.

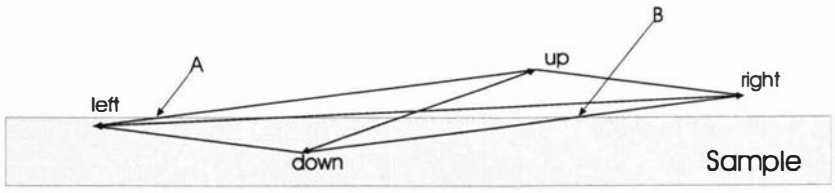


Figure 8.7: **Limits of Actuator Range:** The limits of the possible extension of the center actuator, horizontally and vertically, forms a diamond shape. The vertical directions (up/down) and one pair of horizontal directions (left/right) are indicated in the diagram. The center of the diamond is the relaxed tip position over the sample. To draw the diamond, calculated sensitivities of the actuator were used. The diamond is distorted, because the right electrode was taken to have double the piezoelectric sensitivity than the left one. As a result the tip cannot follow the sample beyond the points A and B indicated in the diagram. This explains the image in figure 8.6.

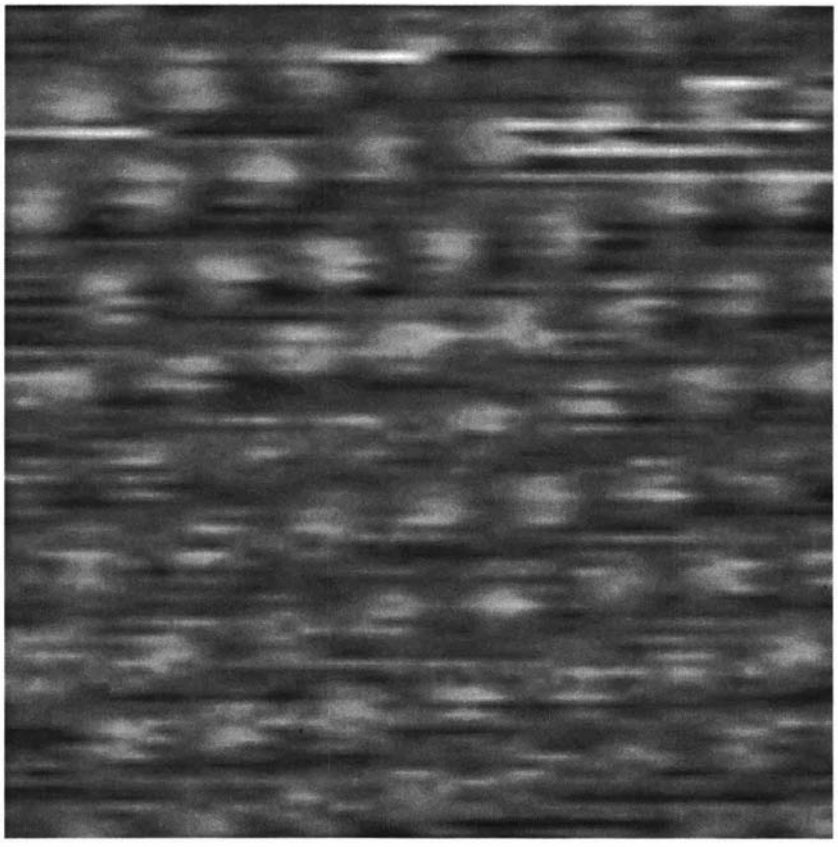


Figure 8.8: **Graphite:** The figure shows an image taken of highly orientated pyrolytic graphite in air. The image shows raw data. The side length calculated from the horizontal sensitivity of the actuator is 2.4 nm, while the lengths calculated from the graphite lattice distance is 1.80 nm in x direction 1.95 nm in y direction. Using the vertical sensitivity of the actuator, the measured corrugation is about 0.2 nm. The current set point was $I_t = 2.2$ nA at a bias voltage of $V_{bias} = +0.3$ V. The number of data points were 100×100 with $N = 15$ samples taken at each point.

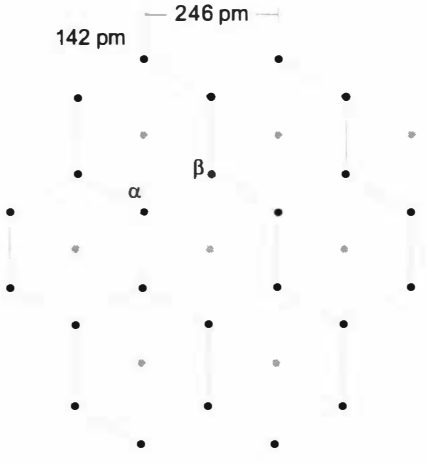


Figure 8.9: **Graphite Structure:** Two planes of graphite are shown, the surface plane in black and the first underlying plane in gray. Within the honeycomb arrangement there are two different sites. The carbon atoms at site α have a neighbor in the underlying plane, while the ones at β have not.

8.4.2 Image Size and Corrugation

The size of the image in figure 8.8 is 2.4 nm × 2.4 nm, if the horizontal sensitivity of $r_h = 27\,^{\mathrm{nm}}/_{\mathrm{V}}$ is used. The known lattice constant of graphite can be used to calculate the horizontal sensitivities in the x and y directions from the image shown in figure 8.8. The image has the dimensions 1.80 nm × 1.95 nm and the sensitivities are

$$r_x = 20.2\,^{\mathrm{nm}}/_{\mathrm{V}}$$

for the sensitivity in x direction and

$$r_y = 21.9\,^{\mathrm{nm}}/_{\mathrm{V}}$$

for the sensitivity in y direction. This is a surprising result, since different horizontal sensitivities were expected from the measurements of the double-piezoelectric effect in section 3.1.2. These were confirmed with large scale measurements and can also be seen in figure 8.6. However, with less than 90 mV, the scan voltages used in figure 8.8 are noticeable smaller than in the above mentioned experiments. This suggests that the observed effect of unequal horizontal sensitivity is related to the electric field strength. As long as the field strength remains small the effect is not observed. Other non-linearities of piezoelectric material, such as hysteresis and creep, are also known to be stronger with a larger electric field.

The measured corrugation of the image in figure 8.8 was roughly $c_{\mathrm{m}} = 0.2$ nm. The s-wave tip model of Tersoff and Hamann as described in section 2.1.2 does not explain this large corrugation. Tersoff and Lang [182] calculate the expected measured corrugation of graphite using the s-wave model and find a very small corrugation. They attribute the failure of the model to the contributions of the orbitals with higher azimuthal quantum number m to the tunneling current. They also point out that graphite is a special case and that the s-wave model is still generally quite accurate. The large corrugation is probably due to some kind of mechanical interaction between tip and sample. The corrugation enhancement due to elastic deformation of the graphite sample as explained by Soler *et al.* [174] is mentioned in section 2.1.2. In another study of scanning a graphite sample in air, Tiedje *et al.* [183] find soft carbonaceous material on the surface of the tip and explain the giant corrugations by the elastic deformations of this soft material rather than the sample itself. In the same article, Tiedje *et al.* report a sudden change of image quality during a scan. Such an abrupt change was also observed with this microscope and is shown in figure 8.10.

8.4.3 Noise Filtering

Acquired images that are stored digitally can be filtered, if knowledge about the characteristics of signal and noise are available. In most cases it is safe to assume $1/f$ noise in the fast scan direction (x) and white noise in the slow scan direction (y).

There is a certain amount of low-frequency noise in the image shown in figure 8.8. A certain contribution to this noise is $1/f$ noise of mechanical or electric origin as well as noise from external vibrations. However, a great part of the $1/f$ noise seen when tunneling is most likely due to a physical process

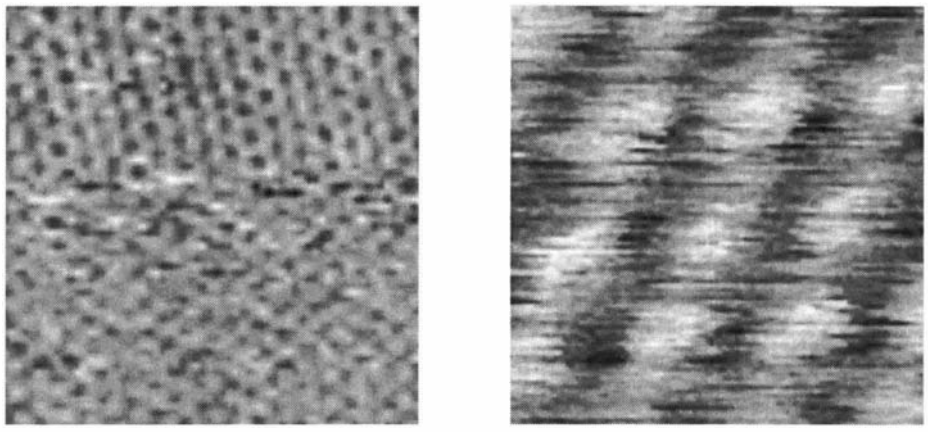


Figure 8.10: **Tip Change and Close-up:** A sudden tip change during a scan is shown in the left image. The current was set to 1.1 nA with a bias voltage of +0.3 V. The image size is about 5×5 nm. The top half of the image shows the graphite pattern, while in the center the image no clear pattern can be seen. This was also observed by Tiedje *et al.* [183] and is explained as a reconfiguration of tip contaminants. The image to the right shows a graphite image at high resolution, the image size is approximately 1.3×1.3 nm. The data are not filtered. The bias voltage was the same as in the left image, the current set point was 5.6 nA. The image corrugation was measured as 0.12 nm.

accompanying the tunneling process, which is also reported by other researchers when tunneling in air under ambient pressure [1, 129, 148, 150, 177, 183]. The study of $1/f$ noise is an intriguing subject and by no means fully understood. Park and Quate [150] already suggest the surface diffusion of adsorbed molecules as a source of this noise. In a more recent study Maeda *et al.* [129] present detailed experimental work on $1/f$ noise in tunneling microscopes and discuss the origin of this noise. They find that the $1/f$ noise is due to fluctuations in tunneling resistance (R_t), which in turn is caused by a fluctuation of both the sample work function Φ_s and the tip work function Φ_t. They suggest desorption and adsorption of gas molecules on the surfaces of tip and sample to be the cause of these fluctuations.

Park and Quate [106] do assume white noise in y direction, while Stoll and Marti [177] point out that instead of uncorrelated white noise there should be a large contribution of $1/f$ noise in y direction since the time between adjacent linescans is larger than between adjacent data point within a linescan. However, the effect of a very slowly decreasing $1/f$ noise in the y direction is essentially the same as assuming white noise.

Jørgensen included a way of removing $1/f$ from images into his display program, SPIP, which is similar to the procedure described by Park and Quate [106] and also Stoll and Marti [177]. All three groups use two-dimensional (2D) Fourier techniques to adjust a Wiener filter, although cruder techniques are also effective as demonstrated in figure 8.11. Noise filtering is done in the spatial frequency domain, where signal and noise can to a certain extent be distinguished. An example of a 2D Fourier transformation is shown in figure 8.11. Each axis of the 2D Fourier transform in figure 8.11 corresponds to a spatial frequency and can be interpreted as a wave vector. For example, the x axis in real space is transformed to the k_x axis in spatial frequency space, where $k_x = (2\pi)/x$. This is similar to the 3D reciprocal space in solid state physics [13]. The signal

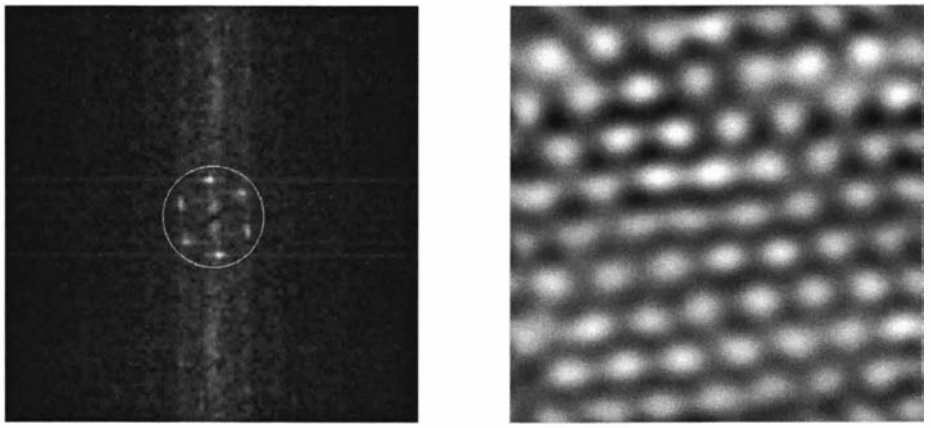


Figure 8.11: **2D Fourier Transformation of Graphite Image:** The left image shows the 2D Fourier transform of the image in figure 8.8. The axes can be interpreted as the direction of the spatial wave vectors. The six peaks represent the six direction of the hexagonal graphite structure. Noise with $1/f$ characteristic can be seen all along the y axis decreasing in both x directions. The right image is the inverse Fourier transformation of the left image, after the high spatial frequencies have been removed by including only spatial frequencies lying within a concentric circle that includes all six peaks in the Fourier image.

content of the image can be seen as the six peaks signifying a repetitive structure in each of three directions. Fast changes in the real-space image are seen at high spatial frequencies in the wave vector image. For instance the irregularities along the real-space y axis are displayed as white noise along the k_y axis in wave vector space in the image in figure 8.11. In the k_x direction this white noise ebbs off towards higher spatial frequencies. This is because the fairly regular horizontal tip movement in the x direction translates the time-related $1/f$ noise into space related noise. Incidentally, if a variable speed mode is used for the horizontal scan, then the horizontal movement would have irregular speed and the $1/f$ noise would be spread out over the k_x direction. Scheuring *et al.* [167] report an improvement in signal-to-noise ratio when using variable horizontal speed. Although the noise is now an overall background noise and cannot be distinguished any longer in the spatial frequency domain, the technique has an advantage if the overall scan speed is so low that it is within the $1/f$ noise region. Since the higher frequencies are filtered out by the system, spreading the noise to higher frequencies increases the overall signal-to-noise ratio. In a system where high scan speeds can be achieved, the $1/f$ noise can be distinguished in the spatial frequency domain and it is probably better to not spread the $1/f$ noise, but rather remove it by filtering.

8.4.4 Terraces on Graphite

A graphite sample was scanned at different length scales and terraces on graphite were found as seen in figure 8.12. On a very small scale only a step is seen, for example in figure 8.13. Figure 8.12 shows how the microscope can operate over different length scales, while zooming in on a particular feature. In figure 8.13 both current and height image are shown. Near the edge the current image shows sudden changes. Such a rapid change in sample topography is a challenge for the controller.

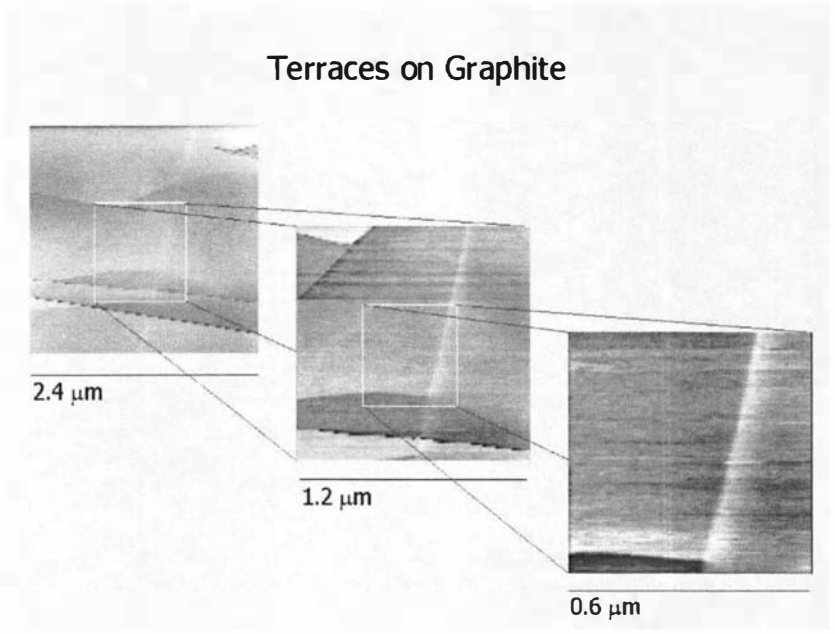


Figure 8.12: **Terraces on Graphite:** The same feature of a graphite sample was scanned at three different length scales. The current was set to 0.3 nA at a bias voltage of +0.3 V.

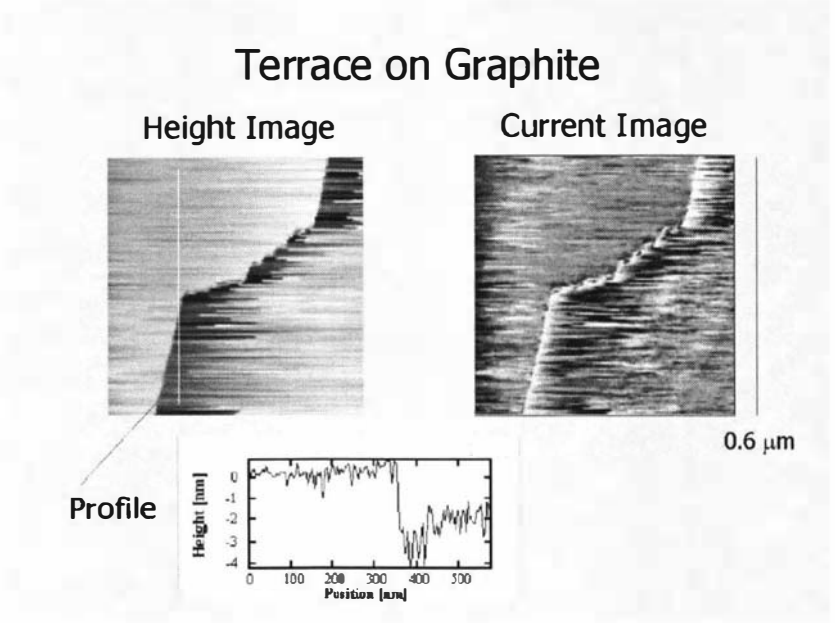


Figure 8.13: **Single Terrace on Graphite:** A single step on graphite was encountered. The scan from left to right is shown, as a height and current image. The bias voltage was +0.3 V with a current setpoint of 0.3 nA.

Chapter 9

Summary and Conclusion

9.1 Summary

9.1.1 Mechanics

The physical dimensions and most of the mechanical, electrical and thermal properties of the microscope and all its essential parts were discussed in chapter 3. Knowing the vertical dimensions of the probe head and sample holder makes it possible to place the tip at the right distance from the sample. The piezoelectric actuators, their operating principle and consequences thereof were discussed in detail. It was found that with 27 nm/V the center actuator has a larger horizontal sensitivity than expected. Later in the experimental section this calculated value was shown to be only a small overestimate, at least when small voltages are applied to the center actuator. The horizontal sensitivity, measured using the graphite lattice, is about 21 nm/V. The vertical sensitivity of the center actuator was calculated to be 3.4 nm/V. The movements of the probe head up and down the sample holder ramp were measured and calibrated, the behavior due to the stick-slip movement had a threshold of about 3 V before considerable motion occurred. If V_s is the step voltage applied to the outer actuators of the probe head, the probe head movement can be described by

$$0.50\,\text{nm/V} \times (V_s - 3.2\,\text{V})$$

going down the ramp and

$$0.27\,\text{nm/V} \times (V_s - 3.3\,\text{V})$$

going up the ramp.

Sharp tips for the microscope could easily be made from tungsten using a loop etching technique. This was the basis for further development of tip making in our research group, where a MSc student built an improved system allowing rapid current shut-off. It has to be admitted that the tungsten tips have an oxide layer and have to be either gently collided with the surface before scanning can commence or be treated with high electric fields. Since the prototype phase for this microscope is considered to be over, tip crashes will not be common any longer and the use of tips made out of Pt/Ir should be contemplated in the future.

The principle of vibration isolation was explained in detail. The vibrational eigen-modes of the probe head and center actuator have been calculated. The lowest horizontal frequency was measured at 2.3 kHz. This was assigned to the lowest eigen-mode of the probe head, which was expected to fall between 1 and 4 kHz. The overall vibration transmissibility of the microscope was calculated and plotted and it was found that the resonance of the tip-sample system at 12 kHz is adequately damped. As a suggestion for future improvement the damping stack can be re-designed to reduce its eigen-frequencies and spread the resonance frequencies of the individual plates further away from each other. To that effect the bottom plates could be made heavier and elastic elements with softer spring constants could be used, e.g. shorter, hollow pieces of viton tubes.

The thermal design of the mechanics was examined and an overall thermal drift of $(+13 \pm 30)\,^{\mathrm{nm}}/_{\mathrm{K}}$ was calculated. The actual drift was measured and the experimental value of $-7\,^{\mathrm{nm}}/_{\mathrm{K}}$ lies within the calculated uncertainty. To minimize the uncertainty of thermal drift, a more compact sample holder design was suggested. The measured drift is comparable to other general-purpose microscopes. That this could be achieved is partly due to the chosen probe head design developed by Besocke [25].

9.1.2 Electronics

The electronics discussion started by considering the microscope as a control system in chapter 4 and continued in chapter 5 with the presentation of the design goals for the electronics and then the actual design. Several reasons were given for making the sampling rate as high as feasible in a computer-controlled system. The shortest possible sampling time in this system is 10 μs, equal to the conversion time of the analog-to-digital converter. The possible closed-loop bandwidth of this system was found to be around 15 kHz. Apart from the sampling time, which determines the bandwidth, the delay time of the sampled-data system is important too, since it determines the phase. The shortest possible delay time in this system is

$$t_{\mathrm{adc}} + t_{\mathrm{dac}} = (10 + 4)\,\mu\mathrm{s} = 14\,\mu\mathrm{s},$$

which leaves a phase margin of 104° at 15 kHz. Related to the delay time is the reaction time of the system, which is the time between logging an event, computing a reaction to it and registering the reaction. The fastest reaction time was calculated to be $2 \times t_{\mathrm{adc}} = 20\,\mu$s.

The actual sampling interval is 12.8 μs, due to the architecture of the counting circuit generating the conversion command. This circuit is currently working asynchronously from the DSP. As a consequence, both the sampling interval and the delay time of the system are not as short as could be possible. However, the performance is still adequate. The closed-loop bandwidth could be as high as 11 kHz. The delay time is between $t_{\mathrm{adc}} + t_{\mathrm{dac}} = 16.8\,\mu$s and $2 \times t_{\mathrm{adc}} + t_{\mathrm{dac}} = 29.6\,\mu$s. These times give phase margins of 110° and 63° at 11 kHz. The reaction time is between 2 and $3 \times t_{\mathrm{adc}}$. In the future the counting circuit should be changed in two ways: the conversion command interval should be reduced to 10.4 μs and the DSP should trigger the circuit. The first measure will increase the bandwidth, while the second will assure the shortest possible phase shift.

The closed-loop bandwidth of the feedback loop is currently limited by the current-to-voltage converter to about 1 kHz. This is mainly due to the fact that the current-to-voltage converter introduces a phase shift of 90° at 5 kHz. With a faster current-to-voltage converter the bandwidth of the closed-loop system would be limited by the lowest vertical probe head resonance at 12 kHz. A bandwidth of 8 kHz should be possible with a faster current-to-voltage converter and a good control algorithm.

The system noise was calculated and measured. The measured noise is in accordance with the calculated noise. However, the noise is larger than was set as a goal. The measured noise lies between -20 and -30 dB below the tunneling current, which is about 7 dB higher than planned. At tunneling currents lower than 0.5 nA the extra noise is due to the noise of the input channel electronics, while at higher tunneling currents it is due to the noise from the high-voltage amplifiers generating extra tip-sample movement via the piezoelectric actuators. The input channel noise is caused mainly by the current-to-voltage converter. Reducing the input capacitance of the current-to-voltage converter would be an important improvement, since the input channel noise would be greatly decreased, while at the same time the bandwidth of the current-to-voltage converter would be increased. Due to the HV module noise, the vertical resolution of the center actuator is about 5 pm instead of 1 pm, while the horizontal resolution is about 40 pm instead of 10 pm. In the future the high voltage power supply should be reduced to a quarter of the present value, which would give the desired resolution and would still allow a scan area of $1.25\,\mu\text{m} \times 1.25\,\mu\text{m}$.

Overall, the fact that the control system is digital has not been the limiting factor for either the speed or noise of the closed-loop control, nor for image data storage and transfer.

9.1.3 Experimental Results

The tip-sample approach was discussed in detail and a maximum approach speed of $2.1\,\mu\text{m/s}$ was calculated. The tip-sample encounter was recorded and faster ways of establishing a stable tip-sample distance after the tip-sample approach were suggested.

Using two different samples the proper operation of the microscope was demonstrated. Atomic resolution was not easily achieved, but graphite could be imaged. A certain amount of $1/f$ noise was encountered and several noise filtering techniques were discussed. With a gold sputtered TEM grating, having somewhat larger features, the total scan range of the microscope could be explored. Different scan methods have been suggested and a few, the 'incremental' and the 'pogo' method, have been used. The currently used PID controller was explained and an improved version has been suggested, which should be investigated in the future. The flexibility of the microscope was demonstrated by recording topographical images as well as several spectroscopic I-V and I-s curves. The 'pogo' method can be used in the future to place the tip quickly in specific positions on the sample, for example to manipulate the sample on a nanometer scale by means of a high voltage or by gentle collision.

9.2 Conclusion

In the introduction, chapter 1, the history, advantages and disadvantages of computer controlled scanning probe microscopes were discussed. This thesis presents the mechanical, electronic and software aspects of such a microscope. In the experimental section its orderly operation and its possible use are demonstrated. Within New Zealand this is the only microscope that allows the researcher full control over all axes of movement for the microscope probe, which is the special feature of this microscope.

This instrument has a lot of potential in that specifically tailored software algorithms can extract more information from a sample than is conventionally possible. When compared with a commercial DSP-controlled SPM, the described STM has the advantage that the code is accessible and all parts of the SPM are well known.

There is also potential for improving the microscope, as well as great opportunities to adapt it to fulfill exciting new tasks such as AFM, or even SPM combined with magnetic resonance.

This microscope should be seen as the first prototype of its kind for our research group. The essential parts had to be brought together to build a fully operational instrument and a great deal of knowledge had to be gained prior to reaching the current level of operation.

The microscope was shown to operate adequately. Following from this a second improved version can be built and a research direction can be chosen.

Appendix A

Mechanics

A.1 Piezoelectric Actuator Properties

The piezoelectric actuators are made from ceramic pressed into a tube shape. All outer actuators are manufactured by Siemens. The center actuator is physically larger and manufactured by Staveley Sensors, who specialize in microscopy applications. The physical dimensions of the center actuator are listed in table A.1. The physical dimensions of the outer actuator are listed in table A.2.

A.2 Silver as an Electrode Material

The electrodes of all actuators in this system are made from silver. There are two commonly used metals for actuator electrodes, silver and nickel. Gold is also used occasionally. These three metals are used, because they can be deposited relatively easily and all three are extremely corrosion-resistant [49]. Silver is also a popular electrode material owing to its good conductivity. It is used for electroplating copper tracks to make them easier to solder to, although silver may dissolve into the molten solder [90]. This problem, solder leaching, can be solved by using silver-containing solder. The company Sensor Technology recommends a tin-lead solder containing 2% silver to solder to silver electrodes. Silver solder, however, usually has a higher melting point, which is a disadvantage when soldering to the heat-sensitive ceramic.

There are a few drawbacks of using silver. Though it is very resistant to oxidation it easily tarnishes even in normal atmosphere at room temperatures. The tarnish is an inorganic surface film consisting of a sulphide [64]. The chemical partner in that reaction, sulfur, is present in traces in normal atmosphere and also on human skin.

When silver is used as a contact material, tarnish may cause the contact to be intermittent or of high resistance. A minimum voltage of 20 V and a contact pressure of 15 grams would be needed to break the tarnish [90]. Although the electrode is not used as a switch, the tarnish will hinder ready soldering.

Solder usually contains flux to remove surface oxides. The solder flux used here, Comweld 965, is recommended for use on tarnished surfaces. It is however better to remove the tarnish for which several methods are possible as mentioned

Center Actuator

property	value
material	PZT-5A
dielectric constant	$\epsilon_r = K_3^T = 1725$
piezoelectric constant	$d_{31} = -0.173\,^{\mathrm{nm}}/_{\mathrm{V}}$
Young's modulus	$Y = (6.3 \pm 0.1) \times 10^{10}\,^{\mathrm{N}}/_{\mathrm{m}^2}$
density	$\rho = (7.5 \pm 0.1) \times 10^3\,^{\mathrm{kg}}/_{\mathrm{m}^3}$
capacitance	$C_{\mathrm{total}} = (2.50 \pm 0.43)\,\mathrm{pF}$
length	$l = (12.70 \pm 0.08)\,\mathrm{mm}$
outer diameter	$d_{\mathrm{o}} = (3.175 \pm 0.05)\,\mathrm{mm}$
width	$t = (0.51 \pm 0.03)\,\mathrm{mm}$
inner diameter	$d_{\mathrm{i}} = d_0 - 2t = (2.2 \pm 0.13)\,\mathrm{mm}$
electrode length	$l_{\mathrm{el}} = (9.9 \pm 0.14)\,\mathrm{mm}$
buried depth	$d_{\mathrm{bc}} = (4 \pm 0.1)\,\mathrm{mm}$
clamp ring	$c_{\mathrm{c}} = (2.00 \pm 0.01)\,\mathrm{mm}$
free length	$l_{\mathrm{f}} = l - c_{\mathrm{c}} = (10.7 \pm 0.08)\,\mathrm{mm}$
cross section	$A = (4.1 \pm 0.5)\,\mathrm{mm}^2$

Table A.1: **Properties of the Center Actuator:** Listed are the geometric properties of the center actuator tube as well as a few material properties of the piezoelectric ceramic PZT-5A used for that tube. The values are given at $T = 298$ K. More property values can be found in the manufacturer's data sheet [175].

Outer Actuator

property	value
material	'VIBRIT' 420
coupling factor	$k_{31} > 0.260$
Young's modulus	$Y = (7.5 \pm 1.5) \times 10^{10}\,^{\mathrm{N}}/_{\mathrm{m}^2}$
capacity	$(1100 \pm 200)\mathrm{pF}$
polarization dir.	positive inside
outer diameter	$d_{\mathrm{o}} = (2.2 \pm 0.1)\,\mathrm{mm}$
inner diameter	$d_{\mathrm{i}} = (1.0 \pm 0.1)\,\mathrm{mm}$
length	$l = (13.0 \pm 0.2)\,\mathrm{mm}$
electrode length	$l_{\mathrm{el}} = (8.6 \pm 0.1)\,\mathrm{mm}$
buried depth	$d_{\mathrm{bo}} = (2.5 \pm 0.1)\,\mathrm{mm}$
clamp ring	$c_{\mathrm{o}} = (1.78 \pm 0.01)\,\mathrm{mm}$
free length	$l_{\mathrm{f}} = l - c_{\mathrm{o}} = (11.2 \pm 0.2)\,\mathrm{mm}$

Table A.2: **Properties of the Outer Actuators:** Listed are geometric and material properties of the outer actuator tubes made by Siemens. Young's modulus is an estimate taken as a typical value for PZT ceramic [137, 152]

above. Since bare silver is not protected against tarnishing it will subsequently tarnish again, which means that the tarnish should be removed just prior to soldering. If the actuator is completely tarnished nothing short of a renewed electrode deposition can restore it.

Another drawback of silver is the electromigration of silver ions in high electric fields under humid conditions. Blech [32] states that silver that is deposited on ceramic will wander between the conductors thereby forming threads. This process can cause a short circuit between the electrodes. Silver is therefore not recommended for use in applications where high dc voltages have to be used.

Silver is also only moderately resistant to mechanical wear, but that is of no consequence for its use as an actuator electrode.

The manufacturer of the actuators used in this microscope [175] favors electroless[1] nickel as the electrode material, because it is thin (0.5 μm), adheres well to the ceramic and is, as mentioned, corrosion-resistant. The nickel is usually coated with gold to allow easy soldering and enhance the conductivity of the electrode.

Silver electrodes have the disadvantage of being thicker (about 20 μm) than nickel electrodes [175]. Despite all these disadvantages, fairly complex electrode patterns, like the one on the center actuator of the probe head used in this project, can only be printed on actuator ceramics using silver. To the best of our knowledge the choice of nickel was only available for the inside electrode of the actuator.

A.3 Capacity of the Actuator

The piezoelectric material used for the center actuator is PZT-5A. The outer and inner surface areas of the tube are electro-plated with a thin metal coating, sketched in figure 2.9. The inner coating forms an inner electrode. The outer coating is cut into four segments forming four outer electrodes. The capacitance between the outer and inner electrode of a tube is given by [120, 195]

$$C = \frac{2\pi\epsilon l_{\mathrm{el}}}{\ln \frac{r_2}{r_1}}, \tag{A.1}$$

where the material property of the ceramic influences the permittivity $\epsilon = \epsilon_0\epsilon_r$, where $\epsilon_0 = 8.85^{\mathrm{pF}}/_{\mathrm{m}}$ is the permittivity of free space and $\epsilon_r = 1725$ is the dielectric constant (or relative permittivity) of PZT-5A. The geometric quantities are the length of the tube electrode $l_{\mathrm{el}} = 9.9\,\mathrm{mm}$, the outer radius of the tube $r_2 = 1.6\,\mathrm{mm}$ and the inner radius $r_1 = 1.1\,\mathrm{mm}$. Inserting these values into equation (A.1) gives

$$\begin{aligned} C &= \frac{55.6^{\mathrm{pF}}/_{\mathrm{m}} \times 1725 \times 9.9 \times 10^{-3}\mathrm{m}}{\ln \frac{1.6}{1.1}} \\ &= \frac{950\,\mathrm{pF}}{0.375} = 2.54\,\mathrm{nF}. \end{aligned}$$

To calculate the error in the capacitance we first have to calculate the error in the inner and outer radius from the values in table A.1. The relative error in

[1] Electroless plating is a way of depositing relatively thick layers of metal (eg Cu, Ni, Ag, Au, Pd) onto non-metals, invented 1944 by Brenner and Riddell.

the outer radius is equal to that of the outer diameter

$$\frac{\Delta r_2}{r_2} = \frac{\Delta d_o}{d_o} = 0.016\,,$$

while the relative error in the inner radius is larger. Since $r_1 = r_2 - t$, the error in r_1 is

$$\Delta r_1 = \left[(\Delta r_2)^2 + (\Delta t)^2\right]^{\frac{1}{2}} = 0.04\,\mathrm{mm}\,.$$

The error in capacitance is then found to be

$$\begin{aligned}\frac{\Delta C}{C} &= \left[(\frac{\Delta \epsilon_r}{\epsilon_r})^2 + (\frac{\Delta l_{el}}{l_{el}})^2 + (\frac{1}{\ln\frac{r_2}{r_1}}\frac{\Delta r_2}{r_2})^2 + (\frac{1}{\ln\frac{r_2}{r_1}}\frac{\Delta r_1}{r_1})^2\right]^{\frac{1}{2}} \\ &= \left[(3\times 10^{-3})^2 + (0.014)^2 + (\frac{1}{0.375}0.016)^2 + (\frac{1}{0.375}0.036)^2\right]^{\frac{1}{2}} = 0.11\,.\end{aligned}$$

The total capacitance of the center actuator is therefore

$$C_{\mathrm{total}} = (2500 \pm 270)\,\mathrm{pF}. \tag{A.2}$$

If we neglect fringe-effects then we get for each of the four segments of the outer electrode a capacitance of

$$C = \frac{1}{4} \times (2500 \pm 270)\,\mathrm{pF} = (640 \pm 70)\,\mathrm{pF}$$

The capacitance of the three outer probe head actuators is given on the data sheet as

$$C = (1100 \pm 200)\,\mathrm{pF} \tag{A.3}$$

A.4 Sensitivity of the Actuators

The sensitivity of the piezoelectric actuators is the length change (Δl) of the actuator per unit of applied voltage (V) across the electrodes of the actuator, i.e.

$$r = \frac{\Delta l}{V} \tag{A.4}$$

A sensitivity can be defined for both horizontal and vertical directions of actuator movement. Since the actuator is a tube, the horizontal sensitivity in the x-direction and y-direction are equal as long as the piezoelectric ceramic is homogeneous.

A.4.1 Vertical Sensitivity

The vertical sensitivity can be calculated using the converse piezoelectric effect from equation (3.6):

$$\frac{\Delta l_z}{l_z} = d_{31} E,$$

where d_{31} is the piezoelectric constant of the material (PZT-5A). The electric field strength is given by $E = V_z/t$, where t is the wall thickness of the tube and

V_z is the voltage between inner electrode and the average voltage of all outer electrodes. The vertical length change is therefore given by

$$\Delta l_z = \frac{d_{31} l_{el} V_z}{t}. \tag{A.5}$$

Note that the length value is again the length of the electrode, since an electric field is only present in this region. The vertical sensitivity will therefore be

$$\begin{aligned} r_z &= \frac{\Delta l_z}{V_z} = \frac{d_{31} l_{el}}{t} \\ &= \frac{-0.173\,\text{nm/V} \times 9.9\,\text{mm}}{0.51\,\text{mm}} = -3.36\,\text{nm/V}\,. \end{aligned}$$

Since the outer actuators are only used in bending mode, the vertical sensitivity of the outer actuators is not listed here.

A.4.2 Horizontal Sensitivity

The horizontal length change can be calculated from a similar equation, which is valid for equal and opposite voltages applied to opposite quadrants. A quadrant is one of the outer electrodes. The horizontal sensitivity is calculated for quadrant $+x$ without losing generality. The horizontal length change of the actuator (x_{el}) is given by

$$x_{el} = \frac{2\sqrt{2} d_{31} l^2 V_x}{\pi d_m t}, \tag{A.6}$$

where d_m is the middle diameter of the tube and V_x is the voltage between the inner electrode and the $+x$ quadrant, while at the same time $-V_x$ is applied to quadrant $-x$. The middle diameter is

$$d_m = \frac{1}{2}(d_o + d_i)$$

with the outer and inner diameters d_o and d_i.

Equation (A.6) is an approximate formula given by Chen [44] as well as by Taylor [179]. The derivation calculates the torque created by the stresses in the two opposing quadrants. The curvature of bending (R) is then calculated from the neutral plane of the bent actuator, which is the plane where the stress is zero. A less accurate estimate for the curvature is to take the far quadrant as the curvature plane and define the curvature angle via the arc length. A diagram is shown in figure A.1.

The radius of curvature is found from

$$\begin{aligned} l + \Delta l &= R\alpha && \text{outer quadrant} \\ l - \Delta l &= (R - d_m)\alpha && \text{inner quadrant} \end{aligned}$$

from which the radius R is found to be

$$R = \frac{d_m}{2}\frac{l}{\Delta l} = \frac{d_m}{2}\frac{t}{d_{31} V}. \tag{A.7}$$

The approximation

$$\Delta l \ll l$$

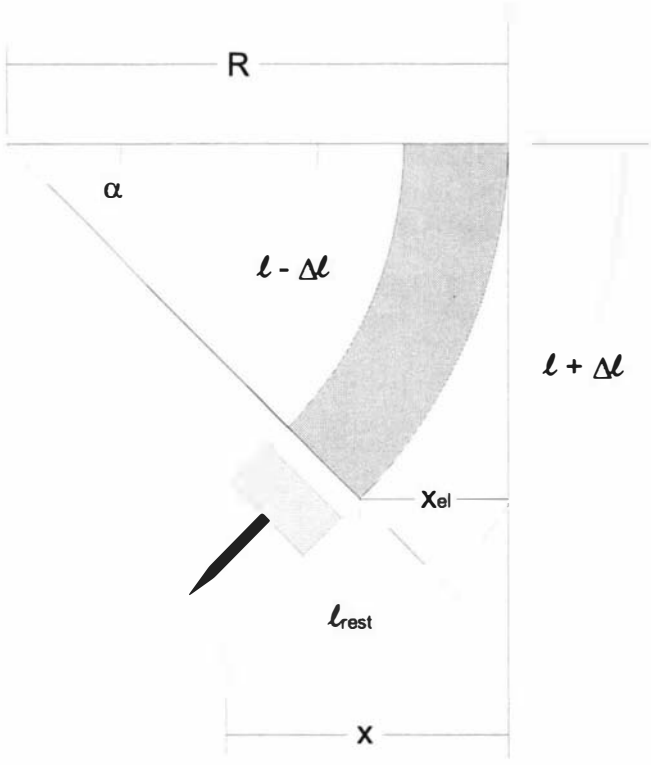


Figure A.1: **Horizontal Actuator Deflection:** Equal and opposite voltages are applied to two opposing quadrants. The bending of the actuator is caused by a torque created by the stress difference in the opposing ceramic walls. The accompanying length change $\pm\Delta l$ is indicated in the diagram. The radius of curvature R is calculated from the angle α and the inner and outer tube lengths. The rest of the actuator, part of the ceramic, the insulator, tip holder and tip, is not curved and follows the tangent of the curved electrodes. To find x_{el} the approximation $x_{el} = l^2/(2R)$ is used. The straight-line equation for the tangent is used to find the horizontal deflection of the tip (x).

is used throughout all calculations in this section. Instead of the factor $d_m/2 = 0.500 \times d_m$ the more accurate calculation delivers $(\pi/(4\sqrt{2})) \times d_m = 0.555 \times d_m$.

The horizontal deflection x_{el} is found from

$$x_{el} = R - R\cos\alpha \approx \frac{l^2}{2R}. \tag{A.8}$$

Substituting R from equation (A.7) above gives the equation for horizontal sensitivity of the ceramic in equation (A.6).

An important point to make is that the total length of the scanning actuator is considerably longer than the electrode length alone. The total length includes the free length plus the length of the macor insulator, the tip holder and the tip itself. The non-curved part of the actuator follows the tangent of the curved part as shown in figure A.1. The equation for the straight line formed by the tangent is

$$x(l) = \frac{l_{el}}{R} l - \frac{l_{el}^2}{2R}.$$

If the total actuator length is substituted for l then the horizontal tip deflection x can be found as

$$x = \frac{2l_{el}l_{tot} - l_{el}^2}{2R}.$$

Now the radius of curvature R from equation (A.7) can be inserted to find essentially the same equation as equation (A.6) using a different length l. Instead of l_{el}^2 in equation (A.6) an effective length l' is used, with

$$l'^2 = 2l_{el}l_{tot} - l_{el}^2 = l_{el}^2(2\frac{l_{tot}}{l_{el}} - 1) = gl_{el}^2,$$

giving an extra multiplicative factor g

$$g = 2\frac{l_{tot}}{l_{el}} - 1.$$

Center Actuator

The horizontal sensitivity of the actuator alone is found to be

$$\begin{aligned} r_{x\mathrm{el}} &= \frac{\Delta l_x}{V_x} = \frac{2\sqrt{2}}{\pi}\frac{d_{31}l_{el}^2}{d_m t} \\ &= 0.900 \times \frac{-0.173\,\mathrm{nm/V} \times (9.9\,\mathrm{mm})^2}{2.67\,\mathrm{mm} \times 0.51\,\mathrm{mm}} = \frac{15.3\,\mathrm{nm/V}}{1.36} = -11.2\,\mathrm{nm/V} \end{aligned} \tag{A.9}$$

The extra factor g for a tip length of 3 mm is

$$g = 2\frac{16.9\,\mathrm{mm}}{9.9\,\mathrm{mm}} - 1 = 2.41\,.$$

Hence, for the tip the horizontal sensitivity is

$$r_x = gr_{x,\mathrm{el}} = -27.0\,\mathrm{nm/V}. \tag{A.10}$$

Since l_{tu} includes the tip length, the horizontal tip sensitivity r_x also depends on the tip length, which varies with the sample height. The tip length lies in most cases between 3 and 4 mm, for the above value a length of 3 mm was used. For a tip length of 3.5 mm the factor g is 2.52, for a length of 4 mm it is 2.62 .

Outer Actuators

The outer actuators carry a steel ball on the lower end. The unclamped total length (l_{tu}) of the outer actuator is $l_o - c_o = 12.7\,\text{mm}$. The lengths are depicted in figure 3.5. The extra factor g is

$$g = 2\frac{12.7\,\text{mm}}{8.6\,\text{mm}} - 1 = 1.95\,.$$

In equation (A.6) the length l can be substituted with an effective length

$$l' = \sqrt{g}\,l_{el} = 1.40 \times 8.6\,\text{mm} \times 12.0\,\text{mm}.$$

The horizontal sensitivity is found to be

$$\begin{aligned} r_x &= \frac{\Delta l_x}{V_x} = \frac{2\sqrt{2}}{\pi}\frac{d_{31}l'^2}{d_m t} \qquad (A.11)\\ &= 0.90 \times \frac{-0.12\,^{\text{nm}}/_{\text{V}} \times (12.0\,\text{mm})^2}{1.6\,\text{mm} \times 0.6\,\text{mm}} = \frac{18.1\,^{\text{nm}}/_{\text{V}}}{0.96} = -18.9\,^{\text{nm}}/_{\text{V}} \end{aligned}$$

Error Calculation The error in vertical sensitivity of the center actuator is calculated as

$$\begin{aligned} \frac{\Delta r_z}{r_z} &= \left[\left(\frac{\Delta d_{31}}{d_{31}}\right)^2 + \left(\frac{\Delta l_{el}}{l_{el}}\right)^2 + \left(\frac{\Delta t}{t}\right)^2\right]^{\frac{1}{2}}\\ &= \left[(0.029)^2 + (0.014)^2 + (0.06)^2\right]^{\frac{1}{2}} = 0.068\,. \end{aligned}$$

The error in piezoelectric constant has been guessed as $\pm 0.005^{\text{nm}}/_{\text{V}}$, though it might in fact be even greater since the manufacturer states that the values are only nominal.

The vertical sensitivity of the center actuator is

$$r_z = (-3.36 \pm 0.23)\,^{\text{nm}}/_{\text{V}}. \qquad (A.12)$$

The error in horizontal sensitivity of the center actuator is calculated as

$$\begin{aligned} \frac{\Delta r_x}{r_x} &= \left[\left(\frac{\Delta d_{31}}{d_{31}}\right)^2 + \left(2\frac{\Delta l_{el}}{l_{el}}\right)^2 + \left(\frac{\Delta g}{g}\right)^2 + \left(\frac{\Delta d_o}{d_o}\right)^2 + 2\left(\frac{\Delta t}{t}\right)^2\right]^{\frac{1}{2}}\\ &= \left[(0.029)^2 + (0.028)^2 + (0.016)^2 + (0.083)^2 + (0.024)^2\right]^{\frac{1}{2}} = 0.097\,. \end{aligned}$$

The error in the middle diameter was

$$\frac{\Delta d_m}{d_m} = \left[\left(\frac{\Delta d_o}{d_o}\right)^2 + 2\left(\frac{\Delta t}{t}\right)^2\right]^{\frac{1}{2}}$$

and the error in the factor g was

$$\frac{\Delta g}{g} = \frac{g+1}{g}\left[\left(\frac{\Delta l_{el}}{l_{el}}\right)^2 + \left(\frac{\Delta l_{tu}}{l_{tu}}\right)^2\right]^{\frac{1}{2}}$$

The horizontal sensitivity of the center actuator is

$$r_h = (-27.0 \pm 2.6)\,^{\text{nm}}/_{\text{V}}. \qquad (A.13)$$

Due to the approximations involved, there is also a small systematic error associated with the horizontal sensitivity, which is not estimated but has to be borne in mind.

The error for the outer actuator is calculated in the same way as for the center actuator. The error in the middle diameter is calculated differently, since outer and inner diameter are known, not outer diameter and wall thickness. The error in the middle diameter is

$$\frac{\Delta d_{\mathrm{m}}}{d_{\mathrm{m}}} = \left[(\frac{\Delta d_{\mathrm{o}}}{d_{\mathrm{o}}})^2 + (\frac{\Delta d_{\mathrm{i}}}{d_{\mathrm{i}}})^2 \right]^{\frac{1}{2}}$$

The tube width is calculated from the inner and outer diameter as

$$t = \frac{1}{2}(d_{\mathrm{o}} - d_{\mathrm{i}}).$$

The error in the tube width is

$$\Delta t = \left[(\frac{1}{2}\Delta d_{\mathrm{o}})^2 + (\frac{1}{2}\Delta d_{\mathrm{i}})^2 \right]^{\frac{1}{2}} = \left[(0.05)^2 + (0.05)^2 \right]^{\frac{1}{2}} \mathrm{mm} = 0.07\ \mathrm{mm}$$

We get the error in r_x for the outer actuator as

$$\begin{aligned} \frac{\Delta r_x}{r_x} &= \left[(\frac{\Delta d_{31}}{d_{31}})^2 + (2\frac{\Delta l_{\mathrm{el}}}{l_{\mathrm{el}}})^2 + (\frac{\Delta d_{\mathrm{o}}}{d_{\mathrm{o}}})^2 + (\frac{\Delta d_{\mathrm{i}}}{d_{\mathrm{i}}})^2 + (\frac{\Delta t}{t})^2 + (\frac{\Delta g}{g})^2 \right]^{\frac{1}{2}} \\ &= \left[(0.15)^2 + (0.023)^2 + (0.045)^2 + (0.1)^2 + (0.236)^2 + (0.038)^2 \right]^{\frac{1}{2}} \\ &= 0.30. \end{aligned}$$

As a summary the horizontal sensitivity of the outer actuator with error is

$$r_h = (-18.9 \pm 5.7)\,\mathrm{nm/V}. \tag{A.14}$$

Piezoelectric Constant of the Outer Actuator

Instead of the piezoelectric strain constant (d_{31}) in the data sheet for the outer actuator only the electromechanical coupling factor (k_{31}) for the ceramic is given. The electromechanical coupling factor is defined as the square root of the ratio between stored mechanical energy and input of electrical energy [152, 185]. The input electrical energy per unit volume is

$$\frac{1}{2}\epsilon_0\epsilon_r E^2$$

and the stored mechanical energy per unit volume is

$$\frac{1}{2}\left(\frac{\Delta l}{l}\right)^2 \frac{1}{s} = \frac{1}{2}\frac{(dE)^2}{s},$$

where $\Delta l/l$ is the strain, s is the compliance (strain per unit stress) and d is the piezoelectric strain constant. We used the converse piezoelectric effect described in equation (3.6). For the coupling factor k we get

$$k^2 = \frac{\frac{1}{2}\frac{(dE)^2}{s}}{\frac{1}{2}\epsilon_0\epsilon_r E^2} = \frac{d^2}{\epsilon_0\epsilon_r s},$$

from which one can find

$$d_{31}^2 = k_{31}^2 \epsilon_0 \epsilon_r s.$$

The dielectric constant was found from the given value of capacitance and the dimensions of the tube (listed in table A.2). The capacitance without dielectric constant ϵ_r was calculated according to equation (A.1) as

$$C = \epsilon_r \times 0.61\,\mathrm{pF} = 1100\,\mathrm{pF}.$$

We find $\epsilon_r = 1800$, which is reasonable for piezoelectric material in general and PZT specifically [137].

For the compliance we had to assume a sensible value, since it was not given in the data sheet. The compliance is taken as the reciprocal of Young's modulus. Most piezoelectric ceramics have values in the range $(6 \cdots 9) \times 10^{10}\,\mathrm{N/m^2}$ for Young's modulus [175], so we assume a value of $(7.5 \pm 1.5) \times 10^{10}\,\mathrm{N/m^2}$.

For piezoelectric strain constant d_{31} we find

$$d_{31} = -0.26\sqrt{1800 \times 8.85\,\mathrm{C^2/m^2N} \times 1.33\,\mathrm{m^2/N}} = -0.12 \times 10^{-9}\,\mathrm{C/N} = -0.12\,\mathrm{nm/V}.$$

As the relative error in d_{31} we get

$$\begin{aligned} \frac{\Delta d_{31}}{d_{31}} &= \left[\left(\frac{1}{2}\frac{\Delta s}{s}\right)^2 + \left(\frac{1}{2}\frac{\Delta \epsilon_r}{\epsilon_r}\right)^2 \right]^{\frac{1}{2}} \\ &= \left[(0.10)^2 + (0.11)^2\right]^{\frac{1}{2}} = 0.15, \end{aligned}$$

which gives an absolute error of

$$d_{31} = (-0.12 \pm 0.02)\,\mathrm{nm/V} \tag{A.15}$$

This is a sensible result; Moseley [137] cites typical values between $-0.06\,\mathrm{nm/V}$ and $-0.17\,\mathrm{nm/V}$ for PZT.

A.5 Vibrational Modes

A.5.1 Vibrational Modes of the Actuators

The piezoelectric actuators used in the probe head are tube shaped. The tubes can vibrate longitudinally (stretching mode) or transversely (bending mode). All tubes are clamped in the probe head disk by a clamp ring. When calculating the vibration of the tubes only the unclamped part of the tube, which is freely moving, should be considered.

The complete probe head can vibrate or only the center actuator. In the first case the outer tubes act as springs supporting the probe head disk, which can vibrate longitudinally, laterally and also rotary. In the second case the actuator can also vibrate longitudinally and laterally. The center actuator is loaded with an extra mass at the end, the tip holder.

The vibrational modes of the center actuator are discussed first, then the modes of the probe head.

A.5.2 Longitudinal Vibration

The vibration frequency in the stretching mode depends on the speed of sound (c) in the ceramic, which is

$$c = \sqrt{\frac{Y}{\rho}}, \tag{A.16}$$

where Y is Young's modulus and ρ is the density of the ceramic, both of which are listed in table A.1 for the central actuator. Now since the center actuator consists of three pieces made from different materials, the sound waves in the material will be partially reflected at any of the borders between the pieces. There are two ways to get an estimate of the longitudinal vibration frequencies. If most of the sound wave is reflected at the border, then the more elastic part of the actuator (the ceramic) can be viewed as a spring and the remainder (the tip holder) as a concentrated mass load. If only a small fraction of the sound wave is reflected then the whole actuator can be viewed as a spring.

A standing wave will establish itself in the actuator. The natural frequency of the longitudinal vibration can be either calculated using the speed of sound and the relevant boundary conditions or with the spring constant of the elastic part of the actuator. The speed of sound approach is exact and should be used where the actuator is viewed as a vibrating rod. The spring constant approach should be used where the actuator is viewed as a spring-mass system. The calculation is then a good approximation [176].

Actuator as Vibrating Rod

For the speed of sound in the central actuator we find

$$c = \sqrt{\frac{6.3 \times 10^{10}\ \mathrm{N/m^2}}{7.5 \times 10^{3}\ \mathrm{kg/m^3}}} = 2900\ \mathrm{m/s}.$$

A standing wave in the material is established under the condition that the length (l) of the piece is equal to a quarter wavelength ($\lambda/4$), or all odd multiples of this

$$l = (2k+1)\frac{\lambda}{4} \quad \text{for } k = 0, 1, 2, \ldots.$$

We find the wavelength to be

$$\lambda_k = \frac{4l}{2k+1}.$$

With the speed of sound being

$$c = f\lambda,$$

we get the resulting stretching mode frequencies as

$$f_k = \frac{c}{\lambda_k} = \frac{c}{4l}(2k+1).$$

The center actuator consists of three different materials instead of just the ceramic. The speed of sound differs in all three materials and reflections will occur at the boundaries between the materials. To estimate the lowest order of the stretching mode vibrations we can calculate an effective speed of sound

	l(mm)	material	Y (GPa)	$\rho(10^3\,\mathrm{kg/m^3})$	$c(\mathrm{m/s})$	$t = \frac{l}{c}(\mu\mathrm{s})$
l_f	10.7	PZT-5A	63	7.5	2900	3.66
w	1.2	macor	66.9	2.52	5150	0.23
m	2.0	steel	210	8.1	5100	0.39
l_v	13.9				3220	4.28

Table A.3: **Effective speed of sound in center actuator:** The three materials making up the center actuator are listed along with their lengths, Young's modulus and density. The labels are the same as the ones in table A.6. The speed of sound is found from equation (A.16). The last column presents the time taken for a point of constant phase to pass through the material. The effective speed of sound is calculated from the total actuator length divided by the traveling time, $c_\mathrm{eff} = l/t$. The piezoelectric tube length l_f is the unclamped part of the tube (not buried in the probe head disk). The effective speed of sound for the center actuator is listed in the row for the total length l_v as $c_\mathrm{eff} = 3220\,\mathrm{m/s}$.

from the time it takes a point of constant phase in the stretching wave to pass the individual materials. This has been done in table A.3.

The condition for establishment of a standing wave in the actuator is still valid. With an effective speed of sound of $c_\mathrm{eff} = 3220\,\mathrm{m/s}$ in the actuator and a length of $l_\mathrm{v} = 13.9\,\mathrm{mm}$ we find the fundamental stretching mode vibration to be

$$f_0 = \frac{c}{4l} = \frac{3220\,\mathrm{m/s}}{4 \times 13.9 \times 10^{-3}\,\mathrm{m}} = (58 \pm 1)\,\mathrm{kHz}. \tag{A.17}$$

The error is estimated to be 2% due to 1% errors in c and l.

Actuator as Spring-Mass System

The center actuator can also be interpreted as a vibrating system consisting of an elastic spring and a rigid mass. The longitudinal vibration is well approximated using Rayleigh's method, which is described by Stokey [176]. The natural frequency is

$$\omega = \sqrt{\frac{k}{m + \frac{m_\mathrm{s}}{3}}}, \tag{A.18}$$

where k is the spring constant, m the rigid mass and m_s the mass of the spring.

The effective spring constant k can be derived from Hooke's law

$$X = Y\frac{\Delta l_z}{l_z},$$

where X is the stress, which causes the strain $\Delta l_z / l_z$ and Y is Young's modulus. Substituting

$$X = \frac{F}{A},$$

for the stress, the equation can be rearranged

$$F = \frac{AY}{l}\Delta l = k\Delta l,$$

where the spring constant is

$$k = \frac{AY}{l}. \tag{A.19}$$

In case there is no extra mass, equation (A.18) delivers a good approximation to equation (A.17), the exact equation for a vibrating rod. This can be seen, when the spring constant k in equation (A.18) is substituted with k from equation (A.19). The mass of the beam (m_s) acting as a spring is expressed by the density of the beam (ρ) and its volume $V = Al$, which gives

$$\omega = \sqrt{\frac{k}{\frac{m_s}{3}}} = \sqrt{\frac{3AY}{l\rho Al}} = \sqrt{3}\,\frac{c}{l},$$

where in the last step the definition of the speed of sound, equation (A.16), has been used. The angular frequency calculated from this equation is 10% smaller than the angular frequency calculated from the exact solution of equation (A.17).

In the following there is a rigid mass, which is the tip holder including the tip clamping screw. The macor insulator and the piezoelectric ceramic of the actuator are first treated as the elastic spring of the system. The result for the same calculation taking the macor insulator as part of the mass load is also given. The mass of the spring cannot be neglected, since it is bigger than the rigid mass load.

Spring Constant Firstly the spring constant k from equation (A.19) is calculated for the center actuator.

The section of the center actuator is

$$A = \frac{\pi}{4}(d_o^2 - d_i^2), \tag{A.20}$$

which we find to be $(4.1 \pm 0.5) \times 10^{-6}\,\mathrm{m}^2$, where the values from table A.1, page 224, were used.

The length of the spring is the unclamped length of the ceramic plus the length of the insulator

$$l = l_f + w = 10.7\,\mathrm{mm} + 1.2\,\mathrm{mm} = 11.9\,\mathrm{mm}.$$

The spring constant for the actuator is found to be

$$k = \frac{4.1 \times 10^{-6}\,\mathrm{m}^2 \times 6.3 \times 10^{10}\,\mathrm{N/m^2}}{11.9 \times 10^{-3}\,\mathrm{m}} = (21.7 \pm 2.6) \times 10^6\,\mathrm{N/m}.$$

Masses The masses of the macor insulator and the actuator tube have been calculated from the physical dimensions and known densities. Although actual weighing of the pieces would be more accurate, the center actuator is glued and cannot be disassembled without risk. Moreover the mass of the unclamped part of the actuator can only be found by calculation. The mass of the center actuator is

$$m = l\rho A = l \times 30.8 \times 10^{-6}\,\mathrm{kg/mm}.$$

With a length of $l = 12.7\,\mathrm{mm}$ the total mass of the center actuator is

$$m_p = 3.9 \times 10^{-4}\,\mathrm{kg},$$

while the mass of the free part of the actuator with a shorter length of $l_f = 10.7\,\mathrm{mm}$ is

$$m_f = 3.3 \times 10^{-4}\,\mathrm{kg}.$$

The mass for the macor insulator was found to be

$$m_{\mathrm{m}} = 0.02 \times 10^{-4}\,\mathrm{kg}.$$

The mass of the spring is then

$$m_{\mathrm{s}} = m_{\mathrm{f}} + m_{\mathrm{m}} = (3.32 \pm 0.02) \times 10^{-4}\,\mathrm{kg}.$$

The error is an estimate.

The mass of the tip holder has been found to be

$$m = (1.30 \pm 0.01) \times 10^{-4}\,\mathrm{kg}$$

by weighing a copy of the original with high-resolution scales (Mettler PC 180). The stated error is the resolution of the scales. Just for comparison the weight has also been calculated and found to be 1.27×10^{-4} kg, which compares well with the measured value. The small screw that holds the tip in the tip holder weighs 0.33×10^{-4} kg and should also be included in the total weight of the tip holder, so that this weight is

$$m_{\mathrm{th}} = (1.63 \pm 0.014) \times 10^{-4}\,\mathrm{kg}.$$

The effective mass in equation (A.18) is

$$m + \frac{m_{\mathrm{s}}}{3} = (1.63 + \frac{3.32}{3}) \times 10^{-4}\,\mathrm{kg} = 2.74 \times 10^{-4}\,\mathrm{kg}.$$

Vibration Frequency The spring constant k and the effective mass are inserted in equation (A.18) to obtain

$$\omega = \sqrt{\frac{21.7 \times 10^{6}\,\mathrm{N/m}}{2.74 \times 10^{-4}\,\mathrm{kg}}} = 281\,000\,\mathrm{s}^{-1}. \tag{A.21}$$

The natural longitudinal vibration frequency is

$$f_{\mathrm{cv}} = \frac{1}{2\pi} \times 285\,000\,\mathrm{s}^{-1} = (45 \pm 3)\,\mathrm{kHz}. \tag{A.22}$$

Taking the macor insulator as part of the mass load, a similar calculation yields a natural longitudinal vibration frequency of

$$f_{\mathrm{cv}} = \frac{1}{2\pi} \times 295\,000\,\mathrm{s}^{-1} = (47 \pm 3)\,\mathrm{kHz}. \tag{A.23}$$

The error has been calculated from

$$\frac{\Delta f}{f} = \frac{\Delta \omega}{\omega} \approx \frac{1}{2}\frac{\Delta k}{k} \approx \frac{1}{2}\frac{\Delta A}{A} = 0.06\,, \tag{A.24}$$

since the error in cross section is dominant.

A.5.3 Transverse Vibration

Formulae to calculate the natural frequency of commonly used systems are given in a shock and vibration handbook edited by Harris [91]. To estimate the lowest order natural frequency (ω) in bending mode we used the formula for "massive springs (beams) with concentrated mass loads" for a system with one free end and one fixed end:

$$\omega = \sqrt{\frac{3YI}{l^3(m + 0.23m_b)}}, \tag{A.25}$$

where Y is Young's modulus, I is the area moment of inertia of the beam section, l is the beam length, m is the mass load and m_b is the mass of the beam. The area moment I has to be calculated first. It is defined as

$$I = \frac{\pi}{64}(d_o^4 - d_i^4),$$

and we find its numerical value to be

$$I = \frac{\pi}{64} \times (3.2^4 - 2.2^4) \times 10^{-12}\,\mathrm{m}^4 = 3.84 \times 10^{-12}\,\mathrm{m}^4.$$

Now that the beam consists of three different parts, piezoceramic tube, macor insulator and steel tip holder, the macor insulator can be either treated as part of the beam or as part of the mass load. We calculate the natural frequency for both cases to get an estimate of the systematic error, although one can argue that the insulator should be treated as part of the flexing beam, since Young's modulus for both materials, ceramic and macor, is very similar. Let us consider this as the first case. The length of the beam is then $l = l_f + w = 11.9\,\mathrm{mm}$.

We use $m_b = m_f + m_m$ in the first case. The mass term in equation (A.25) is

$$(1.63 + 0.23 \times 3.32) \times 10^{-4}\mathrm{kg} = 2.39 \times 10^{-4}\,\mathrm{kg}.$$

The natural frequency is

$$\begin{aligned} f &= \frac{1}{2\pi}\sqrt{\frac{3 \times 63 \times 10^9\,\mathrm{N/m^2} \times 3.84 \times 10^{-12}\,\mathrm{m}^4}{11.9^3 \times (10^{-3}\,\mathrm{m})^3 \times 2.39 \times 10^{-4}\,\mathrm{kg}}} \\ &= \frac{1}{2\pi}\sqrt{1.80 \times 10^9\,\mathrm{s}^{-2}} = \frac{1}{2\pi}42\,400\,\mathrm{s}^{-1} = (6.8 \pm 0.3)\,\mathrm{kHz}. \end{aligned}$$

The same calculation for the case where the macor insulator is treated as part of the mass load yields

$$f = (7.9 \pm 0.3)\,\mathrm{kHz}. \tag{A.26}$$

The error is calculated below. The average frequency for the two cases is

$$f = (7.4 \pm 0.6)\,\mathrm{kHz}, \tag{A.27}$$

where we include half the difference as a systematic error. The biggest contribution to the systematic error is the difference in the value of the length, which has a strong influence on the frequency.

Uncertainty in Bending Mode Frequency Since the systematic error in the calculation is relatively high, it is not worthwhile to do a detailed error calculation for the random errors. The following errors were estimated for the main variables:

$$\frac{\Delta I}{I} \approx 4\frac{\Delta d_o}{d_o} = 0.07$$
$$\frac{\Delta m}{m} \approx 0.025$$
$$\frac{\Delta Y}{Y} \approx 0.01$$
$$\frac{\Delta l}{l} \approx 0.014$$

We get

$$\begin{aligned}\frac{\Delta f}{f} &= \left[(\frac{1}{2}\frac{\Delta E}{E})^2 + (\frac{1}{2}\frac{\Delta I}{I})^2 + (\frac{3}{2}\frac{\Delta l}{l})^2 + (\frac{1}{2}\frac{\Delta m}{m})^2\right]^{\frac{1}{2}} \\ &= \left[(0.005)^2 + (0.035)^2 + (0.021)^2 + (0.013)^2\right]^{\frac{1}{2}} = 0.043\,,\end{aligned}$$

which gives us a random error for the frequency of

$$\Delta f = 0.043 \times 7.4\,\text{kHz} = \pm 0.3\,\text{kHz}.$$

The systematic error from above (± 0.6 kHz) is twice the random error and hence of the same order of magnitude.

A.5.4 Vibration of the Probe Head

Longitudinal Mode

The mass of the probe head disk is calculated from its dimensions and density. The volume of the disk is approximated as the volume of two small disks as

$$V = \frac{\pi}{4}(a_1^2 h_1 + a_2^2 h_2) = \frac{\pi}{4}(23^2 \times 2.4 + 18.4^2 \times 1.6) \times 10^{-7}\,\text{m}^3 = 14.2 \times 10^{-7}\,\text{m}^3,$$

where a_1 and a_2 are the disk diameters and h_1 and h_2 are the disk heights. With a density of steel $\rho = 7.8 \times 10^{-3}\,\text{kg/m}^3$ the disk mass is estimated to be

$$m = 7.8 \times 10^3\,\text{kg/m}^3 \times 14.2 \times 10^{-7}\,\text{m}^3 = 11 \times 10^{-3}\,\text{kg}.$$

The outer actuator tubes act like springs. The effective spring constant k is calculated from equation (A.19)

$$k = \frac{AY}{l}. \tag{A.28}$$

The cross section (A) of each outer actuator is calculated with equation (A.20), that is

$$A = \frac{\pi}{4}(d_o^2 - d_i^2). \tag{A.29}$$

The cross section is found to be $(3.0 \pm 0.4) \times 10^{-6}\ \mathrm{m}^2$. The tabulated values in table A.2 were used. For the spring constant of the outer tubes

$$k = \frac{3.0 \times 10^{-6}\ \mathrm{m}^2 \times 7.5 \times 10^{10}\ \mathrm{N/m}^2}{11.2 \times 10^{-3}\ \mathrm{m}} = 20.1 \times 10^{6}\ \mathrm{N/m}.$$

is found.

Since three tubes support the disk the overall spring constant is three times as big. The natural vibration frequency is found to be

$$f = \frac{1}{2\pi}\sqrt{\frac{k}{m}} = \frac{1}{2\pi}\sqrt{\frac{3 \times 20.1 \times 10^{6}\ \mathrm{N/m}}{11 \times 10^{-3}\ \mathrm{kg}}} = (12 \pm 2)\ \mathrm{kHz}. \qquad \text{(A.30)}$$

The error comes mainly from the mass estimate of the probe head, where a 10% error was allowed.

Lateral Mode

The outer tubes can also bend, which can result in lateral or torsional vibration of the probe head.

The vibration can be described as a vibration of a beam. On the disk side the tubes are clamped, which means that one end of the beam is fixed. The other end of the tubes are practically free, since there is only a small lateral friction force between the steel balls on the feet of the tubes and the ramp. Depending on how strong the balls are held to the ramp the other end of the beam can be described as either hinged or free. If the attachment between balls and ramp is strong then the connection behaves as if hinged, otherwise it behaves as if free. The attachment due to static friction depends on the mass of the disk and on the materials of the ball (steel) and the ramp (tungsten carbide). The fixed-hinged mode of vibration has a higher natural frequency than the fixed-free mode. A multiplier of 4.4 for the different modes can be estimated from data given by Harris [91]. The two modes present two extremes. The actual vibration might be somewhere in between being completely free or hinged. The actual bending mode frequency is consequently expected to be between the two extreme frequencies calculated below.

If we assume a fixed-free mode of vibration then the bending mode frequency can again be calculated from equation (A.25), which is

$$\omega = \sqrt{\frac{3YI}{l^3(m + 0.23m_\mathrm{b})}}.$$

From this equation a bending-mode spring constant (k_b) can be defined as

$$k_\mathrm{b} = \frac{3YI}{l^3}.$$

Substituting the area moment I

$$I = \frac{\pi}{64}(d_\mathrm{o}^4 - d_\mathrm{i}^4)$$

the spring constants is

$$k_\mathrm{b} = \frac{3\pi}{64} Y \frac{d_\mathrm{o}^4 - d_\mathrm{i}^4}{l^3}. \qquad \text{(A.31)}$$

Inserting the dimensions and material constants of the outer tube k_b is found to be

$$k_b = \frac{3\pi}{64} \times 7.5 \times 10^{10}\,\mathrm{N/m^2} \times \frac{(2.2^4 - 1^4)}{11.2^3} \times 10^{-3}\,\mathrm{m} = 0.176 \times 10^6\,\mathrm{N/m}.$$

The natural frequency of the bending mode can now be computed. Again, the spring constant is multiplied by three to account for the three tubes. The masses of the tubes is less than 1 g and negligible against the disk mass. The frequency of the bending mode vibration is

$$f_h = \frac{1}{2\pi}\sqrt{\frac{k_b}{m}} = \frac{1}{2\pi}\sqrt{\frac{3 \times 0.176 \times 10^6\,\mathrm{N/m}}{11 \times 10^{-3}\,\mathrm{kg}}} = 1.10\,\mathrm{kHz}. \tag{A.32}$$

If a fixed-hinged type of vibration can be assumed then a natural frequency of around

$$f_h = 4 \times 1.10\,\mathrm{kHz} = 4.41\,\mathrm{kHz} \tag{A.33}$$

is expected.

The torsional bending mode frequency of the probe head is related to the lateral bending mode frequency by [22]

$$f_r = \sqrt{2} f_h.$$

For the fixed-free mode the torsional frequency is

$$f_r = \sqrt{2} \times 1.10\,\mathrm{kHz} = 1.55\,\mathrm{kHz}, \tag{A.34}$$

for the fixed-hinged mode it is found

$$f_r = \sqrt{2} \times 4.41\,\mathrm{kHz} = 6.2\,\mathrm{kHz}. \tag{A.35}$$

Disk Bending

The disk itself can also vibrate. Its lowest natural frequency is given by [91]

$$f = \frac{1}{2\pi} B \sqrt{\frac{Y h^2}{\rho a^4 (1 - \nu^2)}},$$

where B is a tabulated coefficient listed in [91], h and a are the thickness and diameter of the probe head disk and ν is Poisson's ratio. Inserting the material constants for steel from table A.3 ($\nu = 0.28$, $\rho = 7.7 \times 10^3\,\mathrm{kg/m^3}$, $Y = 210 \times 10^9\,\mathrm{N/m^2}$) and the dimensions $h = 4\,\mathrm{mm}$ and $a = 18.4\,\mathrm{mm}$ from above, the flexing frequency of the disk is

$$f = B \times 10.2\,\mathrm{kHz}.$$

The coefficient B is given as 6.09 for a free plate and as 4.35 for a plate clamped at the center. In the case of the probe head disk the disk is clamped in three places, but the two values

$$f = 4.35 \times 10.2\,\mathrm{kHz} = 44.4\,\mathrm{kHz}$$

and

$$f = 6.09 \times 10.2\,\mathrm{kHz} = 62.1\,\mathrm{kHz}$$

give an estimate about what flexing frequency one has to expect from the disk.

Other Modes

It is also possible for the probe head to vibrate in a mode where the center tube oscillates opposite to the lateral bending of the outer tubes. This would result in a vibration with a frequency faster than the pure lateral bending mode (1.06 kHz or 4.24 kHz), but slower than the pure bending of the central actuator (7.4 kHz).

Summary The different vibrational modes of the center tube and the probe head and their natural frequencies are summarized in table A.4.

actuator	mode	condition	f (kHz)
outer	lateral	free	1.1
outer	torsional	free	1.6
mixed			1 to 7
outer	lateral	hinged	4.4
outer	torsional	hinged	6.2
center	lateral		6.8 to 7.9
outer	longitudinal		12
disk	flexing		44.4 to 62.1
center	longitudinal		44.8 to 57.9

Table A.4: **Natural Frequencies of the Probe Head:** The first column labels the type of vibration, while the second gives the calculated natural frequency. The entries are sorted with rising natural frequency.

A.5.5 Beam Vibration

The lowest natural frequency of vibration of a beam, such as the STM tip, depends on the material of the beam and its dimensions. A beam with a circular cross section consisting of only one material is considered in this derivation. In all other subsections of section A.5, the extrinsic properties spring constant (k) and mass (m) of the beam were used. In contrast, to discuss the influence of the beam dimensions length (l) and diameter (d) on the vibrational frequency, the frequency is expressed in terms of its dimensions and its intrinsic material properties, such as density (ρ) and Young's modulus (Y) in this subsection.

Firstly, transverse vibrations of the beam that is fixed on one end and free on the other end are considered. For a beam made from only one material and having a uniform cross sectional area, the frequency can be found from equation (A.25) by assuming no mass load, i.e. $m = 0$. The lowest natural frequency is then

$$\omega = \sqrt{\frac{3Y I}{l^3 0.23 m_b}}, \tag{A.36}$$

where I is the area moment of inertia of the beam section and m_b is the mass of the beam. Now the area moment

$$I = \frac{\pi}{64} d^4,$$

is inserted into equation (A.36) and the mass of the beam is substituted by the density of the beam (ρ). The beam mass is given by

$$m_b = \rho A l = \rho \frac{\pi}{4} d^2 l$$

and equation (A.36) is re-written as

$$\omega = \sqrt{\frac{3Y}{0.23l^3} \frac{1}{16} \frac{d^2}{\rho l}} = \sqrt{\frac{3}{0.23}} \frac{1}{4} \sqrt{\frac{Y}{\rho}} \frac{d}{l^2}.$$

The term $\sqrt{Y/\rho}$ is recognized as the speed of sound (c) from equation (A.16), which leaves the lowest natural transverse frequency of a beam as

$$\omega = \sqrt{\frac{3}{0.23}} \frac{c}{4} \frac{d}{l^2}. \tag{A.37}$$

The numerical term $\sqrt{3/0.23}$ is equal to 3.61 . A different treatment, also in Harris' book [91], states a slightly different number (3.52), but leads otherwise to the same expression as above.

The transverse vibration result can be compared to the longitudinal frequency for the vibration of a beam, which was derived in section A.5.1 as equation (A.17) and is repeated here as

$$\omega = 2\pi \frac{c}{4} \frac{1}{l}. \tag{A.38}$$

Both frequencies are proportional to the speed of sound of the material. The speed of sound depends on Young's modulus and the material density. The stiffer and lighter the material, the higher the speed of sound and the higher the lowest transverse and longitudinal natural frequencies.

With regards to the dimensions of the beam, the beam length is crucial for both frequencies. The longer the beam the lower both frequencies. The beam diameter influences only the transverse frequency, which is proportional to the diameter. The longitudinal frequency does not depend on the diameter.

A.6 Thermal Drift of the STM

The thermal expansion of a material of length l is given by

$$\Delta l = l\alpha \, \Delta T, \tag{A.39}$$

where Δl is the amount of length expansion, α is the thermal expansion coefficient and ΔT is the change in temperature. The material of original length l_1 at temperature T_1 expands to

$$l_2 = l_1(1 + \alpha \Delta T),$$

at temperature $T_2 = T_1 + \Delta T$. The thermal expansion coefficient α is usually given in $10^{-6}\,\mathrm{K}^{-1}$ or ppm/K and a few values for relevant materials are listed in table A.5.

material	α ppm/K
stainless Steel	11
soft Steel	12
Ni	13.4
Brass	18
W	4.3
PTFE	70 ... 120
WC	3.7
Piezo. Ceramic	3 ... 11
Macor	12.3
Graphite	7.9
- pyrolytic	-0.6 (in basal plane)
- pyrolytic	23 (across basal plane)
Al	23.1
Cu	16.5

Table A.5: **Thermal Expansion Coefficients ($T = 293\,\mathrm{K}$):** Several thermal expansion coefficients for the materials used in building the beetle and sample holder are listed as well as possible candidates for improvement. Elements are listed under their chemical abbreviation. The metals and ceramics are all similar in the range of $(10 \ldots 20) \times 10^{-6}\,\mathrm{K}^{-1}$. Plastics have a much higher expansion coefficient and only PTFE (teflon) is listed. Thermal expansion values for piezoelectric ceramics are not readily found. A value for the ceramic used in the central actuator (PZT-5A) was not given in the manufacturer's data sheet [175], instead values for other ceramics are listed with $3 \ldots 4 \times 10^{-6}\,\mathrm{K}^{-1}$. Pohl [156] cites a typical value of $7 \times 10^{-6}\,\mathrm{K}^{-1}$, while the company Physik Instrumente gives a value of $11 \times 10^{-6}\,\mathrm{K}^{-1}$ for PZT ceramic [153]. Moseley [137] lists typical values as being $5 \ldots 7 \times 10^{-6}\,\mathrm{K}^{-1}$. The value for graphite depends very much on the state of the material. Since oriented graphite is highly anisotropic, most physical properties are different across and perpendicular to the basal plane. This is also true for the thermal expansion coefficient. The anisotropic values average out for amorphous graphite. For more information see Kirk-Othmer [119]. Values are compiled from mostly from Kaye and Laby [111], but also from Kuchling [120], Carvill [43], Davis [55] and Harper [89].

To compute the overall thermal coefficient of the tip with respect to the sample, the amount of thermal expansion of the outer piezoelectric actuators, ramp, and sample holder cover is added and the amount of expansion of the center piezoelectric actuator, sample and sample stand is subtracted from that. The individual thermal expansion coefficients are listed in table A.6.

Washers are used for compensation of the gap distance. The material of the washers is most likely zinc plated soft steel, which was inferred from the facts that they are magnetic, their density of $7.26 \times 10^3\,\mathrm{kg/m^3}$ matches that of malleable iron [55] of $7.27 \times 10^3\,\mathrm{kg/m^3}$ and soft steel is recommended as washer material to distribute the exerted load of the nut [55].

The overview in table A.6 shows that the 'beetle' probe head design is well matched with regards to linear thermal expansion. The lengths of the outer and center piezoelectric actuators match to within 0.1 mm, which is equal to the assumed uncertainty. In this calculation the tip holder (length m) was included as part of the center actuator, but the tip itself (length t) not. The thermal expansion of the outer and center actuators is matched within their large uncertainty, the residual thermal drift is $(-1 \pm 28)^{\mathrm{nm}}/\mathrm{K}$.

The sample holder is not as well matched as the probe head. The actual thermal drift depends on the sample thickness (d). One or more steel washers (length s_w) are used to compensate for the different sample thickness values. Depending on the thickness, different drift values result. With the graphite sample no washer was used and we find a thermal drift of $(+34 \pm 30)\,^{\mathrm{nm}}/\mathrm{K}$ for the sample holder. The length compensation for the sample holder is in within the uncertainty of 0.1 mm. The same calculation for the very thin grating sample, where two washers were inserted, gives a thermal drift of $(+13 \pm 30)\,^{\mathrm{nm}}/\mathrm{K}$. The length compensation is $\Delta l = -0.3\,\mathrm{mm}$. The probe head was standing on top of the ramp and the tip was short.

The calculated thermal drift of the sample holder lies within the uncertainty in all three considered cases, but still varies considerably. The thermal drift should ideally be small with small uncertainty and equal in all cases.

The thermal design of the sample holder can be improved. One issue that has to be addressed is the height difference of the sample holder cover and stand and the use of aluminium for both. This difference alone causes a drift of

$$3\,\mathrm{mm} \times 23.1 \times 10^{-6}\,\mathrm{K}^{-1} = +69.3\,^{\mathrm{nm}}/\mathrm{K}. \tag{A.40}$$

While aluminium is an easy material for mechanical preparation it has an undesirable high thermal expansion coefficient. The thermal drift in equation (A.40) could be halved if steel was used instead. The main problem in the sample holder design is that the 3 mm length difference in the sample holder cover is compensated by materials of different thermal expansion coefficient. Tip, sample and washer all have a much smaller expansion coefficient than aluminium, while PTFE, a plastic, has a very high expansion coefficient. If the sample stand has to be adjusted with a washer, the thermal drift changes. The thermal drift is also hard to predict, since large values are subtracted from each other, although the uncertainty in the expansion coefficient of the actuators is also high.

For a future re-design, more reliable data for the piezoelectric actuators need to be found. Instead of PTFE as an electrical insulator, a ceramic should be used. The sample holder and stand should be made small and sturdy. Ceramics have low expansion coefficients and can therefore be considered as construction material. However, they are also brittle, which has to be taken into account.

		l (mm)	α ($10^{-6}\,\mathrm{K}^{-1}$)	$((\alpha l \pm \Delta\alpha l)\,^{\mathrm{nm}}/_{\mathrm{K}})$	
	$+d_\mathrm{s}$	+1.5	11	+17	10
	$+p_\mathrm{o}$	+13	7	+91	18
	$+b$	+1.5	11	+17	5
outer tube	$+l_\mathrm{o}$	+16.0		+125	21
	$-p_\mathrm{i}$	−12.7	7	−89	18
	$-w$	−1.2	12.3	−15	2
	$-m$	−2.0	11	−22	6
center tube	$-l_\mathrm{c}$	−15.9		−126	19
	$-t$	-3 ± 0.5	4.3	−13	1
	$-s_\mathrm{i}$	−15	23.1	−347	5
	$-d$	−1	-0.6	+1	0
	$-s_\mathrm{w}$	−1	11	−12	3
	$-g$	−0.25	100	−25	5
	$+s_\mathrm{o}$	+18	23.1	+416	6
	$+r$	$+1.2 \pm 0.2$	3.7	+4	1
sample holder	$\pm s_\mathrm{d}$	-0.1 ± 0.7		+35	10
	$\pm s_\mathrm{sw1}$	-0.1 ± 0.7		+23	10
	$\pm s_\mathrm{sw2}$	-0.3 ± 0.7		+14	10
gap	$\pm\Sigma_\mathrm{d}$	$+0.0 \pm 0.7$		+34	30
	$\pm\Sigma_\mathrm{sw1}$	$+0.0 \pm 0.7$		+22	30
	$\pm\Sigma_\mathrm{sw2}$	-0.3 ± 0.7		+13	30

Table A.6: **Thermal Expansion of the STM:** The lengths and thermal expansion coefficients of the probe head and sample holder are summed up in the list. Figure 3.16 shows the different lengths and materials used, the labels in the table are the ones used in the figure. The lengths l in the second column should add up to approximately zero. This is when the tip just touches the sample. Tip length (t) and ramp height (r) can vary as indicated to adjust the gap to a small positive gap. There are inside lengths, which are the lengths of the center piezoelectric actuator and the sample stand, and outside lengths, which are the length of the outer piezoelectric actuator and the sample holder cover. All inside lengths are counted as negative, while all outside length are counted as positive. Overall a positive total length indicates a gap between tip and sample. This gap may be up to 0.4 mm wide, the height distance of the ramp. The last columns give the thermal expansion coefficient in $^{\mathrm{nm}}/_{\mathrm{K}}$ for the materials and the estimated uncertainty. The uncertainty in the expansion coefficient contributes mainly to the overall uncertainty. For the piezoelectric ceramic a relatively large error of 20% was assumed. The uncertainty in length was assumed to be an absolute value of 0.1 mm. The list has three intermediate sums, the composite lengths of the center piezoelectric actuator, the outer actuator and the sample holder. The result for the sample holder varies with sample thickness (d), which can be adjusted by including or leaving out one or two steel washers (s_w). The piezoelectric actuators are reasonably well matched, which is a feature of this design. The sample holder is not well matched mostly due to the long lengths used for the cover and stand and the choice of aluminium as a construction material. A positive thermal expansion coefficient means that the gap between tip and sample *increases* with increasing temperature.

Appendix B

Electronics

B.0.1 Cascading Shift Registers

A single shift register 74LS597 has eight parallel data inputs $D_0, \ldots, D_7$. It is possible to handle more than eight input lines by cascading ('daisy chaining') several shift registers in the following manner:

- the parallel input data are divided into groups of eight and connected to one shift register each,
- the serial data outputs Q_S are connected to the serial data inputs D_S of the next shift register,
- all individual control signals ($\overline{\text{SHLD}}$, SHCP, STCP) are tied together to form common controls signal for all shift registers.

With every clock signal on SHCP every shift register clocks out the previously latched data to the serial data output Q_S. A shift register in the middle of the chain is getting serial data into its D_S input while it is clocking out serial data on its Q_S output. The incoming serial data are attached ('tagged') to the latched-in parallel data.

B.0.2 Pulse Stretcher

The following functional blocks are used in several places in the DSP interface board.

Pulse Stretcher

A pulse stretcher produces a synchronized pulse of variable length from a short pulse. As shown in figure B.1, the pulse stretcher is made up from a flip-flop, a counter and logic gates. The variable pulse is usually longer than the short input pulse.

Pulse Tagger

If a pulse of a minimum length is needed after a specific event is signaled, the pulse tagger circuit of figure B.2 can be used. The pulse tagger is a variation of the pulse stretcher. It exploits the fact that the pulse stretcher is edge-triggered.

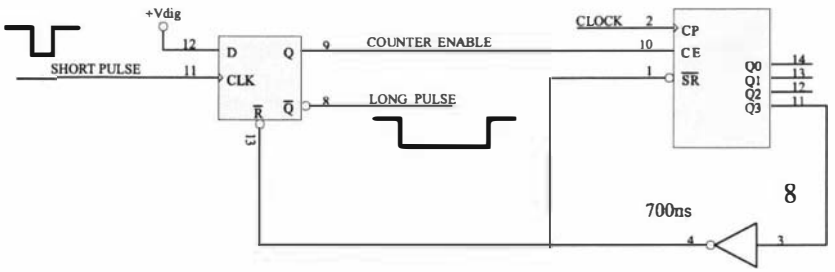


Figure B.1: **Pulse Stretcher Circuit:** The flip-flop to the left is set by the short input pulse, which in turn enables the counter to the right. A certain number of clock cycles are counted until finally the flip-flop is reset again and ready for the next input pulse. The output pulse is one of the flip-flop outputs or its derivative. Depending on which output is used, the pulse can be synchronous with the clock. If n clock cycles are counted, then the length of the output pulse is equal to $n - 1$ cycles. The diagram shows an example of counting $n = 8$ cycles.

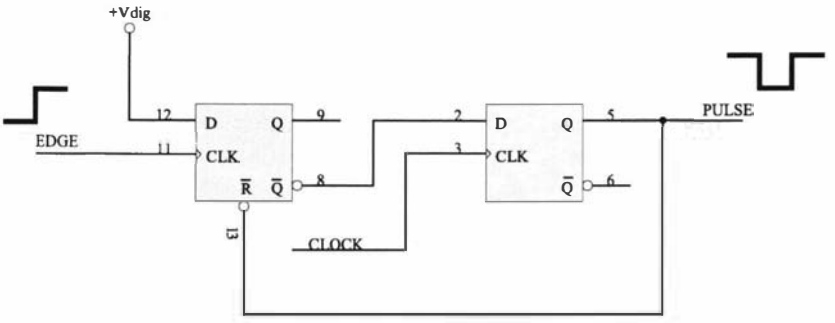


Figure B.2: **Pulse Tagger Circuit:** The pulse tagger is basically the same circuit as the pulse stretcher from figure B.1, but used for a different purpose. This is possible, because the pulse stretcher is an edge-triggered device. The shortest synchronized pulse is one clock cycle long. A simple flip-flop instead of a counter can be used in this case. The pulse taggers in this DSP interface are all made from two flip-flops. The output of the pulse tagger is the output of the second flip-flop, since the first flip-flop is reset asynchronously.

Extra-Pulse Eliminator

The special property of the eliminator is to produce only a single pulse with a minimum length, while the input receives more than one triggering edge. The circuit of the extra-pulse eliminator is the same as that for the pulse tagger shown in figure B.2. The eliminator exploits the memory property of the input flip-flop and the fact that the two flip-flops are decoupled. If the input flip-flop receives more than one rising edge it will not change its state, since the second flip-flop is set by an independent signal, the clock.

B.0.3 Communication Protocol

The communication between DSP and PC depends on the 8255A. In the 'strobed input/output' mode the 8255A has two status signals and expects strobe and acknowledgement signals from the DSP side. The parallel-data register provides these signals for the 8255A and also two status signals for the DSP.

The Write-Channel On the parallel-data register the status signal for the write-channel is one of the DSP flags, FLAG3. The status signal of the 8255A is IBF. The 8255A expects a strobe signal ($\overline{\text{STB}}$) to latch the data. The parallel-data register uses an output enable signal ($\overline{\text{OE}}_{\text{IO}}$) to drive the data lines for a sufficient time.

The communication link that the DSP uses to write data to the PC is referred to as the write-channel; the link that the DSP uses to read data from the PC is the read-channel.

Data are transferred from the DSP to the PC in the following way:

1. The DSP checks that FLAG3 is low (inactive). The DSP writes to the parallel-data register.

2. The data are latched in the write register with a combination of write and select signal of the DSP. The DSP data bus is free. Latching the data, sets FLAG3 high.

3. The output enable signal ($\overline{\text{OE}}_{\text{IO}}$) starts driving the data lines. It enables both the output of the write register and the receiving buffer on the Driver unit.

4. The strobe signal ($\overline{\text{STB}}$) starts at the same time as the output enable signal, but ends earlier. The data are received by the peripheral interface in the PC on the rising edge of the strobe. The output enable signal $\overline{\text{OE}}_{\text{IO}}$ still holds the data bus for a certain hold time.

5. The status signal on the PC side (IBF, 'input buffer full') is set by the strobe $\overline{\text{STB}}$. The PC polls the IBF signal and reads the data from the peripheral interface after IBF has been set. Reading the data in turn resets IBF.

6. The falling edge of the IBF signal resets the flag FLAG3.

7. The DSP writes the next datum, only when FLAG3 is low.

The Read-Channel The status signal for the read-channel available to the DSP is another DSP flag, FLAG2. The 8255A has the status signal denoted $\overline{\text{OBF}}$. The same signal holds the data line drivers. The 8255A expects an acknowledgement signal ($\overline{\text{ACK}}$) after the parallel-data register latched the data.

The data are transferred from the PC to the DSP in a similar fashion as above:

1. The PC checks that the status signal $\overline{\text{OBF}}$ ("output buffer full") is high (inactive). The PC then writes to the peripheral interface. The write command sets the $\overline{\text{OBF}}$ signal low (active).

2. The $\overline{\text{OBF}}$ signal enables the data line driver on the driver unit.

3. With a delay of one clock cycle (100 ns) after the $\overline{\text{OBF}}$ signal starts driving, the DSP status signal FLAG2 is set. The read register on the parallel-data register latches the data on this starting edge of FLAG2.

4. The DSP checks whether FLAG2 is set (active). In case FLAG2 is set, the DSP reads the data from the read register. The DSP enables the output of the read register with the output enable signal $\overline{\text{OE}}_{\text{P}}$, a combination of the select and read signal of the DSP.

5. Following the output enable signal $\overline{\text{OE}}_{\text{P}}$, the parallel-data register activates the acknowledgement signal $\overline{\text{ACK}}$ for the 8255A. The acknowledgement $\overline{\text{ACK}}$ sets the status signal $\overline{\text{OBF}}$ of the 8255A back to high. This in turn resets the status signal FLAG2 of the DSP.

6. The PC writes the next datum only if $\overline{\text{OBF}}$ is high.

B.0.4 Switchboard Connections

Let 1,2 and 3 denote the control signals coming from the three HV modules and let $A, B, \ldots, F$ denote the six different outer actuators electrodes. One connects:

$$\begin{array}{lll} \text{lateral} & A & = C = E = 1 \\ & B & = D = F = 3 \\ \text{circular} & A & = 1 \\ & B & = \text{GND} \\ & C & = E = 2 \\ & D & = 3 \\ & F & = -3. \end{array}$$

No switch is needed for the unchanged connections $A = 1$, $C = E = 2$ and $D = 3$. One double switch each is needed for B,C and F to switch between:

$$\begin{array}{lllll} B & = & D & \text{or} & \text{GND} \\ C & = & 1 & \text{or} & 2 \\ F & = & 3 & \text{or} & -3 \end{array}$$

Noise source	noise density	noise gain	gain density product
Johnson noise R_f	$\sqrt{4kTR_f}$	1	$\sqrt{4kTR_f}$
Johnson noise R_2	$\sqrt{4kTR_2}$	$\frac{R_f}{R_2}$	$\frac{R_f}{R_2}\sqrt{4kTR_2}$
Amplifier voltage noise	e_{na}	$(1+\frac{R_f}{R_2})$	$(1+\frac{R_f}{R_2})\,e_{na}$
Amplifier current noise	i_{na}	R_f	$R_f\,i_{na}$
Amplifier current noise	i_{na}	$(1+\frac{R_f}{R_2})$	$(1+\frac{R_f}{R_2})\,i_{na}$
Source noise	e_s	$(1+\frac{R_f}{R_2})$	$(1+\frac{R_f}{R_2})\,e_s$

Table B.1: **Operational Amplifier Noise:** The first two Johnson noise contributions are the ones from the feedback resistor (R_f) and the second resistor (R_2), which is the input resistor for the inverting amplifier and the resistor to ground for the non-inverting amplifier. Which one of the amplifier current noise is attributed to the source and which one to the feedback circuit also depends on the amplifier configuration. The interesting fact of the feedback circuit current noise is that the generated noise voltage is proportional to $R_f \parallel R_2$ and the noise gain proportional to $(R_f + R_2)/R_2$, so that the product is proportional to just R_f. The source noise density e_s is often, but not always, Johnson noise of the source resistance R_s, i.e. $e_s = \sqrt{4kTR_s}$. In the case of the current-to-voltage converter it is mainly the shot noise of the tunneling current.

The whole switching action can be done with one rotary switch, if the switch connects 12 inputs to 6 outputs for two different switch positions. The 12 inputs would bear the control signal pairs $\pm1, \pm2, \pm3$ and GND, while the 6 inputs bear the $\pm B, \pm C, \pm F$ pairs.

B.1 Probe Electronics

B.1.1 Amplifier Noise Calculation

Consider an operational amplifier in either inverting or non-inverting configuration. The total amplifier noise contribution has been calculated by multiplying the individual noise contributions of the resistors and the operational amplifier noise by the estimated noise gain. The general result is listed in table B.1. The resulting noise voltages for the modules are listed below. For all bandwidth calculations we use the corner frequency of the respective low-pass filter. Since the filters decrease sharply above the corner frequency this is a good approximation.

The high-voltage amplifier is an inverting operational amplifier with $R_f = 190\,\mathrm{k\Omega}$ and $R_{in} = 3.9\,\mathrm{k\Omega}$, giving a gain of $G_{hv} = 47$. The amplifier voltage noise density is approximated as $5\,^{nV}/\sqrt{Hz}$; the amplifier current density was calculated as being the shot noise of the input bias current as $5.1\,^{fA}/\sqrt{Hz}$. The temperature dependence of the bias current has been taken into account and an operating temperature of 65°C was assumed. The resulting amplifier noise is

$$\left(56^2 + 392^2 + 249^2\right)^{\frac{1}{2}}\,^{nV}/\sqrt{Hz} = 468\,^{nV}/\sqrt{Hz}$$

which are the first three entries in table B.1. The current noise terms are below $1\,^{nV}/\sqrt{Hz}$ and negligible. There is no resistance between the non-inverting terminal and ground, so the last list entry is also zero. In a bandwidth of

$f_c = 100\,\mathrm{kHz}$ the noise sum is

$$468\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}} \times \sqrt{f_c} = 468\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}} \times 316\,\sqrt{\mathrm{Hz}} = 148\,\mu\mathrm{V}.$$

The same high-voltage amplifier with $R_f = 62\,\mathrm{k}\Omega$ and $R_{in} = 3.9\,\mathrm{k}\Omega$ has a gain of $G_{hv} = 15.8$. The resulting amplifier noise is now smaller:

$$\left(32^2 + 127^2 + 84^2\right)^{\frac{1}{2}}\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}} = 156\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}}$$

which are the first three entries in table B.1. In a bandwidth of $f_c = 100\,\mathrm{kHz}$ this is

$$156\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}} \times \sqrt{f_c} = 156\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}} \times 316\,\sqrt{\mathrm{Hz}} = 49\,\mu\mathrm{V}.$$

The Sallen-Key low-pass filter of the HV-module is in non-inverting configuration with $R_f = 15\,\mathrm{k}\Omega$ and $R_2 = 56\,\mathrm{k}\Omega$. The gain is $G_{lp} = 1.268$. The amplifier voltage noise density is given as $3.2\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}}$ and the amplifier current density is taken as $0.6\,\mathrm{fA}/\sqrt{\mathrm{Hz}}$. The Sallen-Key type low-pass filter has two input resistors of $R_{in} = 2 \times 27\,\mathrm{k}\Omega$. The capacitors are ignored in the noise calculation. The resulting amplifier noise is

$$\left(16^2 + 8^2 + 9^2 + 4^2 + 41^2 + 38^2\right)^{\frac{1}{2}}\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}} = 60\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}}$$

following the order of table B.1. In the same bandwidth of $f_c = 100\,\mathrm{kHz}$ as above this is

$$60\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}} \times 316\,\sqrt{\mathrm{Hz}} = 19\,\mu\mathrm{V}.$$

In fact the two biggest terms of noise density are generated by the input resistances. Since they only serve to build an *RC*-low-pass filter at the operational amplifier input, one should be able to reduce the noise by choosing smaller resistances and bigger capacitances. The stage ahead of the low-pass filter has to supply more current in that case, which should be no problem if the resistance is kept above $1\,\mathrm{k}\Omega$. A greater error in the corner frequency would also result if a suitable capacitance value cannot be found.

The Sallen-Key low-pass filter of the ADC module is a two-stage low-pass filter as opposed to the one in the HV module. The noise calculation is done in two steps. The first stage has $R_f = 4.7\,\mathrm{k}\Omega$ and $R_2 = 56\,\mathrm{k}\Omega$. The gain is $G_{lp} = 1.084$. The input resistors are $R_{in} = 2 \times 47\,\mathrm{k}\Omega$. Amplifier voltage noise and current density are the ones for the OP-270GP as above ($3.2\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}}$ and $0.6\,\mathrm{fA}/\sqrt{\mathrm{Hz}}$). The resulting amplifier noise is

$$\left(9^2 + 3^2 + 3^2 + 3^2 + 61^2 + 43^2\right)^{\frac{1}{2}}\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}} = 75\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}},$$

where again the noise components are ordered according to table B.1. The bandwidth of the ADC module is $f_c = 50\,\mathrm{kHz}$, the root of which is

$$\sqrt{50\,\mathrm{kHz}} = 223\,\sqrt{\mathrm{Hz}}.$$

We calculate the amplifier noise as

$$75\,^{\mathrm{nV}}/\sqrt{\mathrm{Hz}} \times 223\,\sqrt{\mathrm{Hz}} = 17\,\mu\mathrm{V}.$$

The two biggest terms of noise density are again generated by the input resistances.

The second stage has $R_f = 12\,\mathrm{k\Omega}$ and $R_2 = 15\,\mathrm{k\Omega}$ with a gain of $G_{lp2} = 1.8$. The input resistors are $R_{in} = 2 \times 10\,\mathrm{k\Omega}$. The second part of the dual operational amplifier OP-270GP is used in the second stage, having the same noise characteristics. The noise density is calculated as

$$\left(14^2 + 13^2 + 7^2 + 6^2 + 22^2 + 33^2\right)^{\frac{1}{2}}\,\mathrm{nV}/\sqrt{\mathrm{Hz}} = 45\,\mathrm{nV}/\sqrt{\mathrm{Hz}},$$

where again the noise components are ordered according to table B.1. With equal bandwidth of $f_c = 50\,\mathrm{kHz}$, we calculate the amplifier noise as

$$45\,\mathrm{nV}/\sqrt{\mathrm{Hz}} \times 223\,\sqrt{\mathrm{Hz}} = 10\,\mu\mathrm{V}.$$

Repeating the above calculation for the two-stage Sallen-Key low-pass filter of the bias module we get

$$\left(9^2 + 3^2 + 3^2 + 3^2 + 42^2 + 36^2\right)^{\frac{1}{2}}\,\mathrm{nV}/\sqrt{\mathrm{Hz}} = 57\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$$

for the first stage, and

$$\left(14^2 + 13^2 + 7^2 + 6^2 + 216^2 + 104^2\right)^{\frac{1}{2}}\,\mathrm{nV}/\sqrt{\mathrm{Hz}} = 241\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$$

for the second stage. The two designs only differ in input resistance and bandwidth. The input resistance of the first stage is $R_{in} = 2 \times 33\,\mathrm{k\Omega}$, of the second stage $R_{in} = 2 \times 100\,\mathrm{k\Omega}$. The root of the bandwidth is

$$\sqrt{10\,\mathrm{kHz}} = 100\,\sqrt{\mathrm{Hz}}.$$

The resulting noise is

$$57\,\mathrm{nV}/\sqrt{\mathrm{Hz}} \times 100\,\sqrt{\mathrm{Hz}} = 6\,\mu\mathrm{V}$$

for the first stage and

$$241\,\mathrm{nV}/\sqrt{\mathrm{Hz}} \times 100\,\sqrt{\mathrm{Hz}} = 24\,\mu\mathrm{V}$$

for the second stage.

The noise contribution of the current-to-voltage converter of the AD1862 digital-to-analog converter in the HV module is calculated with the source resistance of $R_s = 2.1\,\mathrm{k\Omega}$ given in the data sheet [8]. The feedback resistance was $R_f = 9.5\,\mathrm{k\Omega}$ and the same amplifier specifications as for the low-pass filter above were assumed. The calculation results in

$$\left(12^2 + 27^2 + 6^2 + 18^2 + 7^2 + 33^2\right)^{\frac{1}{2}}\,\mathrm{nV}/\sqrt{\mathrm{Hz}} = 48\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$$

In the same bandwidth of $f_c = 100\,\mathrm{kHz}$ as above this is

$$48\,\mathrm{nV}/\sqrt{\mathrm{Hz}} \times 316\,\sqrt{\mathrm{Hz}} = 15\,\mu\mathrm{V}.$$

Using only the on-chip resistance of $R_f = 3\,\mathrm{k\Omega}$ and the specifications of the NE5534, which are $3.5\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ (voltage noise density) and $0.6\,\mathrm{fA}/\sqrt{\mathrm{Hz}}$ (amplifier current density), we get

$$\left(7^2 + 8^2 + 2^2 + 8^2 + 3^2 + 14^2\right)^{\frac{1}{2}}\,\mathrm{nV}/\sqrt{\mathrm{Hz}} = 20\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$$

In the same bandwidth of $f_c = 100\,\text{kHz}$ as above this is

$$20\,\text{nV}/\sqrt{\text{Hz}} \times 316\,\sqrt{\text{Hz}} = 6\,\mu\text{V}.$$

The noise contribution of the current-to-voltage converter of the PM-7543 digital-to-analog converter used in the bias module is calculated with a source resistance approximated as $10\,\text{k}\Omega$. The feedback resistance is the on-chip resistance of the same value $R_f = 10\,\text{k}\Omega$. The amplifier specifications of the OP-270 are used. The calculation gives

$$\left[13^2 + 13^2 + 6^2 + 6^2\right]^{\frac{1}{2}}\,\text{nV}/\sqrt{\text{Hz}} = 21\,\text{nV}/\sqrt{\text{Hz}},$$

where the noise contributions are in order of table B.1 above. The last two entries are zero. In a bandwidth of $f_c = 10\,\text{kHz}$ for the bias module this is

$$21\,\text{nV}/\sqrt{\text{Hz}} \times 100\,\sqrt{\text{Hz}} = 2\,\mu\text{V}.$$

The total noise contribution of the ADC-module input buffer is less than $2\,\mu\text{V}$. The individual values are not further listed.

B.1.2 Noise Bandwidth

If white noise is low-pass filtered with a simple RC-filter, then the noise density does not decrease rapidly. To take the gentle roll-off into account, the noise can be integrated in the following way. The noise density (e_n) is written as

$$e_n = \frac{e_{n0}}{\sqrt{1 + (\frac{f}{f_0})^2}}, \tag{B.1}$$

where e_{n0} is the noise density at low frequencies and f_0 is the cut-off frequency of the low-pass filter. Integrating the square of the noise density between f_1 and f_2 yields

$$\begin{aligned} V_n^2 &= \int_{f_1}^{f_2} e_n^2\,\mathrm{d}f \\ &= e_{n0}^2 \int_{f_1}^{f_2} \frac{1}{1 + (\frac{f}{f_0})^2}\,\mathrm{d}f \\ &= e_{n0}^2 \left[f_0 \tan^{-1}\frac{f}{f_0} \right]_{f_1}^{f_2} \\ &= e_{n0}^2 f_0 \left(\tan^{-1}\frac{f_2}{f_0} - \tan^{-1}\frac{f_1}{f_0} \right). \end{aligned}$$

If the integration is taken to infinity then the noise voltage is

$$V_n = e_{n0}\sqrt{f_0(\frac{\pi}{2} - \tan^{-1}\frac{f_1}{f_0})}. \tag{B.2}$$

If the complete bandwidth from zero to infinity is taken, then the noise voltage is

$$V_n = e_{n0}\sqrt{f_0\frac{\pi}{2}}. \tag{B.3}$$

This differs from a 'brick-wall' type of low-pass filter by an extra factor of $\sqrt{\pi/2} = 1.25$. Roughly speaking, if the gentle roll-off is not taken into account, the noise is underestimated by a quarter of its value.

B.1.3 Step Rise Time

Frequency and time behavior are linked in that the bandwidth of an operational amplifier is related to its time constant τ. In the simplest case bandwidth and time constant are related like

$$\tau = \frac{1}{\omega} \tag{B.4}$$

as it is for a simple RC low-pass filter. The bandwidth of the operational amplifier might indeed be an RC low-pass filter, which could be the feedback impedance. Or the operational amplifier is internally compensated, which means that an internal RC low-pass filter is incorporated in the circuit. In general the behavior of operational amplifiers can also be more complicated, but equation (B.4) works well as an estimate. The equation allows us to estimate how much time it takes an amplifier to reach a certain fraction of its final value after a step change has been applied to the input. Often the time it takes to reach 90% is referred to as the step-rise time. For a RC low-pass filter this time is $\ln 10\,\tau$. Since there is usually more than one bandwidth limited element in the analog signal processing circuits, the fact has to be taken into account that for a second stage the input does not change abruptly anymore but smoothly. How the output of a RC low-pass filter looks like when driven by another, buffered, RC low-pass filter can be seen in figure B.3.

The graph is the result of solving the differential equation for the second stage, which gives

$$V_{c2} = V_0(1 - (1 + \frac{t}{\tau})e^{-\frac{t}{\tau}}),$$

where V_{c2} is the output voltage of the second stage measured across the capacitor, V_0 is the input step voltage and τ is the time constant of both circuits. From the graph we see that the time constant of two cascaded RC low-pass filters is 2.16τ, i.e. more than 2 times the individual time constants, which is not surprising for a voltage change has to travel through the first stage first to reach the second.

Table B.2 shows that the step rise time of one or two bandwidth limited stages already takes up a sizeable fraction of the $10\,\mu$s long sampling interval. This is hardly avoidable since smoothing and anti-alias filtering is highly desirable and the low-pass filters have been put into the signal paths with the highest possible bandwidth.

The step rise time for both a 2-pole and a 4-pole Bessel filter have also been listed in table B.2. A Bessel filter is not simply equal to cascading a number or simple RC filters. The cut-off frequency of all stages of the Bessel filter is higher than the desired cut-off of the overall filter. The rise time of these filters is therefore not as long as the one for several cascaded RC filters with a cut-off frequency at the desired bandwidth. Since the difference between a 2-pole and 4-pole Bessel filter is not large, the response time would not be greatly reduced by reducing the filter order. Whether or not more than one stage should be bandwidth limited is a question of trading off noise reduction against speed.

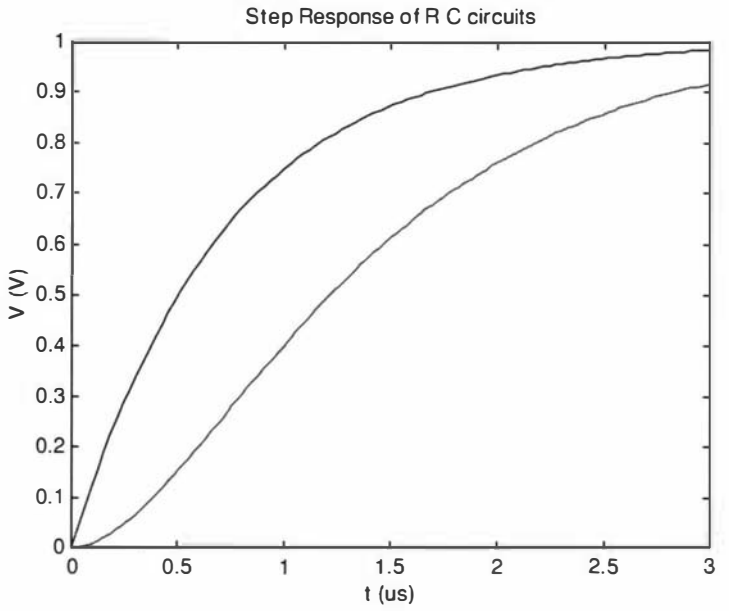


Figure B.3: **Step Response of two RC low-pass filters:** The step response of two cascaded RC low-pass filters was calculated and the result shown in the figure. The input to the first RC low-pass filter was a voltage step of $1\,\mathrm{V}$ at $t = 0$. The time constant of both RC low-pass filters was $\tau = 0.72\,\mu\mathrm{s}$. This was chosen to give a settling time of $5\,\mu\mathrm{s}$ (i.e. to reach 99.9% of its final value). The output of the first stage is the faster rising one and is also the input of the second stage. The second stage needs slightly more than $2\ln 10\,\tau$ to reach the same output voltage level as the first one. From the graph the time taken was read as $1.554\,\mu\mathrm{s} = 2.16\,\tau$. The step rise time (0 to 90%) for the two graphs are read as $1.67\,\mu\mathrm{s}$ and $2.82\,\mu\mathrm{s}$. The solution of the differential equation was $V_{c2} = V_0(1 - (1 + t/\tau)\exp(-t/\tau))$.

module	f	τ	t_{sr1}	t_{B2}	t_{B4}	t_{sr2}	t_{sr3}
	(kHz)	all (μs)					
HV	100	1.59	3.66	**4**	5	7.92	12.1
ADC	50	3.18	7.32	8	**10**	15.8	24.1
Bias	10	15.9	36.6	40	**50**	79.2	121

Table B.2: **Estimate and Comparison of Time Constants:** Listed are the step rise times calculated from the time constants τ, 2.16τ and 3.3τ at different bandwidths used in the probe electronics. The step rise times are obtained by multiplying by $\ln 10$. They are labeled t_{sr1} through to t_{sr3} in the table. The step rise time of a 2-pole and 4-pole Bessel filter according to the design table in [100], p.273, is also given as a comparison. These are t_{B2} and t_{B4}. Numbers in bold are of the actually implemented filters. The time constants indicate the time it takes a step change to travel through one, two or three cascaded bandwidth limited stages. The time behavior of the stages is approximated as simple RC circuits. The second and third stage have longer time constants, because the voltage change has to pass the previous stages first. The factor 2.16 is a result of solving the differential equation for two RC low-pass filters, the third factor 3.3 is calculated from the assumption that the third stage will retard the voltage change as much as the second stage, i.e. $1.16 \times 2.16 = 3.3$. It is probably a slight underestimate.

When taking characteristic V-I curves the relatively slow step rise of the bias module has to be taken into account.

Ramp Response The time lag of a low-pass filter is of interest when approaching the tip towards the sample. Coming closer to the sample at a constant speed results in an exponentially rising voltage at the current-to-voltage converter output. This is approximated as a voltage ramp starting at $t = 0$, which is

$$V_r = wt.$$

The response of a simple RC low-pass filter to this ramp is found by solving the differential equation

$$\dot{V}_c = \frac{1}{RC}(V_r - V_c).$$

The output voltage of the low-pass filter is the voltage across the capacitor V_c. The solution is

$$V_c = w(\tau e^{-\frac{t}{\tau}} + t - \tau),$$

where the time constant τ is as usual $\tau = RC$.

After the exponential has died down, that is after $t \gg \tau$ the low-pass filter output follows the ramp with a time lag of τ. This result is used to estimate how much time the approach mechanism needs to detect the approaching surface.

B.1.4 Choosing Optimum Gain

In all three types of analog electronic modules, the ADC module, bias module and high-voltage module, there is more than one gain stage between source and digitized signal. To a certain degree one has a choice how much gain the individual stages should have. A guideline of how to distribute the gain is the overall signal-to-noise ratio of the signal in question. It is generally true that for any chain of amplification stages the best signal-to-noise ratio can be achieved for chains that have the highest amplification in the earliest stages. This is because the later stages amplify the noise of the earlier stages as can be seen below.

Let us consider the high-voltage module, where the right choice of gain distribution is especially important. The high-voltage (HV) module has three gain stages after the digital-to-analog converter, a current-to-voltage converter, a low-pass filter and a high-voltage amplifier. The digital-to-analog converter puts out a current (I_{dac}), which gets amplified to

$$V_{\mathrm{out}} = G_{\mathrm{hv}} G_{\mathrm{lp}} R_{\mathrm{tr}} I_{\mathrm{dac}},$$

where G_{hv} is the gain of the high-voltage amplifier, G_{lp} the gain of the low-pass filter and R_{tr} the trans-resistance of the current-to-voltage converter. Nothing can change the fact that the digital-to-analog converter noise (I_{nd}) gets amplified by the same gain factor as the signal

$$V_{\mathrm{ndo}} = G_{\mathrm{hv}} G_{\mathrm{lp}} R_{\mathrm{dac}} I_{\mathrm{nd}},$$

where V_{ndo} is the resulting output noise. The other noise voltages at the high-voltage output are

$$V_{\mathrm{nto}} = G_{\mathrm{hv}} G_{\mathrm{lp}} V_{\mathrm{nt}},$$

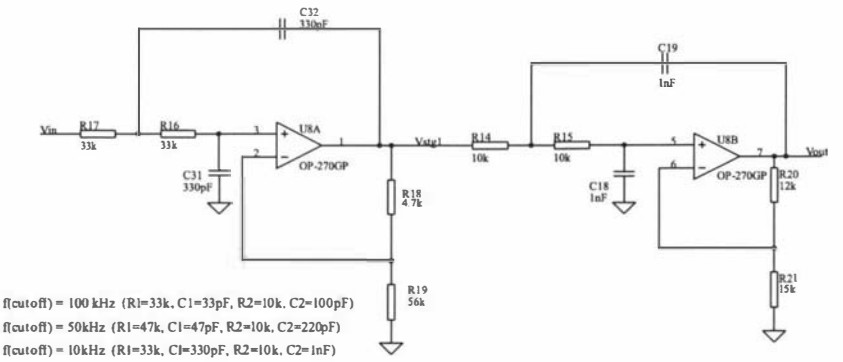


Figure B.4: **Sallen-Key Bessel Low-Pass Filter:** The filter is taken from the Bias module and has a cut-off frequency at 10 kHz. Other values are listed to the bottom left. For the operational amplifier the dual amplifier OP270GP from Analog Devices was chosen.

with V_{nto} being the output noise due to the input noise V_{nt} of the low-pass filter, and

$$V_{\mathrm{nho}} = G_{\mathrm{hv}} V_{\mathrm{nl}},$$

where V_{nho} is the output noise voltages due to the input noise V_{nl} of the high-voltage amplifier. The output noise can be found from the individual noise contributions and can be compared to the signal to find the relative contributions from the different gain stages. The individual noise contributions add with the sum of squares like

$$\left[\left(\frac{I_{\mathrm{nd}}}{I_{\mathrm{dac}}}\right)^2 + \left(\frac{V_{\mathrm{nd}}}{R_{\mathrm{dac}} I_{\mathrm{dac}}}\right)^2 + \left(\frac{V_{\mathrm{nl}}}{G_{\mathrm{lp}} R_{\mathrm{dac}} I_{\mathrm{dac}}}\right)^2\right]^{\frac{1}{2}} \tag{B.5}$$

The relative output noise stemming from the original signal is of course still the same as the relative original noise. Unavoidably, the other gain stages add extra contributions. However, as can be seen from equation (B.5), the relative noise contributions from the later stages become smaller if more gain is assigned to the early stages of amplification.

B.1.5 Sallen-Key Low-Pass Filters

There are low-pass filters in each probe electronics module. All of them have been design from the filter design table in the book of Horowitz and Hill [101]. The disadvantage of this filter type is, that the cut-off frequency is highly sensitive to the actual component values. However, this is not too critical in this application. The cut-off frequency does not have to be exact and components can be hand-picked. Finding four significant figures in the table suggests that precision resistors be used. For all filters the OP-270GP from Analog Devices has been chosen as a dual, low-noise, precision amplifier with sufficient bandwidth [6]. A circuit diagram is shown in figure B.4.

Component Values for Sallen-Key Filters

The operational amplifier stages have to be designed to have the correct gain K and corner frequency f_{c}, where both values are listed in the design table. The

filter	R_1	R_2	K	$1+\frac{R_1}{R_2}$	E
4-1	12 kΩ	15 kΩ	1.800	1.759	+2.3%
4-2	4.7 kΩ	56 kΩ	1.084	1.084	-0.01%
2	15 kΩ	56 kΩ	1.268	1.268	-0.01%

Table B.3: **Gain Values for Sallen-Key Low-Pass Filter:** The resistance values satisfying the filter-design table for the Sallen-Key filters in the book of Horowitz and Hill [101] are presented. The first two columns are the actually used resistance values, the next columns are the desired gain and the calculated gain from equation (B.6).

desired gains for the different probe electronics modules are listed in table B.3, the frequencies are listed in table B.4. To find a specific gain K for a non-inverting amplifier one has to find two resistances with a ratio of

$$K = 1 + \frac{R_1}{R_2} \Rightarrow R_1 = (K-1)R_2\,. \tag{B.6}$$

The task is now to find standard-value resistances that fulfill equation (B.6). The conventional way of finding the resistances is to start with a trial value for say R_2, calculate R_1 from equation (B.6) and see whether this value is close enough to a standard value within a certain tolerable gain error. A second criterion is whether the found resistance value lies within a band of acceptable resistance values of 1 kΩ ... 100 kΩ. Too small resistances would put to high a demand on the current driving capabilities of the previous stage, while too high resistances would generate too much extra noise in the circuit if they are directly in the signal path. The above given limits of 1 kΩ and 100 kΩ are conventionally given as suitable [100]. Another demand that the chosen resistors have to adhere to, is the amount of heat they can safely dissipate. Resistors come with certain power ratings. The standard resistors usually have a rating of 0.25 W, others are available with ratings of 0.5 W and 1 W. Beyond a power rating of 1 W the resistors become very bulky. The power rating is only an issue for large currents and large voltages. With the power rating ($P_{\max}$) and the voltage across the resistor (V) one can calculate the minimum resistance value that can stand a heat dissipation up to that level. The resistance is

$$R > \frac{V^2}{P_{\max}} \tag{B.7}$$

For this project a small program has been written that contains all standard resistance values. The program calculates all possible resistance ratios that come close to the desired value of $(K-1)$ from equation (B.6) and offers a selection of the best fitting combinations. The program can be confined to use only resistance values from within a certain band of values, which can be the one mentioned above, although it is adjustable. Resistance values for the desired gains K that were found to be good are listed in table B.3.

The first-order errors for the low-pass amplifier due to bias current and offset voltage are

$$V_{\mathrm{lp,os}} = (1+\frac{R_{19}}{R_{56}})V_{\mathrm{os}} + R_{19}I_{\mathrm{b1}} - (R_{15}+R_{16})I_{\mathrm{b2}}.$$

f_c	R	C	f_t	E
10 kHz	33 kΩ	330 pF	14.32 kHz	+2%
10 kHz	10 kΩ	1 nF	16.06 kHz	−1%
50 kHz	47 kΩ	47 pF	71.60 kHz	+0.6%
50 kHz	10 kΩ	220 pF	80.30 kHz	−9.9%
100 kHz	27 kΩ	47 pF	127.2 kHz	−1.4%

Table B.4: **Frequency Values for Sallen-Key Low-Pass Filter:** The resistance and capacitance values to find the right target frequency according to the filter-design table in Horowitz and Hill's book [100] are listed. The cut-off frequencies (f_c) of the low-pass filters in the different modules are listed in the first column. From the filter design table a target frequency is calculated, listed in column f_t. The second and third columns are the R and C values that were used to give the stage a roll-off at the target frequency.

The term due to the different input currents is 2 mV. The offset voltage term is $1.3 \times 250\,\mu\text{V} = 330\,\mu\text{V}$. The relative gain error is given by

$$\frac{\Delta G}{G} = \frac{R_1}{R_1 + R_2} \times \sqrt{2} \times \frac{\Delta R}{R},$$

assuming equal resistor tolerances. For metal film resistors, this evaluates to $0.21 \times \sqrt{2} \times 0.01 = 3.0 \times 10^{-3}$, i.e. 0.30% for the values used in the HV module.

A similar procedure is followed when finding a specific corner frequency given by the product of a resistance and a capacitance as in

$$f_c = \frac{1}{2\pi RC}.$$

Since there is a smaller number of standard capacitance values, usually one starts with a trial capacitance value and checks whether the resistance value calculated from

$$R = \frac{1}{2\pi f_c C} \tag{B.8}$$

fulfills the criteria mentioned above. The same program as above contains all standard capacitance values and calculates fitting corner frequencies. Again, both the capacitance values as well as the resistance values can be restricted to a certain band. So far the standard values stored in the program are the ones for the low-precision carbon resistors. With 5% precision resistors the error in K is

$$\Delta K = \frac{R_1}{R_2}\sqrt{2} \times 0.05.$$

Using the program is still worthwhile, since it is faster and automatically handles the value constraints. The program can also be extended to be used with higher precision metal film resistances if need be. The corner frequency values that were calculated in this project are listed in table B.4. With one exception the found resistance and capacitance values calculate to a frequencies that match the target frequency very well; the percentage error is below 2% for most values. The one pair of values with greater deviation of almost 10% was accepted, because of the small resistance value. The alternative of $R = 56\,\text{k}\Omega$ and $C = 33\,\text{pF}$ would generate more Johnson noise and would still be −6.7% below the target

frequency. The relative error in corner frequency due to the uncertainty in component values is

$$\frac{\Delta f_c}{f_c} = \left((\frac{\Delta R}{R})^2 + (\frac{\Delta C}{C})^2 \right)^{\frac{1}{2}},$$

where the error in resistance is 5% or 1% depending on the type of resistor used. The error in capacitance is strongly dependent on the specific type of capacitor used and has to be checked with the specifications of the manufacturer.

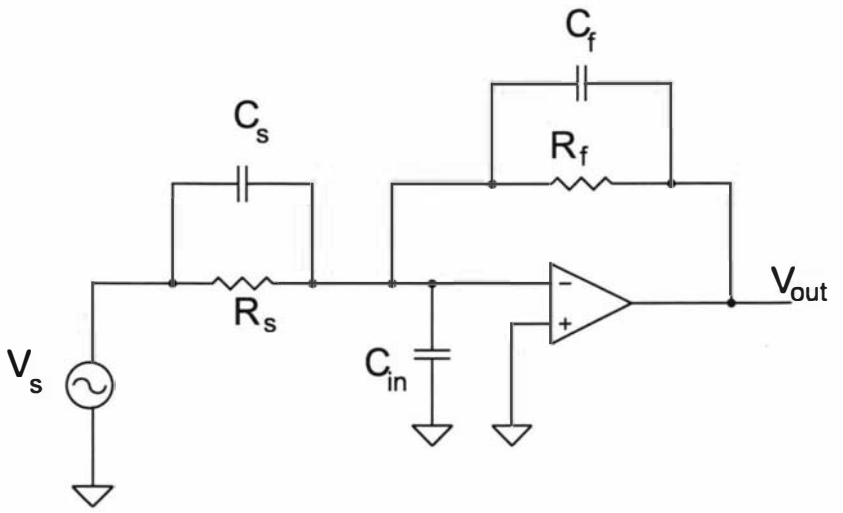


Figure B.5: **Inverting Operational Amplifier:** An operational amplifier in inverting configuration is shown, including input capacitance ($C_{\rm in}$) and feedback resistor ($R_{\rm f}$) in parallel with a feedback capacitor ($C_{\rm s}$). The voltage source ($V_{\rm s}$) has a source resistance ($R_{\rm s}$) in parallel with a capacitor ($R_{\rm s}$).

B.2 Inverting Operational Amplifier

B.2.1 Voltage Signal Gain

For any operational amplifier in inverting amplifier configuration, the inverting terminal depends on the source voltage ($V_{\rm s}$) and the output voltage of the amplifier ($V_{\rm out}$). A circuit diagram is shown in figure B.5. This can be expressed as

$$V_- = aV_{\rm out} + bV_{\rm s}\,, \tag{B.9}$$

where a and b depend on the source impedance ($Z_{\rm s}$), feedback impedance ($Z_{\rm f}$) and the input capacitance ($C_{\rm in}$) in the most general case. The terms a and b are fractions of impedances given by the voltage divider equation.

The operational amplifier governing equation is

$$V_{\rm out} = A(V_+ - V_-).$$

Substituting for V_- from equation (B.9) and setting $V_+ = 0$ gives the output voltage

$$V_{\rm out} = -\frac{A}{1+Aa}\,bV_{\rm s}, \tag{B.10}$$

where it can be seen that the fraction of the output voltage feedback, i.e. a, is affected by the amplifier gain, while the fraction b is not.

For an ideal amplifier, $A \to \infty$ and equation (B.10) reduces to

$$V_{\rm out} = -\frac{b}{a}V_{\rm s} \tag{B.11}$$

The inverting amplifier shown in figure B.5 includes the three parasitic capacitances of the source, input and the feedback resistor.

According to equation (B.9), the output voltage (V_{out}) can be calculated by determining the voltage at the inverting input of the operational amplifier, which is

$$\begin{aligned} V_- &= \frac{Z_s \parallel C_{in}}{Z_s \parallel C_{in} + Z_f} V_{out} + \frac{Z_f \parallel C_{in}}{Z_f \parallel C_{in} + Z_s} V_s \\ &= \frac{Z_s}{Z_s + (1 + j\omega C_{in} Z_s) Z_f} V_{out} + \frac{Z_f}{Z_f + (1 + j\omega C_{in} Z_f) Z_s} V_s, \end{aligned} \quad \text{(B.12)}$$

where Z_s and Z_f are the source and feedback impedances, C_{in} is the input capacitance, V_s is the source voltage, ω is the angular frequency of the source voltage and j the imaginary unit.

The fractions a and b are found to be

$$a = \frac{Z_s}{Z_s + Z_f + j\omega C_{in} Z_s Z_f}, \quad \text{(B.13)}$$

$$b = \frac{Z_f}{Z_s + Z_f + j\omega C_{in} Z_s Z_f}. \quad \text{(B.14)}$$

Although the input capacitance influences both fractions a and b, giving them frequency characteristics of low-pass filters, the overall effect cancels in the fraction b/a. This is because the denominators of both fractions a and b are equal and cancel in equation (B.11) above so that V_{out} is given by

$$V_{out} = -\frac{b}{a} V_s = -\frac{Z_f}{Z_s} V_s. \quad \text{(B.15)}$$

While this is the case for a high open-loop gain A, C_{in} does play a role in determining the frequency response of the amplifier, if A is not high enough.

B.2.2 Real Amplifier

Signal Gain

The open-loop gain A of a real operational amplifier is modeled with a single roll-off corner frequency as

$$A = \frac{A_0}{1 + j\omega \frac{A_0}{\omega_u}}, \quad \text{(B.16)}$$

where A_0 is the dc open-loop gain and ω_u is the unity-gain frequency. For the corner frequency of the open-loop gain the approximation

$$\omega_c = \frac{\omega_u}{\sqrt{A_0^2 - 1}} \approx \frac{\omega_u}{A_0}$$

was used. Substituting this result in equation (B.10)) gives

$$V_{out} = -\frac{1}{a + j\frac{\omega}{\omega_u}} b V_s. \quad \text{(B.17)}$$

In this equation the same approximation $A_0 \gg 1$ was used. Hence the influence of the operational amplifier is the addition of an extra corner frequency at $\omega = \Re(a)\omega_u$, where $\Re(a)$ is the real part of a. With an *ideal* operational amplifier the frequency response is only determined by the source and feedback impedances.

The fractions a and b can be inserted into equation (B.17) to obtain the general result in terms of the various impedances

$$G = -\frac{V_{out}}{V_s} = \frac{1}{(1 + j\frac{\omega}{\omega_u})\frac{Z_s}{Z_f} + j\frac{\omega}{\omega_u} - \frac{\omega^2}{\omega_u}C_{in}Z_s}. \tag{B.18}$$

For an ideal operational amplifier with $\omega_u \to \infty$ the equation correctly reduces to $G = Z_f/Z_s$.

With a non-ideal operational amplifier it can be seen that the frequency dependence of the circuit gain G is influenced by the unity-gain bandwidth. Especially interesting is the real part of the $\omega^2 C_{in} Z_s/\omega_u$ term, which causes a $-40\,^{dB}/_{dec}$ roll-off at high frequencies. This term shows the interaction of the input capacitance (C_{in}) and the transfer function of the operational amplifier. With ideal operational amplifiers the input capacitance may be disregarded, but with a real amplifier a high input capacitance causes an early gain roll-off, which is undesirable. A high input capacitance is also undesirable, because it increases the noise gain of the amplifier, therefore one important implication of equation (B.18) is that the input capacitance should be as small as possible. If this is not feasible then the influence of the input capacitance can be diminished by increasing the unity-gain bandwidth (ω_u) by choosing an amplifier with broader bandwidth and higher gain or by employing two amplification stages.

To analyze the amplifier circuit in figure B.5 in more detail the impedance

$$Z_R \parallel Z_C = \frac{Z_R Z_C}{Z_R + Z_C} = \frac{R}{1 + j\omega RC}$$

for a resistor in parallel with a capacitor is used for both the source and the feedback impedance. This can be inserted into equation (B.18), which after some algebra reads

$$G = \frac{1 + j\omega R_s C_s}{\frac{R_s}{R_f} - \frac{\omega^2}{\omega_u}R_s C_{tot} + j\frac{\omega}{\omega_u}(1 + R_s C_f \omega_u + \frac{R_s}{R_f})}, \tag{B.19}$$

where C_{tot} is the total capacitance defined as $C_{tot} = C_s + C_{in} + C_f$. The result is an expression that is similar to the equation of a damped driven oscillator. The ratio of the gain (G) of the amplifier to the dc gain (G_{dc}) is

$$\frac{G}{G_{dc}}\frac{1}{1 + j\omega R_t C_s} = \frac{\omega_0^2}{\omega_0^2 - \omega^2 + j\omega 2\delta}, \tag{B.20}$$

where ω_0 and δ are defined as

$$\omega_0^2 = \frac{\omega_u}{R_f C_{tot}} \tag{B.21}$$

and

$$2\delta = \frac{1}{(R_s \parallel R_f)C_{tot}} + \omega_u \frac{C_f}{C_{tot}}. \tag{B.22}$$

The only difference between this expression and that of a resonance circuit is an extra factor on the left hand side containing $R_t C_s$, which is related to the corner frequency of the source impedance. The second order term in equation (B.20) is used in other sections and conveniently abbreviated as P by

$$P = \frac{\omega_0^2}{\omega_0^2 - \omega^2 + j\omega 2\delta}. \tag{B.23}$$

B.2.3 Noise Gain

To calculate the noise gain of an operational amplifier in inverting configuration equation (B.12) is used, which was written in short as

$$V_- = aV_{\text{out}} + bV_{\text{s}}.$$

Including the input voltage noise (V_{na}) between ground and non-inverting terminal, the operational amplifier action is given as

$$V_{\text{out}} = A(V_{\text{na}} - V_-)$$

and the output voltage due to noise is

$$V_{\text{out}} = \frac{A}{1 + aA} V_{\text{na}}. \tag{B.24}$$

The noise gain of the amplifier is

$$G_{\text{n}} = \frac{V_{\text{out}}}{V_{\text{na}}} = \frac{A}{1 + aA}. \tag{B.25}$$

For high open-loop gain A the noise gain reduces to

$$G_{\text{n}} = \frac{A}{1 + aA} \overset{A \to \infty}{\longrightarrow} \frac{1}{a}. \tag{B.26}$$

To find the ideal noise gain the fraction a has to be inserted from equation (B.13).

Ideal Noise Gain

The fraction a was given in equation (B.13) as

$$a = \frac{Z_{\text{s}}}{Z_{\text{s}} + Z_{\text{f}} + j\omega C_{\text{in}} Z_{\text{s}} Z_{\text{f}}}.$$

When calculating the signal gain, the denominator would cancel with the denominator of the fraction b. In contrast the denominator of a is important when calculating the noise gain, since the denominator strongly influences the frequency response of the noise gain. Inserting the parallel resistor and capacitor impedances as above, gives the ideal noise gain as

$$G_{\text{n}} = \frac{1}{a} = G_{\text{n0}} \frac{1 + j\omega(R_{\text{t}} \parallel R_{\text{f}}) C_{\text{tot}}}{1 + j\omega R_{\text{f}} C_{\text{f}}} \tag{B.27}$$

where

$$G_{\text{n0}} = \frac{R_{\text{f}} + R_{\text{s}}}{R_{\text{s}}} \tag{B.28}$$

is the noise gain at zero frequency.

Real Noise Gain

Inserting a from equation (B.13) and A from equation (B.16) into equation (B.25) for the noise gain gives G_{n} as

$$G_{\text{n}} = \frac{\frac{Z_{\text{s}}}{Z_{\text{f}}} + 1 + j\omega C_{\text{in}} Z_{\text{s}}}{\frac{Z_{\text{s}}}{Z_{\text{f}}} - \frac{\omega^2}{\omega_{\text{u}}} C_{\text{in}} Z_{\text{s}} + j\frac{\omega}{\omega_{\text{u}}}\left(\frac{Z_{\text{s}}}{Z_{\text{f}}} + 1\right)}$$

Inserting the parallel resistor and capacitor impedances as above, gives the noise gain as

$$G_n = G_{n0}\frac{1 + j\omega(R_s \parallel R_f)C_{tot}}{1 - \omega^2\frac{R_f C_{tot}}{\omega_u} + j\frac{\omega}{\omega_u}\frac{R_f+R_s}{R_s} + j\omega R_f C_f} = G_{n0}(1 + j\omega(R_s \parallel R_f)C_{tot})P, \tag{B.29}$$

where G_{n0} is defined in equation (B.28) and P is defined in equation (B.23).

B.2.4 Transimpedance

The transimpedance of a transimpedance amplifier can be calculated from equations already derived in appendix B.2, since an operational amplifier in inverting configuration is also used here. However, the voltage source has to be replaced by a current source as shown in figure 5.10. The basic equation for the transimpedance amplifier is

$$V_- = aV_{out} + (Z_f \parallel Z_t)I_t.$$

with essentially the same factor

$$a = \frac{Z_t}{Z_f + Z_t} \tag{B.30}$$

as in equation (B.12). The impedance Z_t includes the tunneling resistance (R_t) and the input capacitance (C_{in}), while the impedance Z_f consists of the feedback resistance (R_f) in parallel with the feedback capacitance (C_f). The current source sees the two resistances in parallel as well as both capacitances, which leads to the expression $Z_f \parallel Z_t$.

The action of the operational amplifier is

$$V_{out} = -AV_-$$

and the output voltage is

$$V_{out} = -\frac{A}{1 + aA}(Z_f \parallel Z_t)I_t. \tag{B.31}$$

The transimpedance of the amplifier is defined as

$$T = \frac{V_{out}}{I_t} = \frac{A}{1 + aA}(Z_f \parallel Z_t). \tag{B.32}$$

The first term is equal to the noise gain derived in equation (B.25). For high open-loop gain A the transimpedance reduces to

$$T = \frac{1}{a}(Z_f \parallel Z_t) = -Z_f. \tag{B.33}$$

If the current source is a good current source in the sense that $Z_t \ll Z_f$ then even without assuming high open-loop gains, equation (B.32) simplifies to

$$T = -\frac{A}{1 + A}Z_f. \tag{B.34}$$

Ideal Transimpedance

Inserting the parallel resistor and capacitor impedances into equation (B.33) gives the ideal transimpedance as

$$T = -\frac{R_f}{1 + j\omega R_f C_f}. \tag{B.35}$$

The transimpedance at zero frequency (T_{dc}) is given by the feedback resistance

$$T_{dc} = -R_f. \tag{B.36}$$

Real Transimpedance

The real transimpedance can be found from equation (B.32) by inserting a from equation (B.30) and A from equation (B.16), which gives

$$T = T_{dc} P. \tag{B.37}$$

The factor P is given in equation (B.23) and T_{dc} is defined in equation (B.36).

B.2.5 Amplifier Voltage Noise

To calculate the contribution of amplifier voltage noise one has to integrate the noise density (e_n) multiplied by the noise gain over the frequency band of interest. The rms value is calculated by taking the square root of

$$V_{na}^2 = \int_0^{\omega_1} (G_n e_n)^2 d\omega.$$

The amplifier noise density e_n consists of white noise with an extra $1/f$-type of noise at lower frequencies. Integration is done in three parts, over the $1/f$-noise band, the frequency band with rising noise gain and the remaining noise integrated to ∞.

The voltage noise density of the AD549 is modeled as a sum of white noise and $1/f$-noise,

$$e_n = e_w + e_f' \frac{1}{\sqrt{f}},$$

where $e_w = 35\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ is the white noise of the AD549 and $e_f' = 174\,\mathrm{nV}$, calculated from a known value $e_n = 90\,\mathrm{nV}/\sqrt{\mathrm{Hz}}$ at 10 Hz from the data sheet. The noise density has to be multiplied with the noise gain, which for low frequencies is taken as having only a single rising corner frequency, i.e. the noise gain is approximated as

$$G_n = G_{n0}\sqrt{1 + (\frac{f}{f_{tft}})^2},$$

where the noise gain is expressed in terms of frequency in Hertz, with $f_{tft} = 2\pi\omega_{tft}$. Multiplication results in the product consisting of two parts. The white noise part is

$$G_{n0} e_w \sqrt{1 + (\frac{f}{f_{tft}})^2}$$

and the $1/f$-part is

$$G_{n0}e'_f\sqrt{\frac{1}{f}+\frac{f}{f_{tft}^2}}.$$

To find the voltage noise one has to integrate the square of the sum. The integral of the mixed product, i.e. the one containing $e_w e'_f$, will be zero since the two noise sources are not correlated. The remaining two non-zero parts are

$$(G_{n0}e_w)^2(f_2+\frac{f_{tft}}{3}(\frac{f_2}{f_{tft}})^3)$$

for the white noise part and

$$(G_{n0}e'_f)^2(\ln(f_2)+\frac{1}{2}\left(\frac{f_2}{f_{tft}}\right)^2)$$

for the $1/f$-part. The integration went over a frequency band from 0 to f_2. For this upper frequency we choose the corner frequency $f_f = 2\pi\omega_f$. The leading term of the lower-frequency noise is proportional to $(f_f/f_{tft})^{3/2}$, which shows the importance to keep the ratio f_f/f_{tft} small.

For the current values of the current-to-voltage converter $R_f = 90\,\mathrm{M\Omega}$ and $C_f = 0.25\,\mathrm{pF}$ the frequency f_f is 7.07 kHz. Both G_{n0} and f_{tft} depend on the tunneling resistance. The highest expected tunneling resistance is 3 GΩ, which results in $G_{n0} = 1.03$ and $f_{tft} = 28\,\mathrm{Hz}$. This value of tunneling resistance is arrived at under the conditions of high bias voltage $V_{bias} = 150\,\mathrm{mV}$, which is the linear limit and low tunneling current $I_t = 0.05\,\mathrm{nA}$. A value for the smallest expected tunneling resistance of 5 MΩ can be estimated with low bias voltage $V_{bias} = 5\,\mathrm{mV}$ and high tunneling current $I_t = 1\,\mathrm{nA}$ that is still non-destructive. The resulting noise is the square root of the sum of the two terms, which is 617 μV. The same calculation for the smallest expected tunneling resistance of 5 MΩ is hardly different and gives an integrated noise of 620 μV. The dc noise gain G_{n0} is 19 in this case and the corner frequency f_{tft} is 509 Hz.

At higher frequencies the voltage noise of the amplifier is the constant white noise e_w. The product of white noise and the noise gain can be approximately described as having a certain level dropping off with $-20\,\mathrm{dB/dec}$ at the frequency where the noise gain meets the open-loop gain. The noise gain level is the high-frequency level of the ideal noise gain, which is $G_{peak} = C_{tot}/C_f$, the drop-off corner frequency is at $f_{nA} = f_u/(C_{tot}/C_f)$. The product of white noise and noise density is

$$G_{peak}e_w\frac{1}{\sqrt{1+(\frac{f}{f_{nA}})^2}},$$

which integrates to

$$G_{peak}e_w\sqrt{\frac{\pi}{2}f_{nA}-f_{nA}\tan^{-1}\frac{f_l}{f_{nA}}}$$

for frequencies from f_l to ∞, using equation (B.2). In case $f_l \ll f_{nA}$ the term $f_{nA}\tan^{-1}(f_l/f_{nA})$ can be approximated to f_l using $\tan^{-1}x \approx x$. This should be used for a small input capacitance. For the present current-to-voltage converter this is not the case, the corner frequency $f_{nA} = 3.78\,\mathrm{kHz}$ is even smaller than

$f_f = 7.07\,\mathrm{kHz}$. To get an approximation for the noise we choose $f_1 = f_{nA}$ and get $713\,\mu\mathrm{V}$. This value does not depend on the tunneling resistance. The overall noise contribution of the amplifier is

$$\sqrt{(620\,\mu\mathrm{V})^2 + (713\,\mu\mathrm{V})^2} = 945\,\mu\mathrm{V},$$

where we used the bigger one of the two lower frequency noise terms.

The same calculation can be repeated for a smaller input capacitance of $C_{tot} = 5\,\mathrm{pF}$, where one gets the values $75\,\mu\mathrm{V}$ for the lower frequencies, $187\,\mu\mathrm{V}$ for the high frequencies and $202\,\mu\mathrm{V}$ for the total. This is for the biggest tunneling resistance ($3\,\mathrm{G\Omega}$). The values for the smallest tunneling resistance ($5\,\mathrm{M\Omega}$) are $47\,\mu\mathrm{V}$ for the lower frequencies, $187\,\mu\mathrm{V}$ for the high frequencies and $193\,\mu\mathrm{V}$ for the total.

Conclusion The leading term of the low-frequency amplifier noise is proportional to

$$\begin{aligned} G_{n0}\left(\frac{f_f}{f_{tft}}\right)^{\frac{3}{2}} &= \frac{R_t + R_f}{R_t}\left(\frac{R_t}{R_t + R_f}\frac{C_{tot}}{C_f}\right)^{\frac{3}{2}} \\ &= \left(\frac{R_t}{R_t + R_f}\right)^{\frac{1}{2}}\left(\frac{C_{tot}}{C_f}\right)^{\frac{3}{2}}, \end{aligned} \tag{B.38}$$

while the amplifier noise at high-frequencies is proportional to

$$G_{peak} = \frac{C_{tot}}{C_f}.$$

Both noise contributions can be reduced without penalty by decreasing the input capacitance, which makes up most of C_{tot}. Increasing the feedback capacitance C_f also reduces the noise, but reduces the signal bandwidth at the same time.

If the feedback resistance is $R_f \gg R_t$ then the low-frequency signal-to-noise ratio increases with rising R_f. Otherwise the choice of tunneling resistance has hardly an influence on the amplifier voltage noise.

The results of this calculation are depicted in figure B.7 and B.6, which present the amplifier voltage noise as a function of tunneling resistance and total capacitance. As can be seen in figure B.6 the influence of the total capacitance on the noise is very pronounced. The tunneling resistance as a parameter is not varied, since a variation can not be seen on this scale. A change in tunneling resistance is only influential in the region where the tunneling resistance (R_t) changes between 1 and $10\,\mathrm{M\Omega}$, i.e. for $R_t \ll R_f$. This can be seen in figure B.7.

B.2.6 Future Options

Manipulating the Feedback Fraction The feedback fraction a plays an important role for both signal and noise gain. Equation (B.10) shows the influence of the feedback fraction on the ideal signal gain, while in equation (B.25) it is seen that the feedback fraction a determines the noise gain of the operational amplifier. The inverse of the feedback fraction $1/a$, which is equal to the ideal noise gain (G_n), is given in equation (5.10) as

$$G_n = G_{n0}\frac{1 + j\omega(R_t \parallel R_f)C_{tot}}{1 + j\omega R_f C_f}.$$

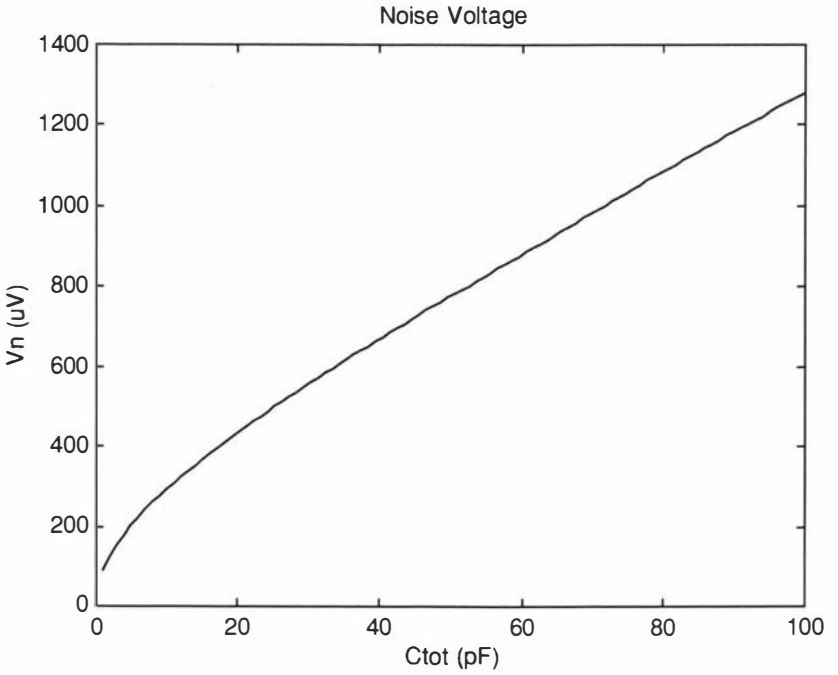


Figure B.6: **Amplifier Voltage Noise:** The strong dependence of the voltage noise on the total capacitance is shown in the graph, where C_{tot} is varied from 0 to 100 pF. The product of noise voltage density and amplifier noise gain was integrated to find the noise voltage. The resulting noise voltage reaches values of over 1 mV for higher values of total capacitance, which is considerable. The variation of the tunneling resistance is negligible and therefore not shown.

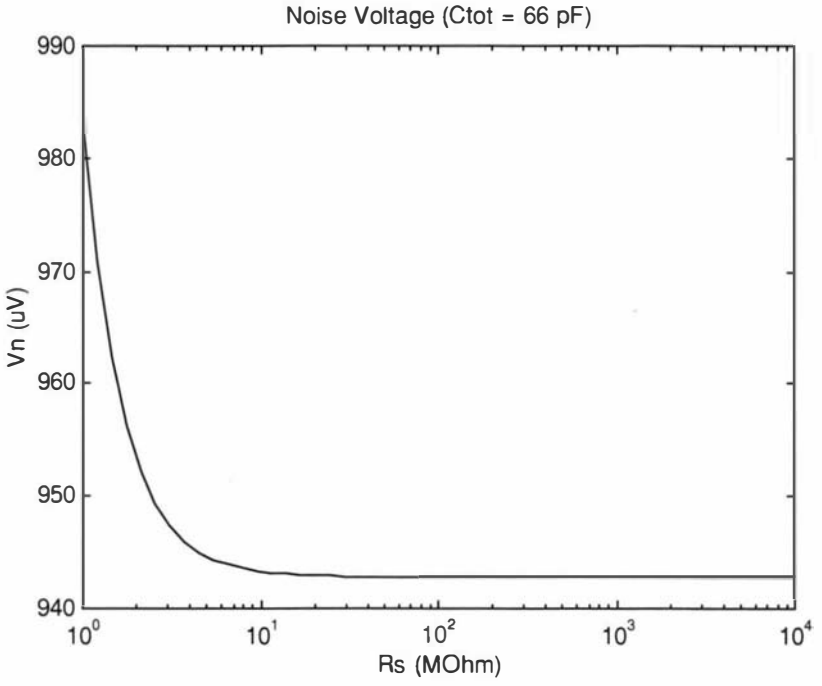


Figure B.7: **Amplifier Voltage Noise:** The tunneling resistance influences the noise gain of the amplifier and therefore also the amplifier voltage noise. As above the product of noise voltage density and amplifier noise gain was integrated to find the noise voltage. Changes in voltage noise are greatest for tunneling resistances smaller than the feedback resistance, which is $R_f = 90\,\mathrm{M}\Omega$. The total capacitance for this graph was $C_{tot} = 66\,\mathrm{pF}$. In comparison to the influence the total capacitance has on the voltage noise the change due to variation of the tunneling resistance is only small.

The gain G_n has two corner frequencies, a low frequency (f_{tft}) after which $1/a$ rises and a high frequency (f_f) after which $1/a$ is leveling again. The low-frequency level of the noise gain is $G_{n0} = (R_f + R_t)(R_t)$, from equation (5.11); the high-frequency level of the ideal noise gain is $G_{nh} = C_{tot}/C_f$ from equation (5.12).

When discussing the voltage noise contribution and the bandwidth of the amplifier we already found that a decrease in C_{in} is paramount in diminishing high-frequency noise gain G_{nh} and increasing the lower corner frequency f_{tft} by reducing C_{tot}.

Looking at the resistance values R_f and R_t one finds that it beneficial for the signal-to-noise ratio to have

$$\frac{R_f}{R_t} \gg 1,$$

i.e. having a large feedback resistance and a small tunneling resistance. The ratio between signal and noise gain at dc is then approaching a value of one, which is its maximum value.

A large feedback resistance value will decrease both corner frequencies of G_n, which is detrimental for the high-frequency signal-to-noise ratio and the signal bandwidth. On the one hand the noise starts increasing already at lower frequencies, on the other hand both signal and noise gain are limited by a low value of f_f.

Since the ideal signal gain begins to roll off at f_f, while the ideal noise gain only reaches a constant level, the operational amplifier is mainly amplifying noise at higher frequencies. To achieve the best possible signal-to-noise ratio one should amplify only in frequency regions where most of the useful signal lies. This can either be done within the feedback loop by actively decreasing the loop gain with a second amplifier or after the feedback loop with the help of a low-pass filter.

A T-network may be considered to replace the single feedback resistor. The T-network is configured in different ways depending on whether the noise or bandwidth of the amplifier is to be controlled. Either a capacitor is placed in parallel with R_2 or a capacitor is placed in series with R_3. The first option decreases the noise gain at higher frequencies, but this is at cost of raised low-frequency noise. The second option acts as a compensation network for the main feedback resistor extending the bandwidth of the feedback resistance by a certain amount. In that it alters the feedback fraction a of the amplifier, the compensation network increases both the signal and the noise bandwidth. This is only useful when the high-frequency level of $G_{nh} = C_{tot}/C_f$ is small and the corner frequency f_f is very low, i.e. for a very high feedback resistor (for instance 1 GΩ) and small input capacitance ($C_{in} < 5$ pF). As soon as the noise gain intercepts the open-loop gain on a rising slope the current-to-voltage converter will oscillate, which means that the compensation network can only extend the bandwidth until that point is reached. This can be at a low frequency of about 10 kHz depending on input capacitance and feedback resistance.

The STM system as a closed-loop controller requires a broad bandwidth. If by some means the signal bandwidth can be extended up to an acceptable frequency then both signal and noise of the current-to-voltage converter can be rolled off at that frequency and the greater noise bandwidth of the current-to-voltage converter would not play a role anymore. This can be achieved by

equipping the current-to-voltage converter with a small feedback resistance and following the current-to-voltage converter with a second bandwidth-limited gain stage. The small feedback resistance decreases the low-frequency signal-to-noise ratio, but the two stages together suppress the high frequency noise. Under the condition that the input capacitance is small this would be a way to attain reasonable bandwidth and signal-to-noise ratio.

B.2.7 T-Network

The bandwidth of a conventional current-to-voltage converter with a single large feedback resistor is in the best case determined by the parasitic capacitance of the feedback resistor. Equation (5.8) of the ideal signal gain shows a low-pass filter behavior above frequencies of $f_f = 1/(2\pi R_f C_f)$. The bandwidth can be increased by using a T-network of resistors in place of the single large resistor. Since the resistors used in the T-network have smaller value, the influence of any parasitic capacitance is greatly diminished.

Amplifier Gain with T-network To describe the behavior of the T-network shown in figure B.8 and consisting of three resistors (R_1, R_2 and R_3), the Thevenin equivalent of the T-network is calculated.

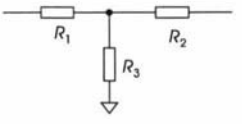


Figure B.8: **T-Network:** Three resistors form a T-network.

The resistors R_2 and R_3 form a voltage divider for the output voltage. This gives the Thevenin voltage as

$$V_T = \frac{R_3}{R_2 + R_3} V_{out}.$$

The Thevenin resistance is

$$R_T = R_1 + (R_2 \parallel R_3).$$

If all capacitances are disregarded for the time being then the amplifier gain with the T-network is

$$\frac{V_{out}}{V_s} = -\frac{R_T}{R_t}\frac{R_2 + R_3}{R_3}, \tag{B.39}$$

where equation (B.11) was used, which was written as

$$V_{out} = -\frac{b}{a} V_s.$$

The fraction a used in equation (B.11) for the T-network is

$$a = \frac{R_t}{R_t + R_T}\frac{R_3}{R_2 + R_3}.$$

The attenuation factor of the T-network is

$$\frac{V_T}{V_{out}} = \frac{R_3}{R_2 + R_3}.$$

The idea of the T-network is that the inverse of the attenuation factor is the multiplication factor for resistor R_1 to replace the single large feedback resistor with

$$\frac{R_2 + R_3}{R_3} R_1.$$

The multiplication factor is usually chosen so that

$$V_T \ll V_{out},$$

which means $R_2 \gg R_3$. For the parallel combination this means $(R_2 \parallel R_3) \approx R_2$. Since the resistor R_1 is usually greater than or equal to R_3, we can approximate the Thevenin resistance with $R_T \approx R_1$. Equation (B.39) can be approximated to

$$\frac{V_{out}}{V_s} \approx -\frac{R_1}{R_t}\frac{R_2}{R_3}.$$

The effective feedback resistance of the T-network is

$$R_f = R_1\frac{R_2}{R_3}.$$

Typical values that could be chosen for the current-to-voltage converter to effect a 1 GΩ resistance would be $R_1 = 10\,\text{MΩ}$ with $R_2 = 100\,\text{kΩ}$ and $R_3 = 1\,\text{kΩ}$.

Effect of the T-network on Bandwidth and Noise The increase in bandwidth can be found by examining the feedback fraction a, seeing where the reciprocal of a intercepts the open-loop gain of the used amplifier. For a given gain-bandwidth product that will be at

$$f_T = f_u\frac{R_3}{R_2},$$

where f_u is the unity-gain frequency, which is equal to the gain-bandwidth product. In our example from above and with $f_u = 1\,\text{MHz}$ that would be

$$f_T = 1\,\text{MHz}\frac{1}{100} = 10\,\text{kHz}.$$

This is but a minor improvement to the current circuit, which achieved about the same bandwidth by reducing the feedback resistor from 1 GΩ to 90 MΩ.

The disadvantage of the T-network is the very decrease in feedback fraction a, which causes an increase in $1/a$ and thereby a multiplication of both offset and noise voltage by the T-network factor of approximately R_2/R_3. The output voltage due to any noise (V_n) at the amplifier input is now

$$V_{out} = \frac{1}{a}V_n = \frac{R_t + R_1}{R_t}\frac{R_2}{R_3}V_n.$$

The same is true for any other voltage at the input, especially the offset voltage of the operational amplifier. The temperature coefficient of the offset voltage is now also multiplied by the same factor R_2/R_3.

Noise Control The T-network can be of merit when used to control the high-frequency noise of the current-to-voltage converter. If we place a capacitor in parallel with R_2 then the T-network becomes frequency dependent and R_2 has to be exchanged for

$$Z_2 = R_2 \parallel C_2 = \frac{R_2}{1 + j\omega R_2 C_2}.$$

This is shown in figure B.9. The inverse of the Thevenin voltage attenuation

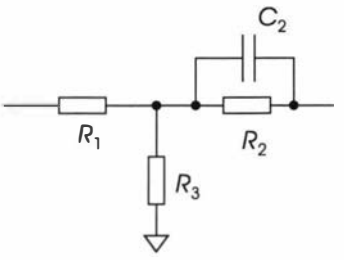


Figure B.9: **T-Network:** This T-network can be used for noise control.

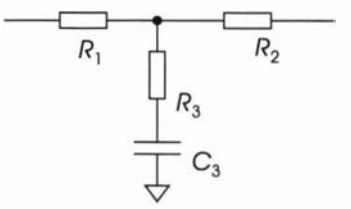


Figure B.10: **Compensation Network:** This T-network can be used for frequency compensation.

factor now reads

$$\frac{V_{\text{out}}}{V_{\text{T}}} = \frac{R_2 + R_3}{R_3} \frac{1 + j\omega(R_2 \parallel R_3)C_2}{1 + j\omega R_2 C_2}.$$

At low frequencies this is the same as the original factor and the T-network provides the same service as before. However, at high frequencies the factor decreases to unity. The roll off starts at $\omega = 1/(R_2 C_2)$ and ends at $\omega = 1/((R_2 \parallel R_3)C_2)$. The effective feedback resistance is now just R_1, which is much smaller than the low-frequency effective feedback resistance, thereby greatly reducing noise gain at higher frequencies than $\omega = 1/((R_2 \parallel R_3)C_2)$.

Compensation Network Another frequency-dependent variant of the T-network is the so-called compensation network, which has the effect of shifting the low-pass filter corner frequency of the feedback resistor from a low to a high frequency. The circuit is shown in figure B.10. The purpose of the network is to compensate for the early gain roll-off of a large feedback resistance due to parasitic capacitance. In the compensation network the lower resistor R_3 is complemented by a series capacitor, so that in the T-network equations one has to replace R_3 with

$$Z_3 = R_3 + \frac{1}{j\omega C_3}.$$

For the inverse of the feedback fraction we get

$$\begin{aligned} \frac{1}{a} &= \frac{R_{\text{t}} + R_{\text{T}}}{R_{\text{t}}} \frac{R_2 + Z_3}{Z_3} \\ &= \frac{R_{\text{t}} + R_{\text{T}}}{R_{\text{t}}} \frac{1 + j\omega(R_2 + R_3)C_2}{1 + j\omega R_2 C_2}. \end{aligned}$$

The compensation network acts in the opposite way to the T-network used for noise control. At low frequencies it has hardly an effect and $1/a$ is given by

$$\frac{1}{a} = \frac{R_{\text{t}} + R_{\text{T}}}{R_{\text{t}}}.$$

Since the compensation network is designed to increase the bandwidth of the feedback resistor one would still have a very large R_1 in the compensation network, i.e. $R_{\text{T}} \approx R_1$. At frequencies higher than $\omega = 1/(R_3 C_3)$ the inverse feedback fraction is boosted by the familiar factor of

$$\frac{R_2 + R_3}{R_3}.$$

The increase starts at $\omega = 1/((R_2 + R_3)C_3)$. If this frequency is placed at the drop-off frequency of the feedback resistance then the overall frequency response is flat until $\omega = 1/(R_3C_3)$. In practice the compensation network is used to shift a low-frequency roll-off at a few kilohertz, for instance 2 kHz or 3 kHz, to a frequency of around 10 kHz. It is not useful to extend the frequency to much higher values since $1/a$ is equal to the noise gain, which is now rising even further until $\omega = 1/(R_3C_3)$ is reached. If the noise gain is extended too much, oscillation of the current-to-voltage converter might result.

B.3 Logarithmic Conversion

B.3.1 Signal Rectifier

To get the absolute value of the buffered and filtered input voltage, the balanced demodulator AD630JN from Analog Devices is used. It is connected so that the incoming signal is modulated with itself:

$$V_{\text{abs}} = \begin{cases} V_{\text{lp}} & : \quad V_{\text{lp}} \geq 0 \\ -V_{\text{lp}} & : \quad V_{\text{lp}} < 0, \end{cases}$$

which is equivalent to

$$V_{\text{abs}} = |V_{\text{lp}}|.$$

A single integrated circuit is used, for it is convenient, precise and linear at frequencies up to and above the used signal bandwidth. The offset error of the AD630JN is given as 500 μV, its input offset current as 50 nA. With an internal feedback resistor of 10 kΩ, the total offset error is 1 mV.

B.3.2 Transconductance Amplifier

Two stages, with one operational amplifier each, are used to make up the transconductance amplifier. A schematic diagram is shown in figure B.11. The input current to the logarithmic converter I_{in} is proportional to the rectified voltage V_{abs}. The first stage serves to reference the input of the second stage to the positive supply voltage $+V_{\text{ana}}$, so that overall the generated current is independent of the particular value of the supply voltage. The output current I_{out} of the transconductance amplifier is given by the voltage across the output pull-up resistor R_{33} divided by its resistance value:

$$I_{\text{out}} = \frac{+V_{\text{ana}} - V_{\text{S2}}}{R_{33}}.$$

V_{D1} is determined by the first stage to be

$$V_{\text{D1}} = +V_{\text{ana}} - \frac{R_{31}}{R_{30}} V_{\text{abs}}.$$

Since the action of the amplifier is $V_{\text{D1}} = V_{\text{S2}}$, the output current ($I_{\text{out}}$) is then

$$I_{\text{out}} = \frac{+V_{\text{ana}} - (V_{\text{ana}} - \frac{R_{31}}{R_{30}} V_{\text{abs}})}{R_{33}} = \frac{R_{31}}{R_{30}R_{33}} V_{\text{abs}}.$$

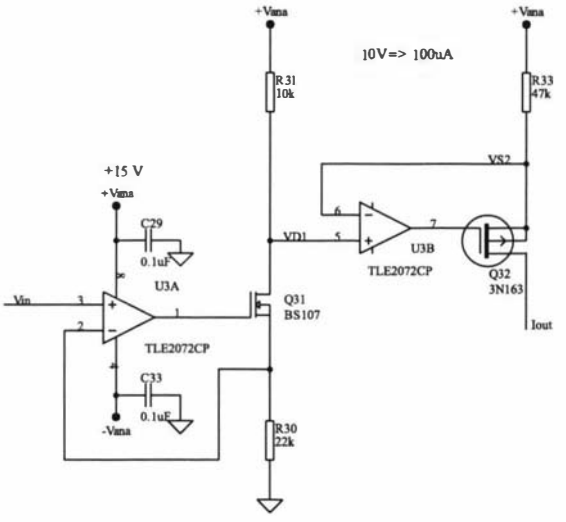


Figure B.11: **Transconductance Amplifier:** The circuit diagram shows a two-stage transconductance amplifier to convert an input voltage (V_{in}) to a proportional current (I_{out}). The output current is the drain current of transistor Q_{32} given mainly by the voltage drop across R_{33}. The first stage serves to translate the input voltage V_{in} to a voltage V_{D1} related to the supply as $V_{\text{D1}} = V_{\text{ana}} - (R_{31}/R_{30})V_{\text{in}}$.

The resistor values have been chosen as

$$I_{\text{in}} = \frac{10\,\text{k}\Omega}{22\,\text{k}\Omega \times 47\,\text{k}\Omega} V_{\text{abs}} = \frac{V_{\text{abs}}}{103\,\text{k}\Omega},$$

which results in a transconductance value of $\Lambda_{\text{vic}} = 10^{-5}\,\text{V/A}$.

The feedback loop of the second stage is established via a MOSFET, which is more accurate in this case than a bipolar transistor, because there is virtually no gate current. Because feedback is not controlling the output current directly, the base current would cause a slight error. The MOSFET used is a 3N163, which is a p-channel enhancement transistor to assure an overall negative feedback for the amplifier U3b.

The equation for the errors of the transconductance amplifier due to offset voltages and bias currents of the operational amplifiers is given by:

$$I_{\text{os}} = \frac{V_{\text{os2}}}{R_{33}} + \frac{R_{31}}{R_{30}R_{33}} V_{\text{os1}} + \frac{R_{31}}{R_{33}}(I_{\text{b1}} + I_{\text{b2}}) \tag{B.40}$$

The indices 1 and 2 indicate the first and second stage of the transconductance amplifier respectively. To achieve very little overall offset output current, it is most important to choose an operational amplifier with low input bias current, since the current terms in the equation do not cancel each other. In this circuit it is also important to consider the input and output range of the operational amplifier. With an input range of up to 10 V and a typical gate-drain voltage for the MOSFET in the first stage (Q_{31}) of 3 V, the output of the first operational amplifier has to be able to deliver up to 13 V. The second operational amplifier has to have an input range of up to its own supply voltage, since the first operational amplifier might close the gate of the first MOSFET. The first restriction can be avoided by using a *pnp* bipolar junction transistor instead of

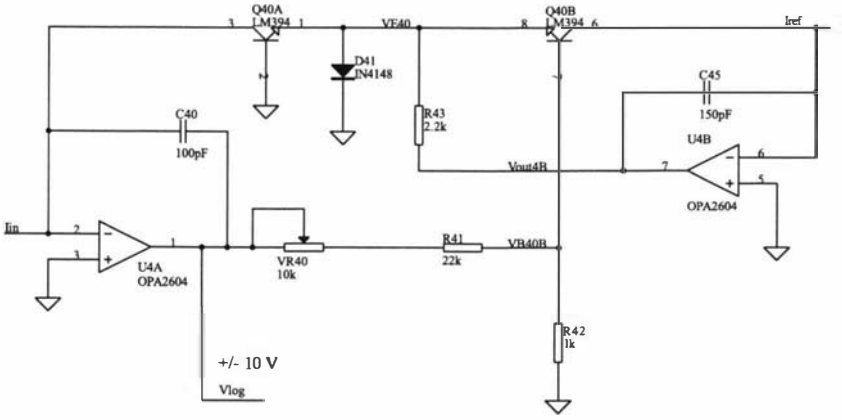


Figure B.12: **Logarithmic Converter:** The emitter voltage V_{E40} is related to the input current due to the logarithmic action of the first transistor (Q_{40a}). The base voltage V_{B40b} of the second transistor Q_{40b} depends on both the input current and the reference current. The logarithmic output V_{log} is the output of the first operational amplifier U4a, which is proportional to the base voltage V_{B40b} via the voltage divider given by R_{41} and R_{42}. Components not related to logarithmic conversion are the capacitors, which stabilize the operational amplifiers, and the diode (D_{41}), which protects the transistors in case of negative input currents.

a MOSFET in the first stage. The disadvantage of that would be a slight error due to the base current of that transistor. Considering the demand on low input bias current and wide input and output voltage range, the dual operational amplifier TLE2072CP was chosen. It has a fairly high offset voltage of 7.8 mV, an input bias current of 5 nA with 1.4 nA offset current. From equation (B.40), the overall offset output current using this operational amplifier is $I_{os} = 240\,\text{nA}$. This is taken into account by an offset trim circuit discussed in section B.3.3.

The bandwidth of the transconductance amplifier is determined by the open-loop gain of the operational amplifiers, their output impedance and the gate capacitances of the transistors. The measured value lies at about 10 kHz, which means that the transconductance amplifier can only be used for slow scans. It is possible to provide an input current for the logarithmic converter from the output of the rectifier by simply using a resistor with a resistance of 100 kΩ. A single resistor would offer higher bandwidth and slightly less noise at the cost of lower precision. The resistor would introduce a gain error in the logarithmic converter due to the finite transistor resistance in the feedback path of the logarithmic converter.

B.3.3 Logarithmic Converter

The logarithmic converter comprises two operational amplifiers as seen in figure B.12. A transistor used as a diode in the feedback path of an operational amplifier causes this configuration to perform in a logarithmic fashion according to the exponential characteristic curve of a diode. In this circuit a second transistor is used for temperature compensation of the strongly temperature dependent offset of the base-emitter voltage V_{BE}. The two transistors used are fabricated on a single chip, the 'supermatch pair' LM394 from National Semiconductor. All transistor properties are better matched than for two single-packed transistors, especially important is the fact that both transistors will have their

body temperatures matched as close as possible. The Ebers-Moll equation relates the collector current I_C and the saturation current I_s to the base-emitter voltage V_{BE} of a transistor as [100]

$$V_{BE} = \frac{kT}{e} \ln(\frac{I_C}{I_s} + 1), \tag{B.41}$$

with k being Boltzmann's constant, T the absolute temperature and e the electron charge. The factor $(kT)/e$ is equal in dimension to a voltage and is about $V_T \approx 25\,\text{mV}$. From equation (B.41), it can be seen that temperature causes a gain error and an offset error via the strong temperature dependence of the saturation current I_s. One could compensate for the gain error with the help of a resistor with a certain positive temperature coefficient. However, for lack of a convenient supply, only an ordinary resistor was used.

To derive an expression for V_{log} in terms of I_{in} the first transistor (Q_{40a}) is considered first. Since the base of the first transistor (Q_{40a}) is at ground, the emitter voltage (V_{E40}) of Q_{40a} is given by

$$V_{E40} = -\frac{kT}{e} \ln(\frac{I_{in}}{I_{s40a}} + 1), \tag{B.42}$$

since the collector voltage of Q_{40a} is at virtual ground. The base voltage V_{B40b} of the second transistor is related to the reference current I_{ref} by

$$V_{B40b} = V_{E40} + \frac{kT}{e} \ln(\frac{I_{ref}}{I_{s40b}} + 1). \tag{B.43}$$

Substituting equation (B.42) into equation (B.43) gives

$$V_{B40b} = -\frac{kT}{e} \ln\left(\frac{I_{s40b}}{I_{s40a}} \frac{I_{in} + I_{s40a}}{I_{ref} + I_{s40b}}\right)$$

Since the transistors are matched, we can assume that $I_{s40a} = I_{s40b}$. We also assume that I_{s40b} is much smaller than both I_{in} and I_{ref}, since the transistors are forward biased. Then the voltage V_{B40b} is found to be

$$V_{B40b} = \frac{kT}{e} \ln(\frac{I_{ref}}{I_{in}}) \tag{B.44}$$

The output voltage (V_{log}) is calculated from V_{B40b} with the voltage divider equation as

$$V_{log} = \frac{R_{42} + R'_{41}}{R_{42}} V_{B40b}.$$

Substituting equation (B.44), the output voltage is finally found to be

$$V_{log} = \frac{R_{42} + R'_{41}}{R_{42}} \frac{kT}{e} \ln(\frac{I_{ref}}{I_{in}}) = -V^0_{log} \ln\left(\frac{I_{in}}{I_{ref}}\right).$$

Both the reference current and I_{ref} the gain factor V^0_{log} have been chosen to give the desired input and output ranges, so that the minimum and maximum input current will give full scale output of the converter. The reference current was set according to:

$$I_{ref} = \sqrt{I_{min} I_{max}} = \sqrt{0.1\,\text{mA} \times 0.1\,\mu\text{A}} = 3.16\,\mu\text{A}.$$

The gain was chosen with the help of:

$$V_{\log}^{0} = \frac{V_{\log,\max}}{\ln\left(\frac{I_{opt}}{I_{max}}\right)}$$

With a resistance value of $R'_{41} = 120\,\mathrm{k\Omega}$ the gain factor will therefore be:

$$\begin{aligned} V_{\log}^{0} &= \frac{R_{42} + R'_{41}}{R_{42}} \frac{kT}{e} \\ &\approx \frac{1\,\mathrm{k\Omega} + 120\,\mathrm{k\Omega}}{1\,\mathrm{k\Omega}} \times 25\,\mathrm{mV} \\ &= 3.0\,\mathrm{V}. \end{aligned}$$

Both the offset and slope of $V_{\log}$ can be adjusted on the circuit board with trim potentiometers. The slope can be manipulated by changing the potentiometer VR_{40}, which changes R'_{41}. The output offset can be trimmed by changing the reference current with VR_{41} in figure B.13. The lower leg of the output voltage divider (R_{43}) could have a positive temperature coefficient to compensate for the temperature dependence of the gain. The value of the coefficient should be roughly equal to $+1/T = +3.4 \times 10^{-3}\,1/\mathrm{K}$.

The major error terms for the logarithmic converter are:

$$\begin{aligned} V_{\log} = & -V_{\log}^{0} \ln\left(\frac{I_{s40b}}{I_{s40a}} \frac{I_{in} - I_{b1} + I_{s40a}}{I_{ref} - I_{b2} + I_{s40b}}\right) + \\ & + \left(1 + \frac{R(40a)}{R(\mathrm{vic})}\right) V_{os1} + \left(1 + \frac{R(40b)}{R(\mathrm{ref})}\right) V_{os2}. \end{aligned}$$

Within the equation R_{40a} and R_{40b} are the effective resistances of the respective feedback paths, R_{vic} is the resistance of the transconductance amplifier and R_{ref} is the resistance of the current reference. For the logarithmic converter a FET input operational amplifier was chosen to minimize the offset error due to input bias current. The output offset current of the transconductance amplifier was found from equation (B.40) to be 240 nA and would have to be added to the input bias current I_{b1}. To adjust this offset a current injection circuit has been included at the inverting input of the logarithmic converter. The current offset adjust can sink or supply currents of $\pm 15\,\mu\mathrm{A}$, which is 15% percent of the maximum input current to the logarithmic amplifier. The leakage currents of the feedback transistor are negligible against the output offset current of the transconductance amplifier. The offset voltage terms show the influence of the internal resistances of the current sources and accordingly the advantages of current sources versus single resistors.

The operational amplifier used in this circuit is the OPA2604AP from Burr Brown, which has an offset voltage of 3 mV and an input bias current of 100 pA with 4 pA offset current. The input bias current is negligible when compared to the output offset current of the transconductance amplifier.

B.3.4 Reference Current

The transistor circuit shown in figure B.13 is used to produce the reference

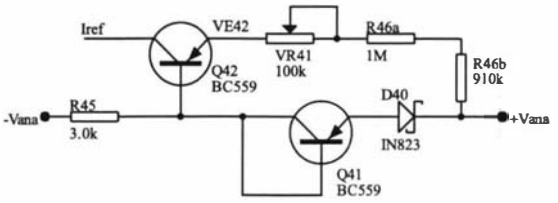


Figure B.13: **Reference Current:** Transistor Q_{42} is the pass transistor. The reference current $I_{\rm ref}$ is mainly given by the Zener voltage across the Zener diode D_{40} and the total resistance of the resistors R_{46a}, R_{46b} and VR_{41}. The second transistor (Q_{41}) is compensating the reference current against variation in temperature. Resistor R_{45} sets the correct current for the Zener diode.

current $I_{\rm ref}$. The output current through the pass transistor (Q_{42}) is given by the voltage (V_Z) across the Zener reference diode D_{40}, which is across the emitter resistance

$$R'_{46} = R_{46a} + R_{46b} + VR_{41}$$

of the pass transistor, plus a correction term:

$$I_{\rm ref} = \frac{V_Z}{R'_{46}} + I_{\rm corr}\,.$$

The desired overall resistance value of the resistors R'_{46} is determined to be

$$R'_{46} = \frac{V_Z}{I_{\rm ref} - I_{\rm corr}} = \frac{6.2\,\mathrm{V}}{3.16\,\mu\mathrm{A} - 0.1\,\mu\mathrm{A}} = 2.0\,\mathrm{M\Omega}.$$

The resistors consists of two series components (R_{46a} and R_{46b}) both of metal film construction so as to achieve a small temperature drift of the reference current. The variable resistor VR_{41} allows a 10% variation of the reference current. The temperature dependence of the saturation current of Q_{42} is compensated by the second transistor (Q_{41}) in diode configuration. To derive the formula for the reference current, one compares the emitter voltage ($V_{\rm E42}$) of Q_{42} given by

$$V_{\rm E42} = +V_{\rm ana} - V_{\rm Z} - V_{\rm EB41} + V_{\rm EB42}$$

with $V_{\rm E42}$ given by

$$V_{\rm E42} = +V_{\rm ana} - R'_{46} I_{\rm E42}.$$

Assuming $V_{\rm EB41} \approx V_{\rm EB42}$ and $I_{\rm E42} \approx I_{\rm ref}$, one gets the above given formula:

$$I_{\rm ref} = \frac{V_{\rm Z}}{R'_{46}}$$

The transistors used are both BC559, low-noise, *pnp*-transistors. The ac characteristics of the selected transistors are not significant in this application. The voltage across the pass transistor is $V_{\rm CE42} = V_{\rm E42} = 9\,\mathrm{V}$, which is within the maximum rating of the BC559 of 30 V.

The Zener diode itself is the temperature compensated reference diode 1N823. The Zener voltage is stated as $V_Z = 6.2$ V at a current of $I_{\rm ZT} = 7.5\,\mathrm{mA}$. The temperature coefficient under these conditions is approximately $5 \times 10^{-5}\,1/\mathrm{K}$. To get such a low temperature coefficient the reference diode has to be provided

with a constant current of 7.5 mA, which is achieved with resistor R_{45}. Ideally R_{45} is

$$R_{45} = \frac{2V_{ana} - V_Z - V_{EB41}}{I_{ZT}} \approx \frac{23\,V}{7.5\,mA} = 3.07\,k\Omega.$$

It has been chosen to be a 3 kΩ resistor, which results in a current of $I_{ZT} = 7.7$ mA.

The mentioned correction value I_{corr} arises owing to the difference in the emitter-base voltages of the two transistors. The complete equation for the reference current is:

$$I_{ref} = \frac{V_Z}{R'_{46}} + \frac{V_{EB41} - V_{EB42}}{R'_{46}} - I_{B42}$$

The error due to base current is calculated with an h_{fE} value of 100 for the BC559 to 0.022 μA and is negligible. The emitter-base voltage term on the other hand is found to be $I_{corr} = 0.10\,\mu A$, which is not negligible. The difference in emitter-base voltages is:

$$V_{EB41} - V_{EB42} \approx V_T \ln\left(\frac{I_{C41}}{I_{C42}}\frac{I_{s42}}{I_{s41}}\right)$$

The collector currents are significantly different:

$$\frac{I_{C41}}{I_{C42}} = 7.7\,mA/3.16\,\mu A = 2.4 \times 10^3.$$

The higher current in the compensation transistor also causes a small difference in saturation currents of $I_{s42}/I_{s41} = 0.88$ due to self heating[1]. With $V_T = 25$ mV, the difference in emitter-base voltages is now found to be 190 mV. This results in a correction of $I_{corr} = 0.10\,\mu A$ to the reference current. The corrected reference current is (3.20±0.17) μA. The error in the reference current is mainly due to the relative error in V_Z of 0.05 given in the data sheet. The reference current can be adjusted with the help of a variable resistor of 200 kΩ over a relative range of 0.1. The variable resistor also serves as a offset trim, since the change in reference current also affects the output offset voltage of the logarithmic amplifier.

If ΔV_{EB} is defined as $\Delta V_{EB} = V_{EB41} - V_{EB42}$, then the temperature coefficient of the reference current is given by

$$\frac{1}{I_{ref}}\frac{dI_{ref}}{dT} = \frac{1}{V_Z}\frac{dV_Z}{dT} + \frac{1}{V_Z}\frac{d(\Delta V_{EB})}{dT} - \frac{V_Z + \Delta V_{EB}}{V_Z}\frac{1}{R'_{46}}\frac{dR'_{46}}{dT},$$

This evaluates to

$$\frac{1}{I_{ref}}\frac{dI_{ref}}{dT} = (5 + 1.4 - 10.3) \times 10^{-5}\,1/K = -3.9 \times 10^{-5}\,1/K,$$

[1]The temperature rise is calculated as follows: The power dissipated in the compensation transistor Q_{41} is $P = 0.7\,V \times 7.5\,mA = 4.2\,mW$. The BC559 data sheet states a thermal resistance from the junction to the ambient surroundings as 200 K/W. The dissipated power will raise the temperature of Q_{41} by 4.2 mW × 200 K/W = 0.8 K. The power dissipation in the pass transistor is negligible. The temperature change will cause a relative saturation current change of

$$\frac{I_s(T + \Delta T)}{I_s(T)} = \frac{V_{gap}}{V_T}\frac{1}{T} \approx 0.18\frac{1}{K} \times \Delta T$$

at room temperature, which is 1.14 for $\Delta T = 0.8$ K.

which is a very small temperature coefficient and means, that the quality of the reference diode has not been compromised by the circuit.

Appendix C

Software

The following text appeared in the January 2000 edition of *Review of Scientific Instruments*. The references therein are self-contained.

REVIEW OF SCIENTIFIC INSTRUMENTS VOLUME 71, NUMBER 1 JANUARY 2000

NOTES

BRIEF contributions in any field of instrumentation or technique within the scope of the journal should be submitted for this section. Contributions should in general not exceed 500 words.

A task-switching system for digital signal processor-based scanning probe microscopes

B. D. Hall[a)]
Measurement Standards Laboratory of New Zealand, Industrial Research Ltd., Lower Hutt, New Zealand

H. Klank and C. D. Eccles
Institute of Fundamental Sciences, Massey University, Palmerston North, New Zealand

(Received 16 March 1999; accepted for publication 3 October 1999)

A simple software framework to manage the concurrent execution of small, programmable, tasks is described. The system is intended for use with a digital signal processor-based scanning probe microscope and is written in ANSI C. The design provides independent memory management, message passing, and structured programming of tasks. © *2000 American Institute of Physics.* [S0034-6748(00)02701-5]

Dedicated digital signal processor (DSP) systems in laboratory-built scanning probe microscopes (SPMs) are becoming increasingly common[1] (for a general reference to DSP systems and applications see Ref. 2). They offer an attractive degree of flexibility to the user because the microscope's behavior can be defined in software, thereby enhancing the possibility of optimizing observations on a particular sample.

In addition to the design of the electronics, software must be considered; ultimately, flexibility in programming the microscope will greatly enhance its usefulness. To date, noncommercial DSP-based SPM systems have taken advantage of manufacturer's evaluation-board systems, which are usually supported with software tools including an ANSI C compiler. In this note, we report on a simple task-scheduling system, written in C, that provides a versatile framework for writing programs to drive the microscope. Our framework has been developed for a SPM system using an Analog Devices ADSP-21020 floating-point DSP (ADSP-21020 EZ-LAB™ Evaluation Board),[3] but is essentially hardware independent in design.[4]

In essence, any DSP-based SPM will operate as a sampled-data control system: it measures some quantity at regular time intervals, and outputs a response. The DSP will continually cycle over at least one task, although several may be "active" concurrently. For example, a probe's height over a sample may be under servo control while its lateral position is varied. It is desirable to write software that recognizes the independent nature of tasks, yet allows them to be combined, as required, during operation. In this note, we describe a scheduling system to cycle and manage tasks, regardless of what those tasks might be.

Our system was inspired by a simple generic control system described by Brown.[5] A list is used to store information about active tasks, using a C structure for each task's list entry. In Brown's system, these structures contain a function pointer and an integer: the function pointer is used to call a task-specific function, and the integer is passed to that function, by reference, allowing it to track its progress between successive calls.

If one considers Brown's task-scheduling system from the programmer's point of view, a number of extensions are desirable. First, task functions cannot define local variables that retain their value. Second, if such storage were available, there is no provision for initialization of task-function parameters, nor can task functions receive information once activated. Third, a procedural approach to task function programming is not supported. Our extension of Brown's design, which we refer to as BOS, provides the following features: arbitrary initial task parameters; arbitrary persistent variables; constructs for structured task programming; and inter-task communication.

To BOS, a task is comprised of three functions, each with distinct roles: a constructor, to configure local memory and initialize a task; a destructor, to release memory and handle any cleanup; and an executor—to perform the task itself. The constructor and destructor are called only once each: as the task begins execution and as it finishes. The executor will be called as many times as is required to complete the task in hand. Each of these functions has a standard format, allowing BOS to use pointers to manipulate them.

Arguments to the constructor function may initialize a task, and the constructor can also allocate memory for use during task execution. If memory is reserved, it must be released on completion; a destructor function handles this, as well as dealing with early task termination. The latter might occur if a human operator signaled a "reset," in which case the active tasks must be shut down in an orderly fashion.

A task executor function must be designed to operate while continually starting and stopping. Any task is broken down into small steps, each of which will complete within the control loop cycle-time available (BOS is not a preemp

a)Electronic mail: b.hall@irl.cri.nz

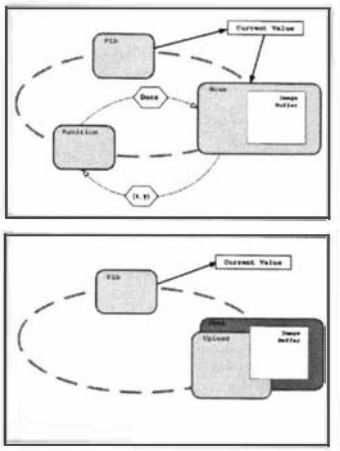


FIG. 1. These two diagrams show the relationship between tasks during acquisition of an image (upper figure) and during uploading of data to a host on completion of scanning (lower figure). Tasks (strictly, task executors) are represented by gray boxes; with the BOS task list shown as a thick, broken, dark-gray line, linking each active task. In both figures, PID maintains the probe above the sample surface, in a way that is specific to the type of microscope; it also stores the most recently sampled value in a global variable (Current Value). During scanning (upper figure), Position receives messages from Scan containing new x-y probe coordinates (messages are shown as angle-ended boxes and a dotted line); an acknowledgment is sent when the probe is in position. Scan can therefore coordinate probe movement while it collects image data, which are stored in a private memory buffer. When data acquisition is complete, the situation will be as shown in the lower figure: PID continues, independently; Position has ceased activity (as a result of a message from Scan); Scan has activated a child task, Upload and given it access to data in the image buffer. When data transfer to the host is complete (not shown), Upload ceases, Scan's destructor releases the memory buffer leaving PID as the only active task.

tive system—it does not intervene and stop a task that is taking too long). At every call, an executor must determine the appropriate step to take, generally using some persistent local variable to do so.

In some cases, task programming can become complicated. For example, different concurrent tasks may need access to a single physical resource (e.g., a communication channel with the PC host), in which case there needs to be some arbitration over access using a common protocol. To ease programming, BOS allows one task to *substitute* another "child" task in its place on the execution list: the "parent" resumes activity only when the child has finished. This allows a structured procedural approach to task programming, and makes it possible to share common procedures between tasks.

One other feature of BOS is its mechanism for passing of information between tasks, whether active or suspended (parents). Each task has a unique identifier, allowing unambiguous delivery of arbitrary "messages" between tasks. A message stack is maintained for each task by BOS and new messages can be retrieved each time the executor is activated. This feature allows synchronization of tasks, as well as sharing of data.

To illustrate BOS task design, consider the case of generic SPM raster imaging of a surface. We can assume that the probe is already within working distance of a surface and under servo control, so a task, called "PID," is active (see Fig. 1). The PID's executor can be designed to store the most recently sampled value in a global memory location. In that way, other active tasks can read the current value.

To obtain a raster image, we use a task called "Scan." Its constructor is provided with scan dimensions, it then allocates a memory buffer to store the image data, and activates a *concurrent* task, "Position," to handle horizontal probe displacement. A convenient way to send x-y coordinates to Position is with messages. Hence, during execution, Scan generates a raster pattern of coordinates for the probe tip, and sends them to Position: at each point, Scan copies a sample value to the buffer. When the image acquisition is complete, Scan uses a child task to upload the image data, from the buffer, to a host computer for analysis and display.

If raster scans were the only mode of operation, then BOS offers no advantage over a more direct approach to programming of the SPM. However, where flexibility is desired, our system offers some significant features. First, tasks are independent: for example, there could be a number of control algorithms, interchangeable with PID. Second, the choice of task to activate is made while the DSP is running; re-compilation of the main control structure is not required, because pointers to tasks are used by BOS. Hence, for example, different types of scan (assuming there is a range available) could be selected during an experiment. Third, tasks can operate in parallel. For example, one might wish to change an experimental parameter during a scan (e.g., sample-tip bias, for a scanning tunneling microscope). This can be achieved by using short-lived tasks which the operator could activate at will (in terms of Fig. 1, this implies that other tasks are simply added to the BOS loop, which should not affect those already there).

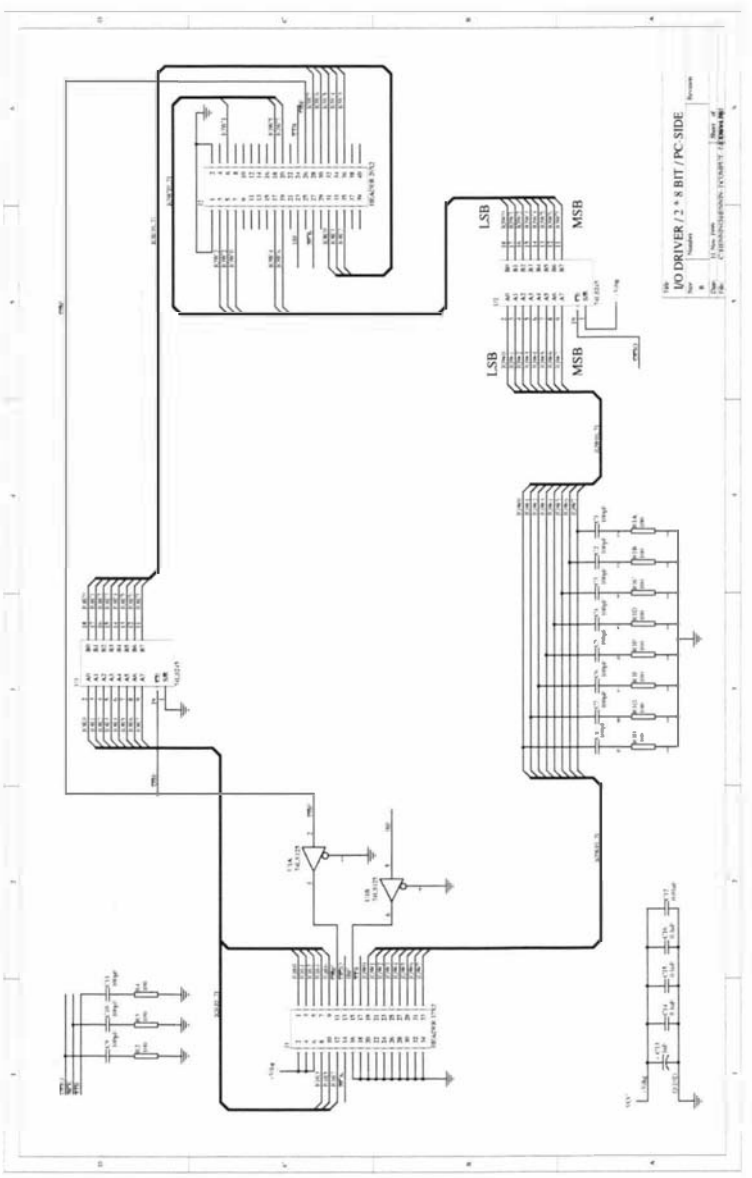
BOS separates the inherent cycling of the control loop operation from specific task implementation. It also recognizes, and supports, the distinct nature of individual tasks. These features make programs more flexible, modular, reliable and readable. The size of BOS is modest: it occupies 6k words of program memory (out of the 32k words available). In terms of performance, the overhead of going once through the control cycle is approximately 170 machine cycles, which on our 33.3 MHz system is about 5 μs. The overhead of a single task-executor function call is about 12 machine cycles (0.36 μs). This allows BOS to be useful for systems with a sampling rate in the tens of kilohertz.

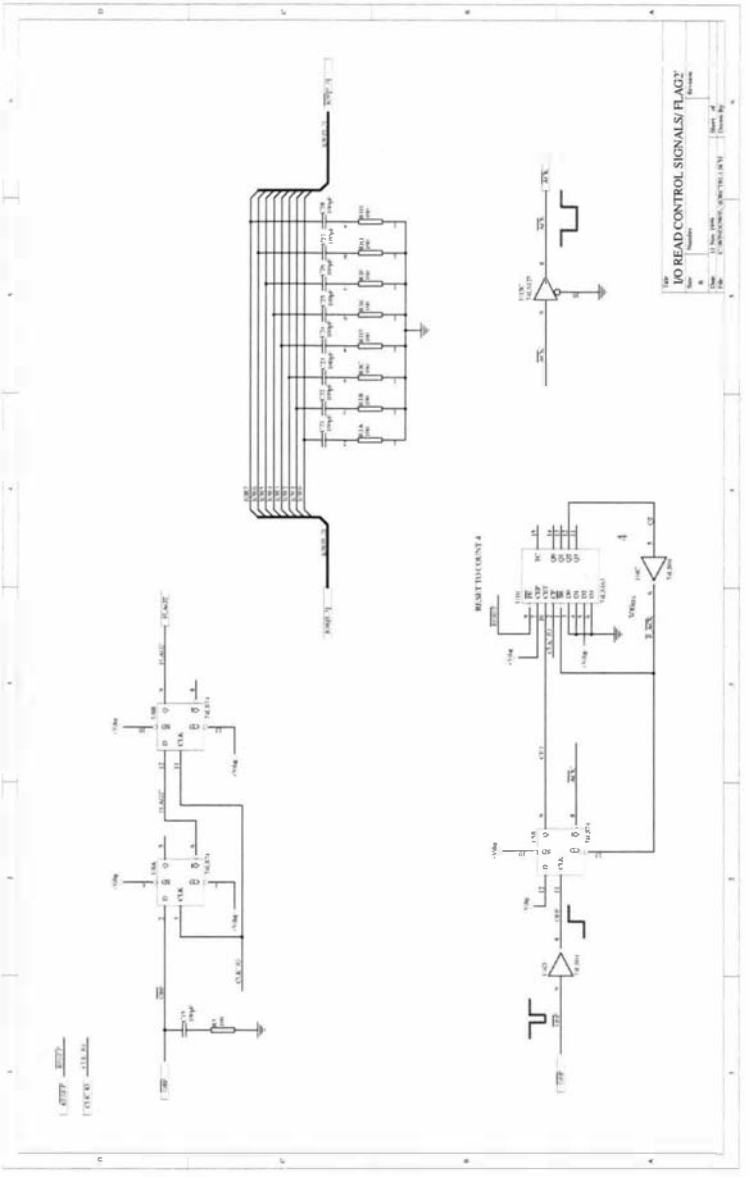
[1] J. Scheuing, W. Clauss, and D. P. Kern, Rev. Sci. Instrum. 69, 4191 (1998); B. Paillard and R. Tang, *ibid* 69, 1770 (1998); D. R. Baselt, S. M.Clark, M. G. Youngquist, C. F. Spence, and J. D. Baldeschwieler, *ibid* 64, 1874 (1993).
[2] R. J. Higgins, *Digital Signal Processing in VSLI* (Prentice–Hall, Englewood Cliffs, NJ, 1989).
[3] H. Klank, B. D. Hall, and C. D. Eccles (to be published).
[4] Source code and documentation are available on request: interested readers should contact B. D. Hall.
[5] R. Brown, Electronics Design News 35, 215 (1990).

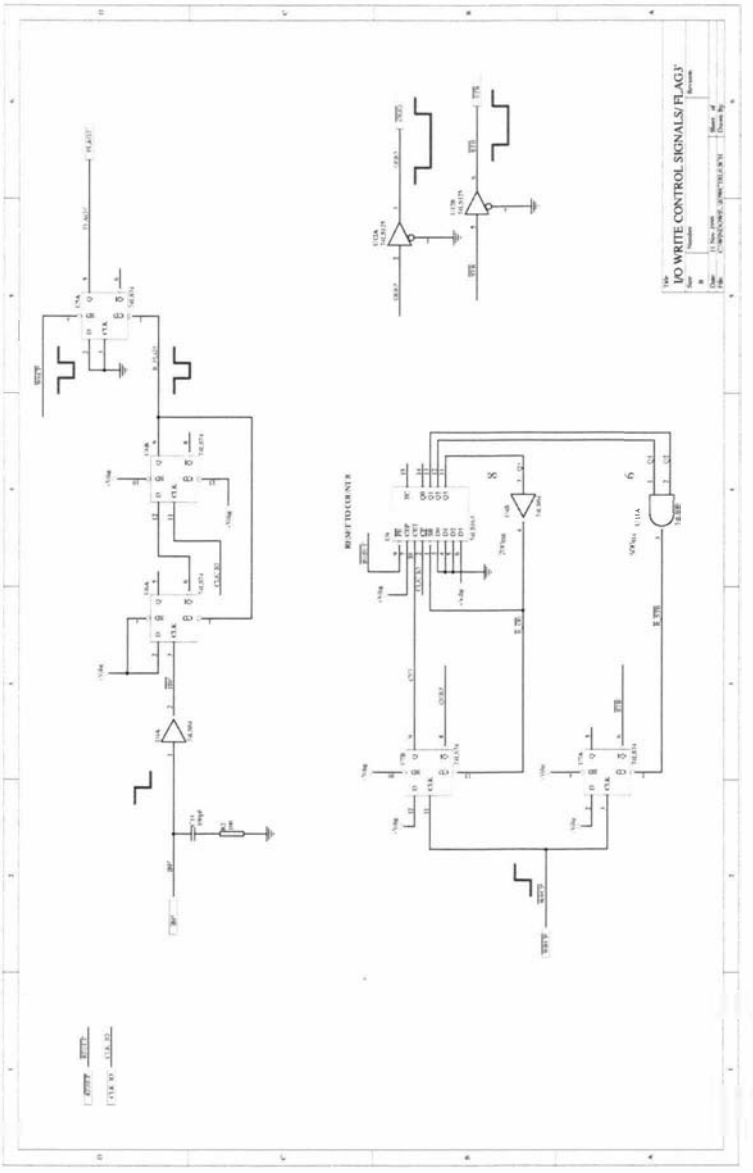
Appendix D

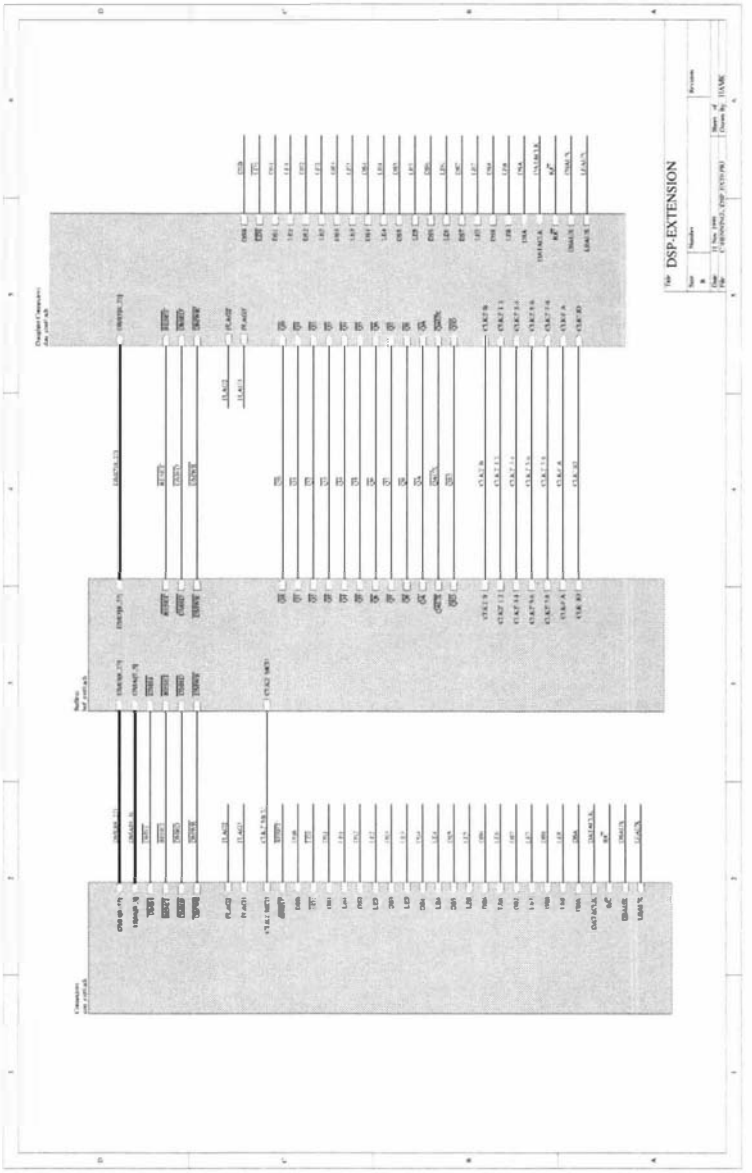
Circuits

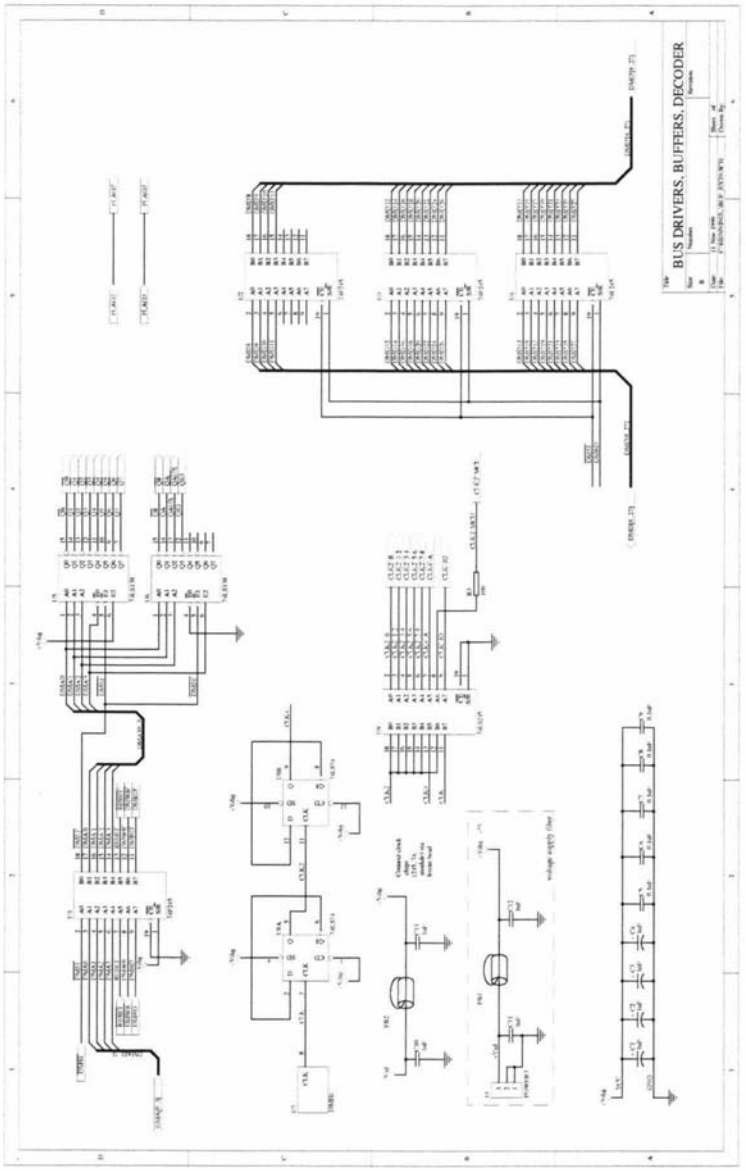
The schematic diagrams of the DSP System Interface and the probe electronics is shown on the following pages.

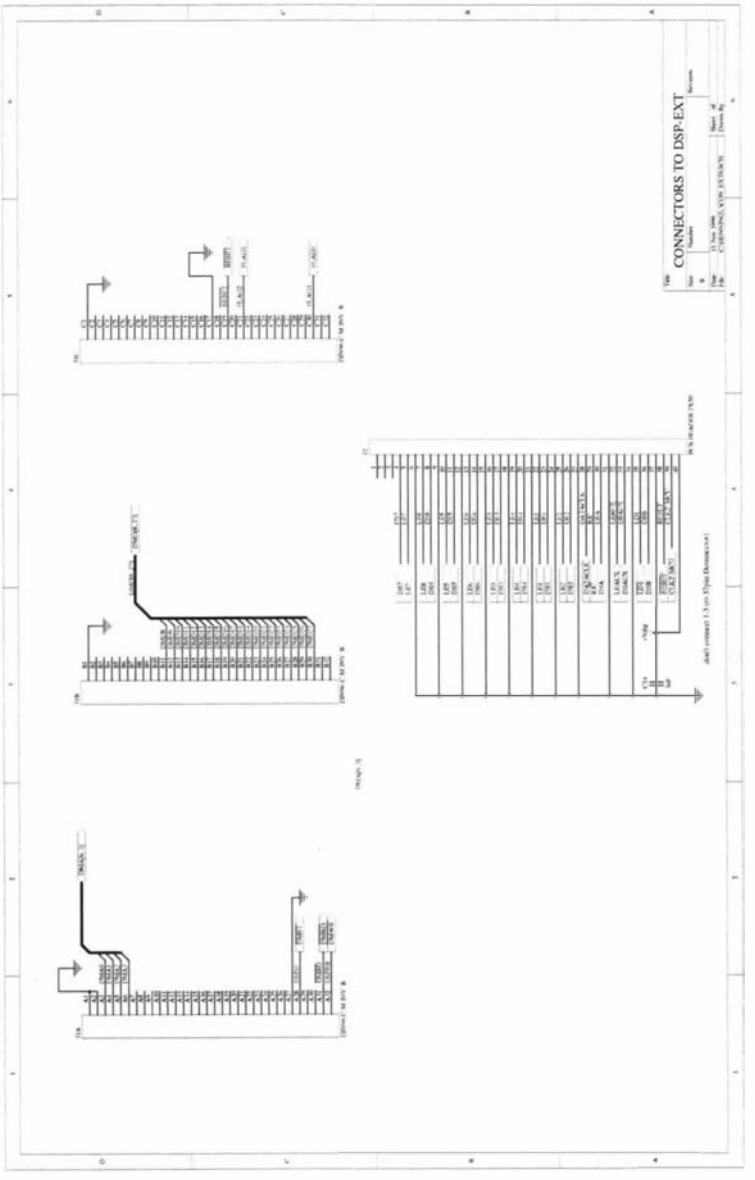

I/O DRIVER / 2 * 8 BIT / PC-SIDE
LSB
MSB
LSB
MSB

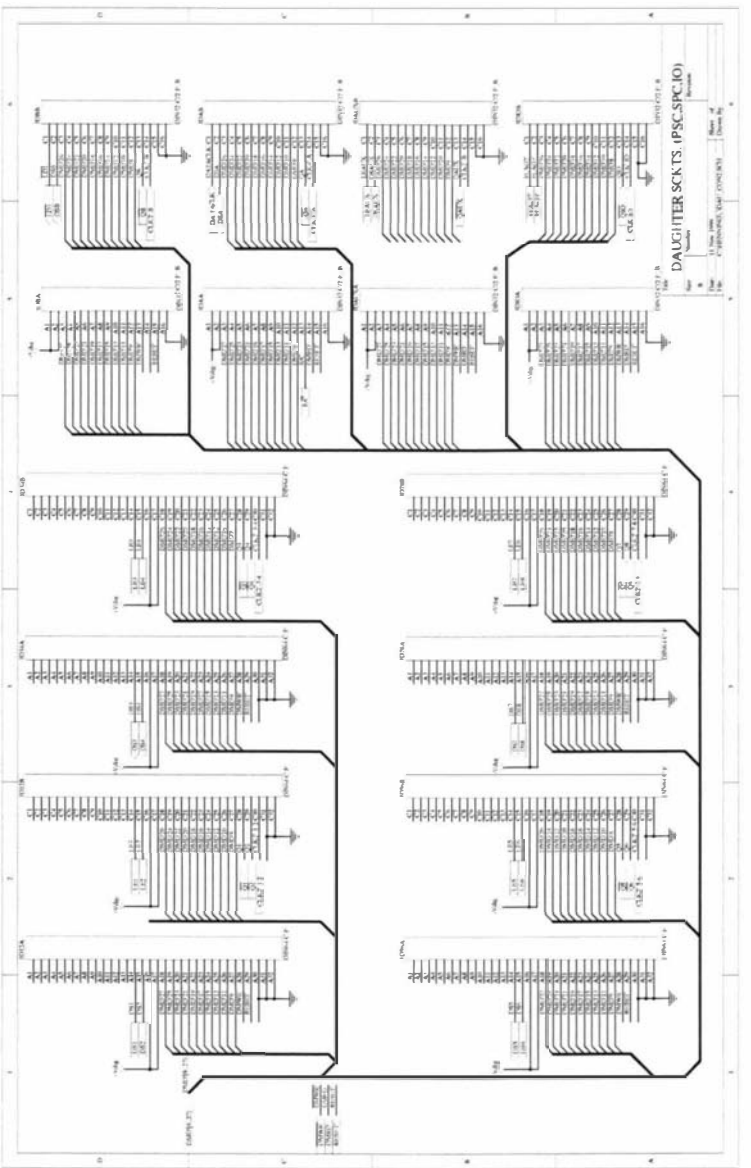

I/O READ CONTROL SIGNALS/ FLAG2'
5555

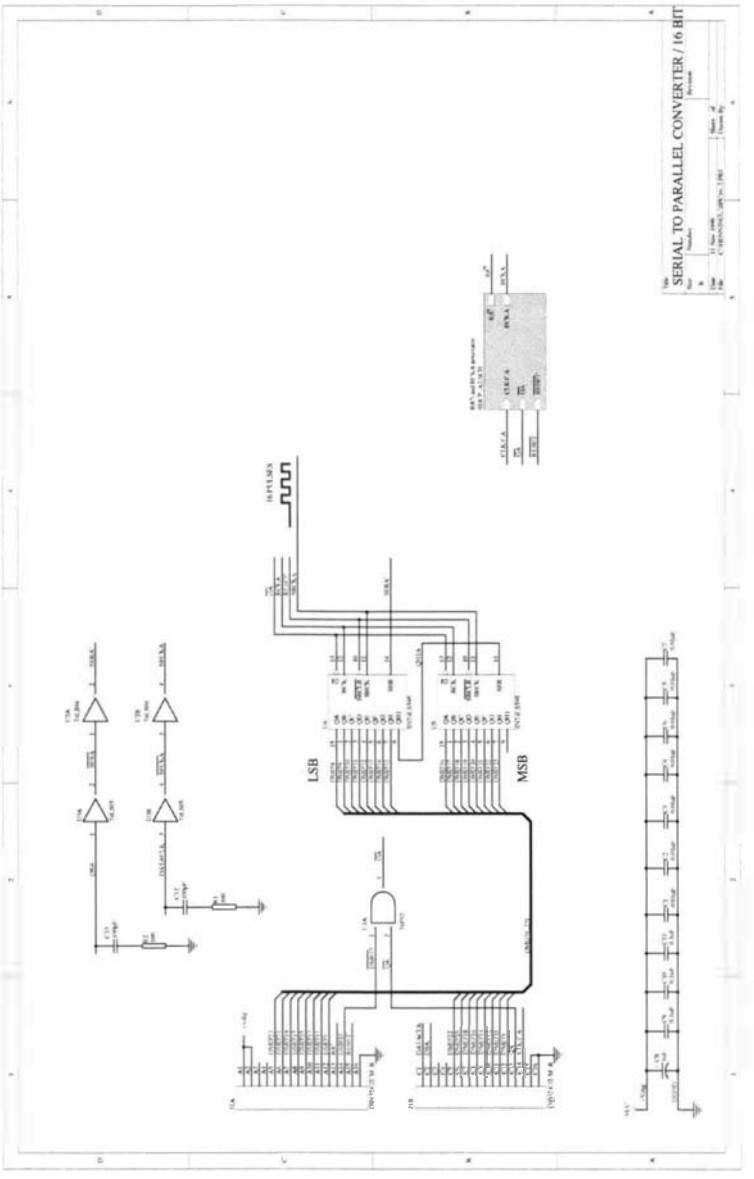

I/O WRITE CONTROL SIGNALS/ FLAGS

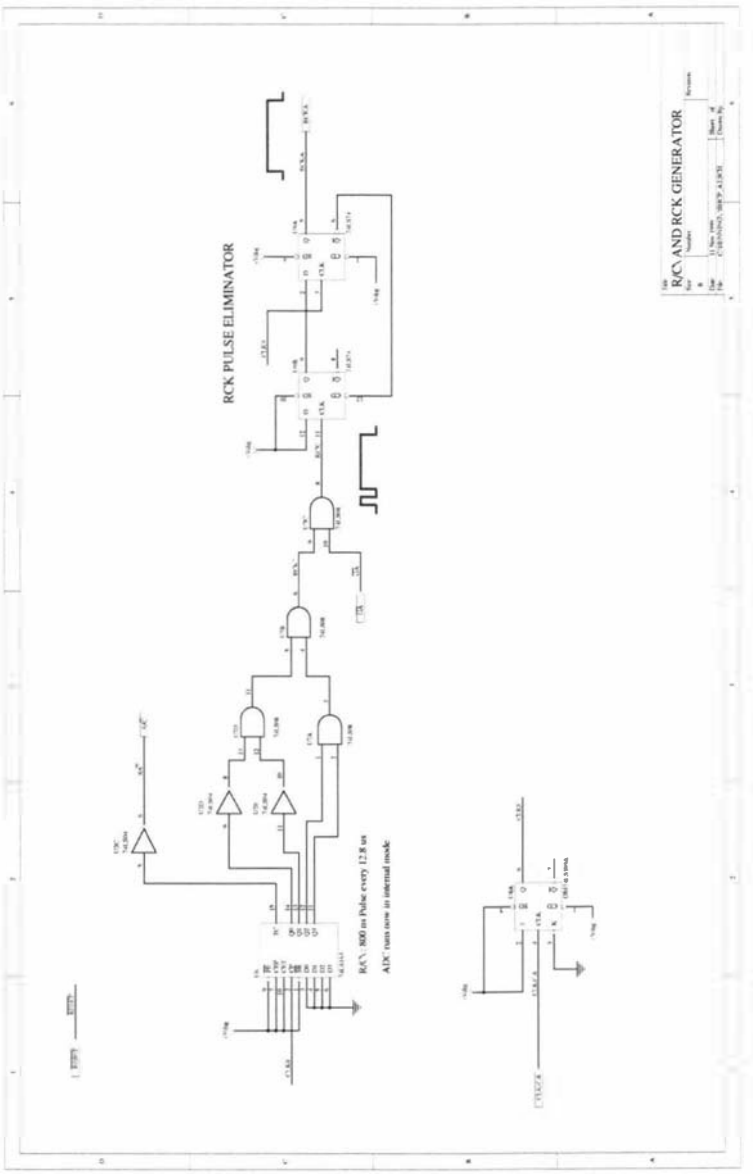

DSP-EXTENSION

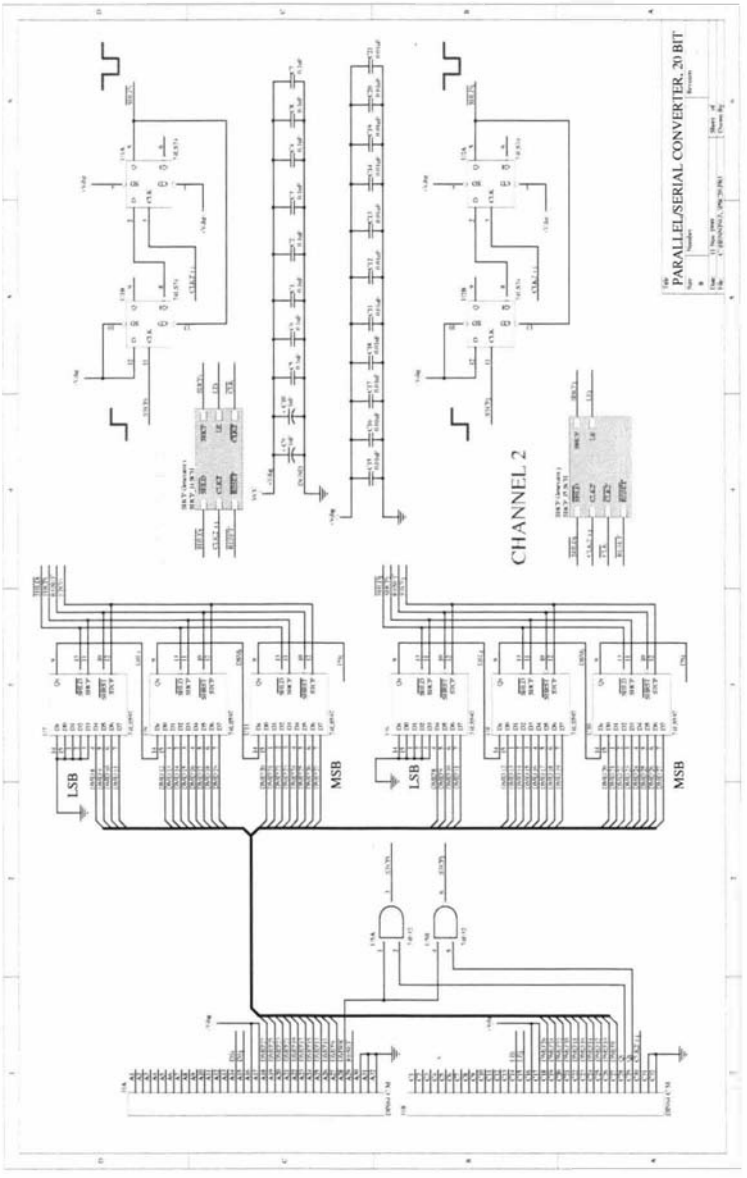

BUS DRIVERS, BUFFERS, DECODER

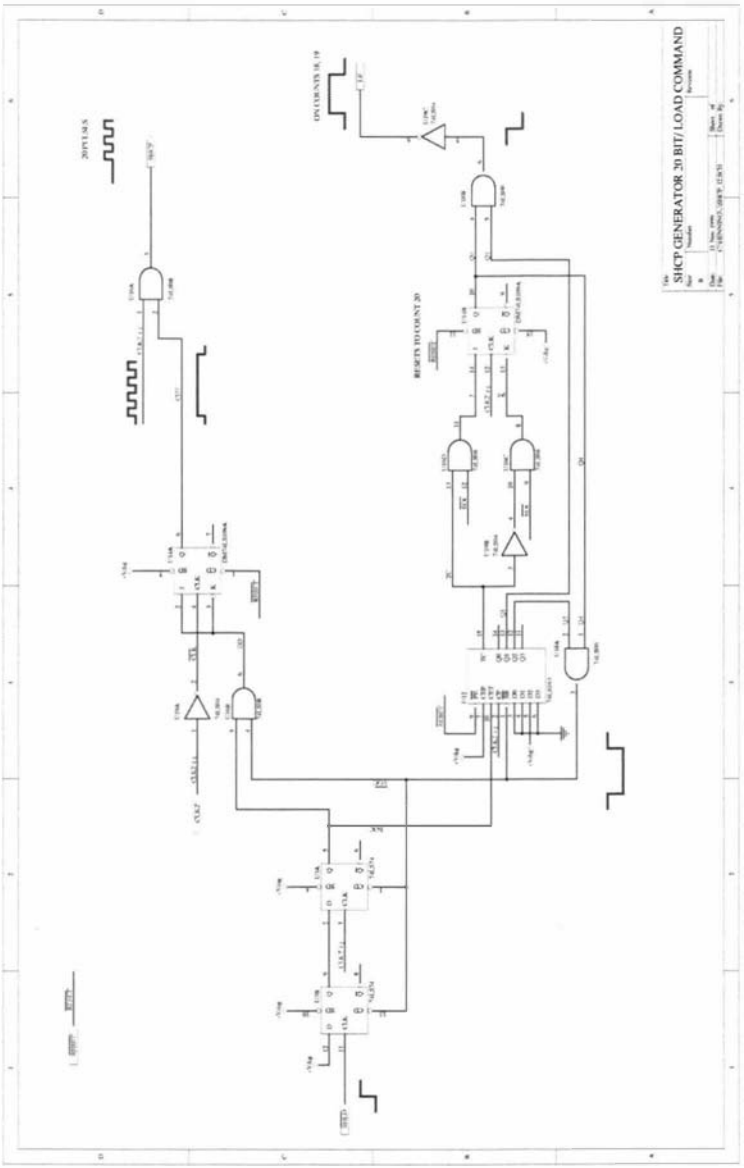

CONNECTORS TO DSP-EXT

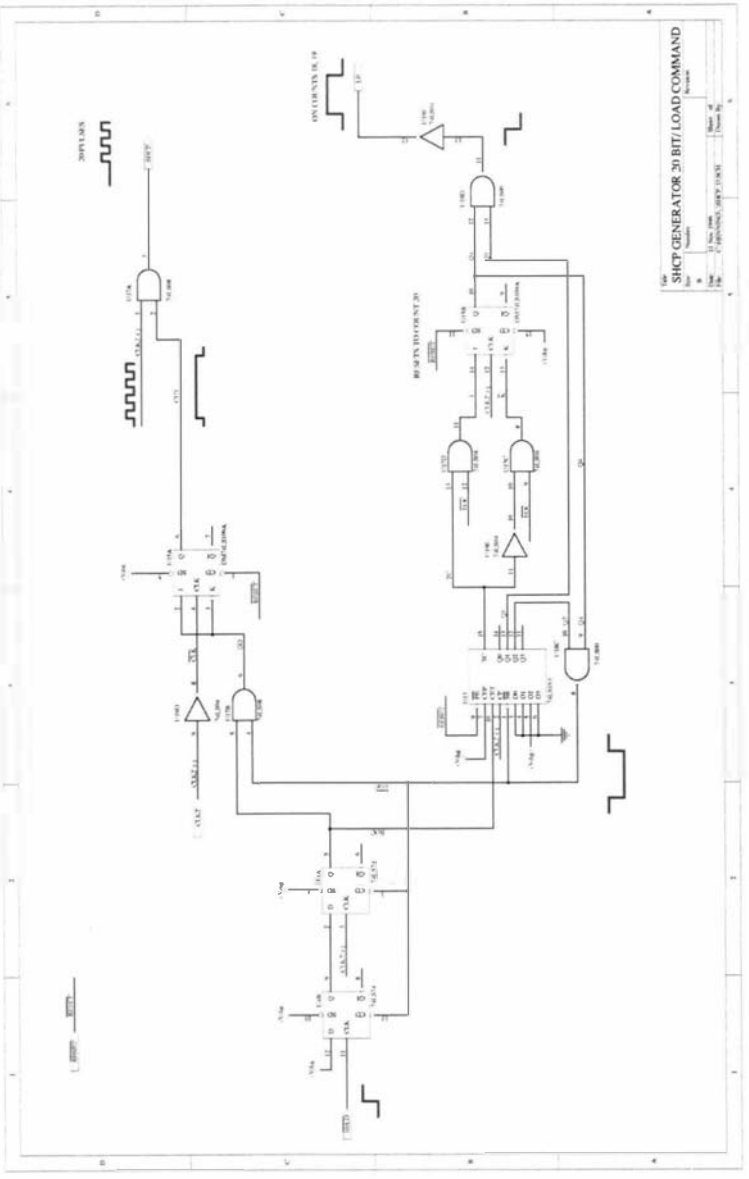

DAUGHTER SOCKTS. (PSC,SPC,IO)

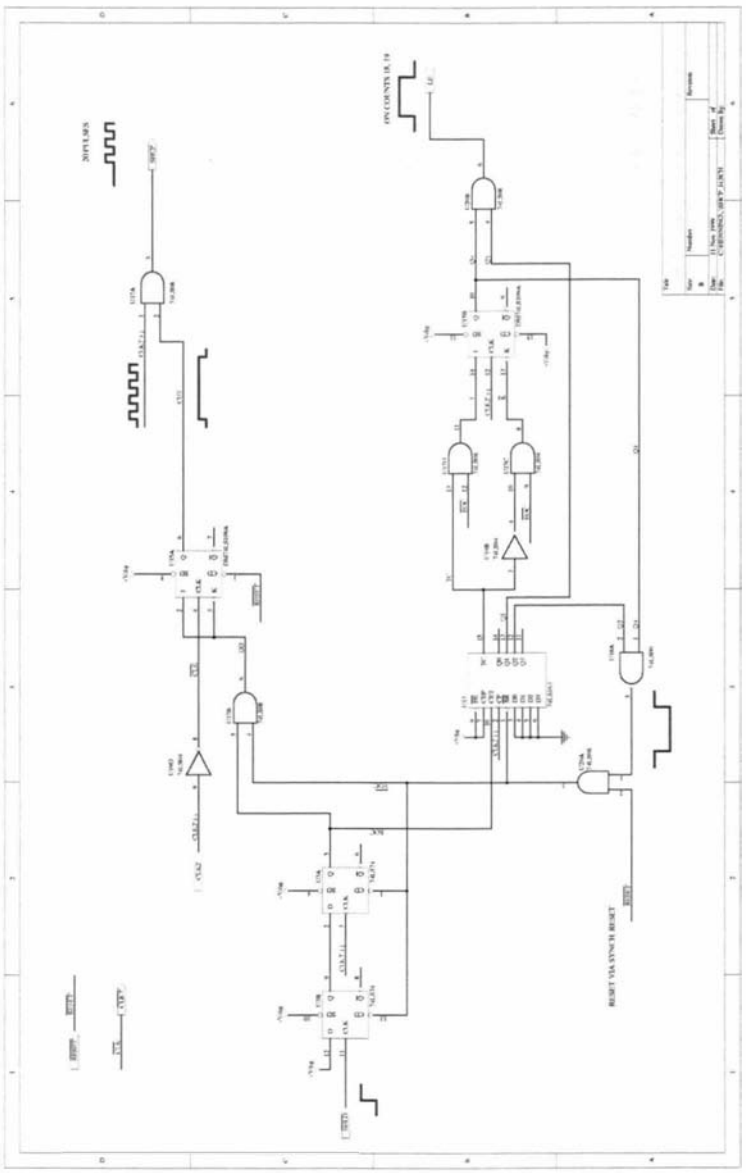

SERIAL TO PARALLEL CONVERTER / 16 BIT
LSB
MSB

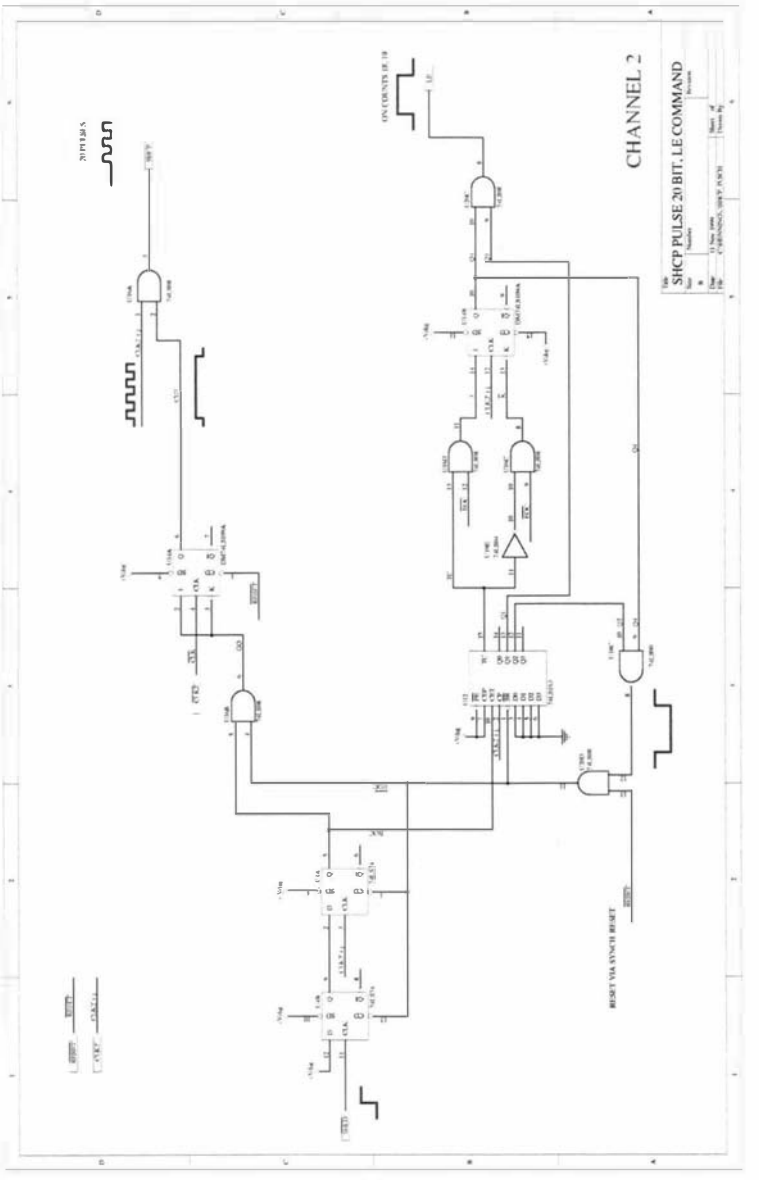

RCK PULSE ELIMINATOR
R/C\ AND RCK GENERATOR
8555
R/C\ 800 ns Pulse every 12.8 us
ADC runs now in internal mode

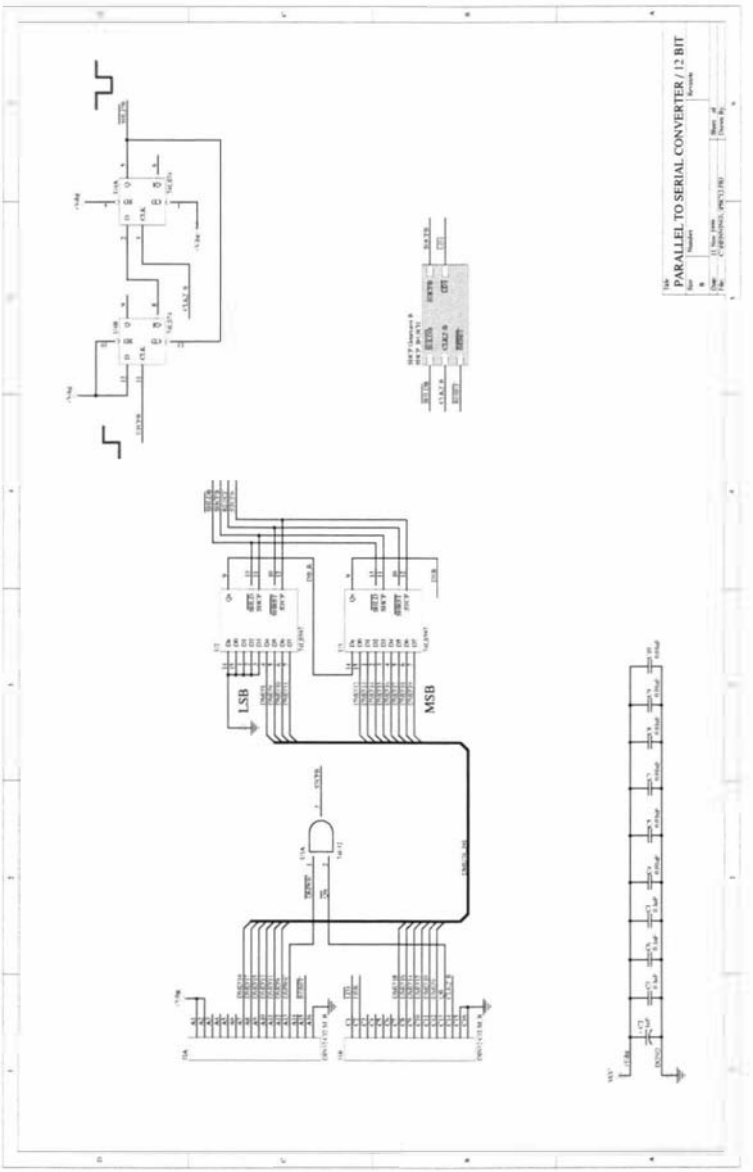

PARALLEL/SERIAL CONVERTER, 20 BIT
CHANNEL 2
LSB
MSB
LSB
MSB

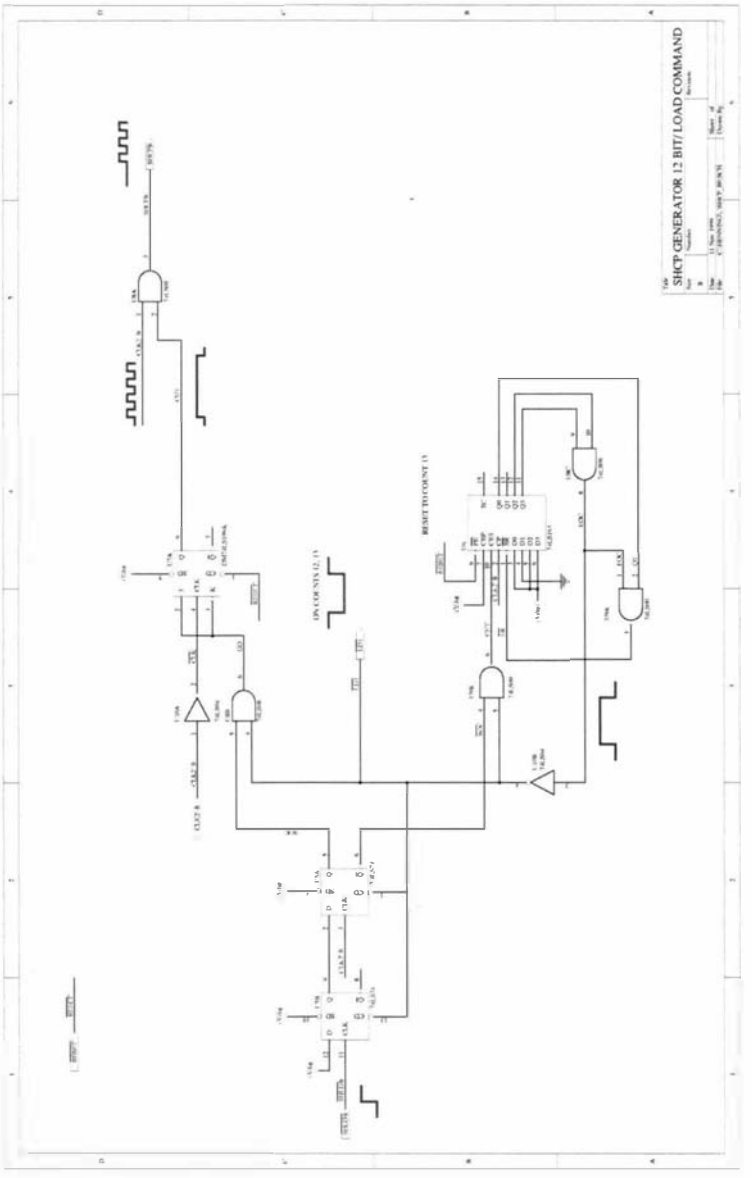

SHCP GENERATOR 20 BIT/ LOAD COMMAND
555

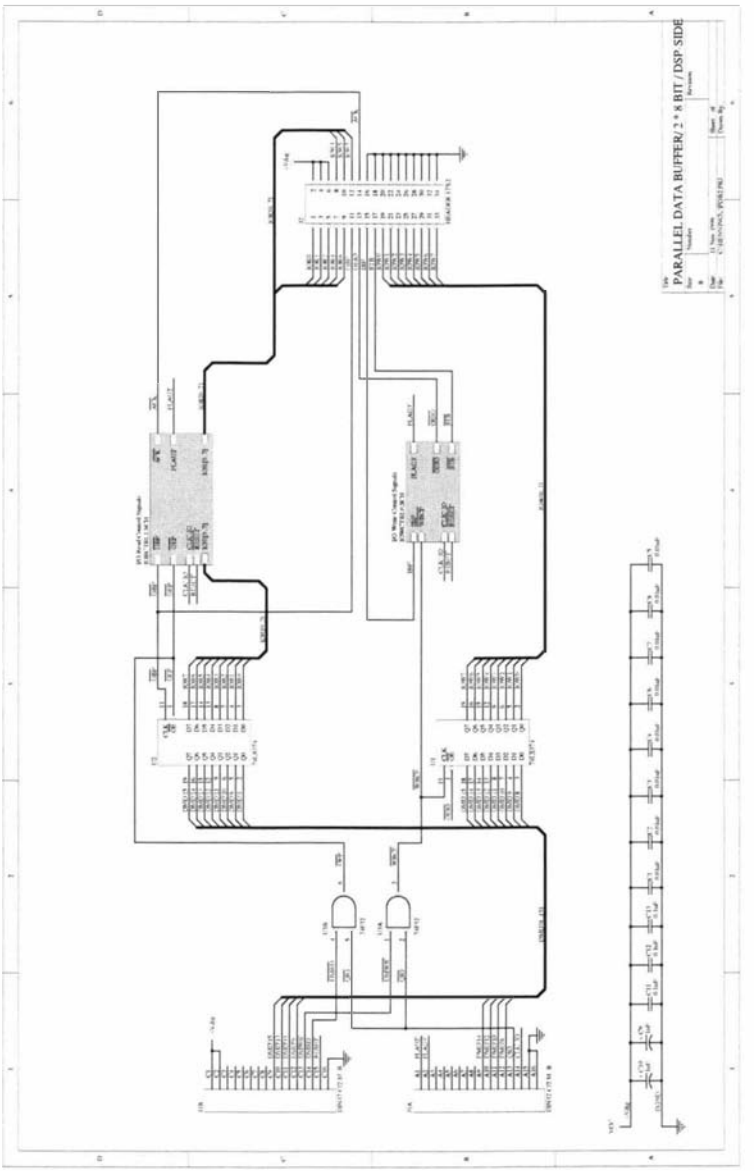

SHCP GENERATOR 20 BIT/ LOAD COMMAND
5555

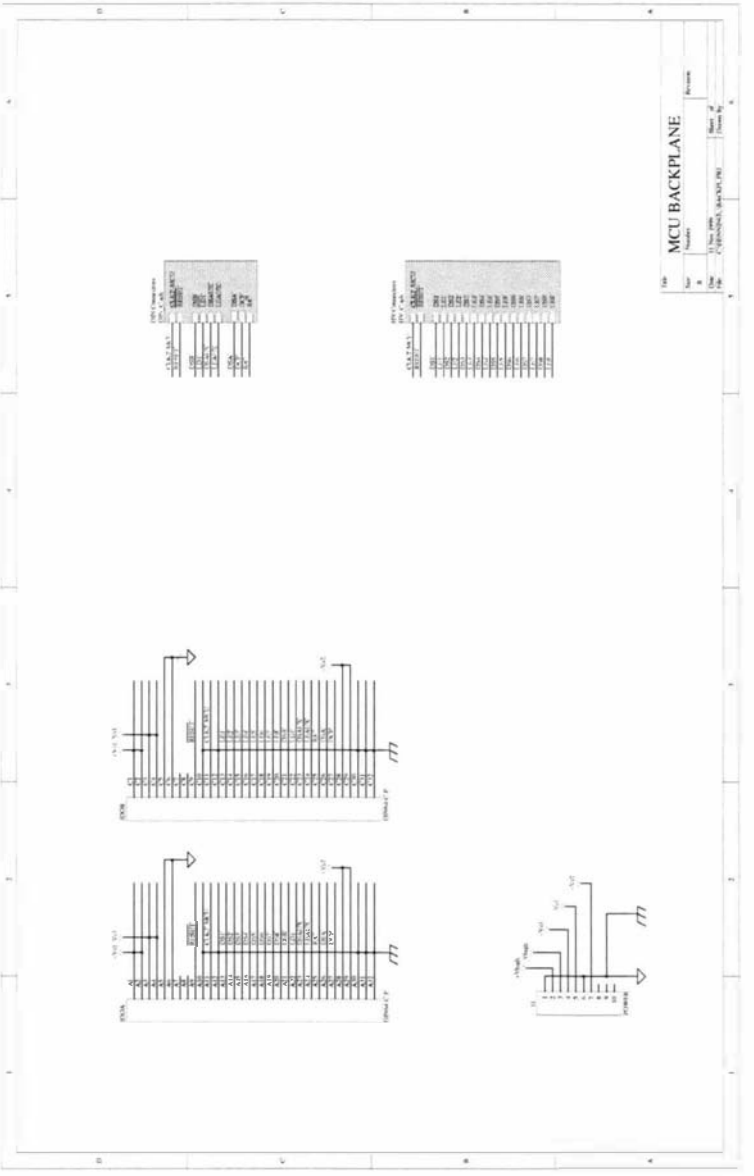


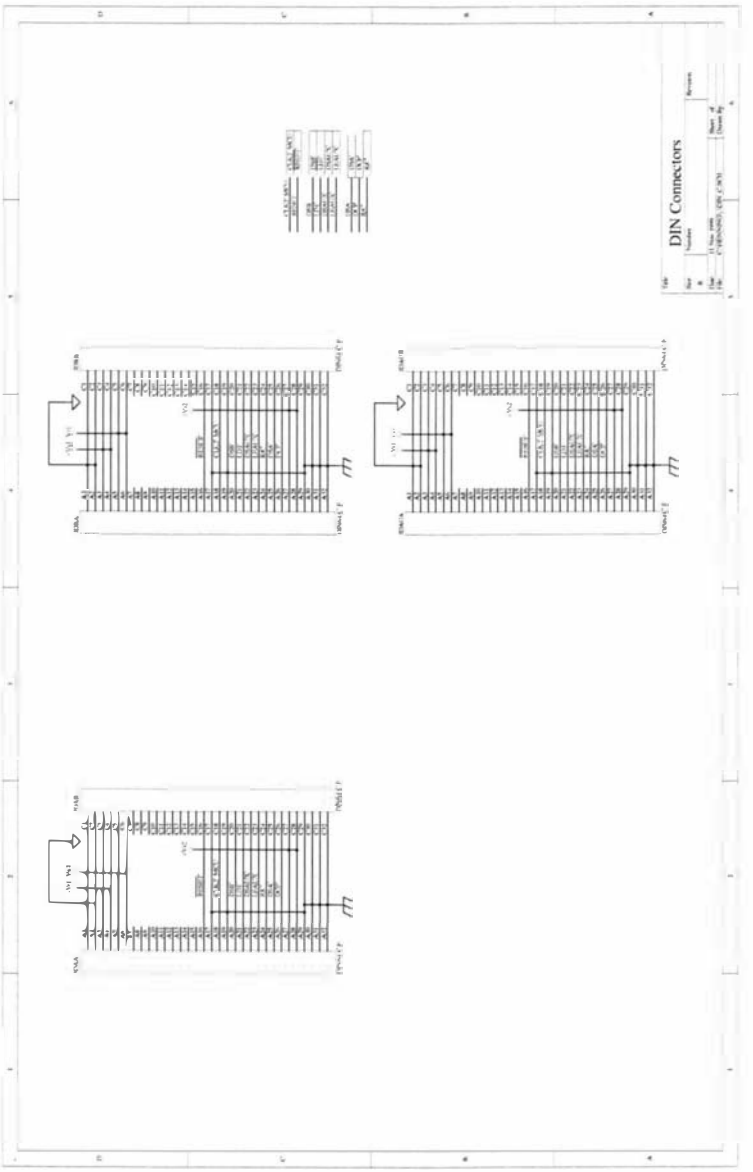

CHANNEL 2
SHCP PULSE 20 BIT. LE COMMAND
5555

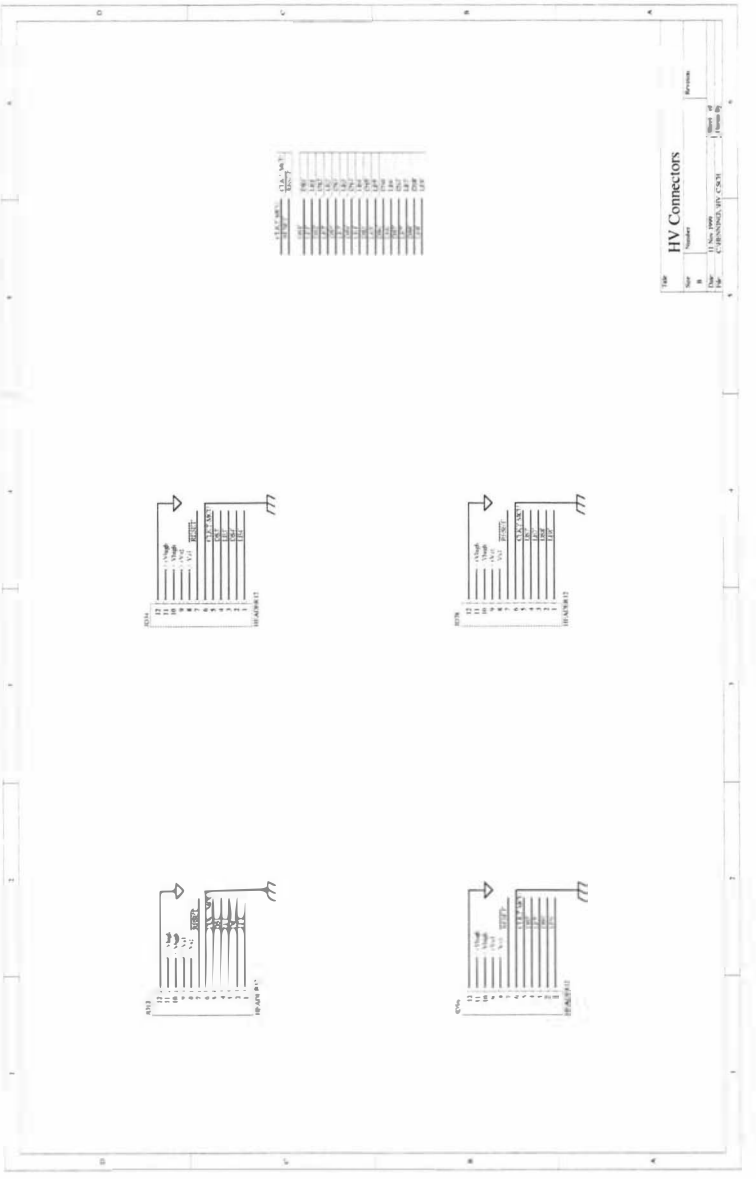

PARALLEL TO SERIAL CONVERTER / 12 BIT
LSB
MSB

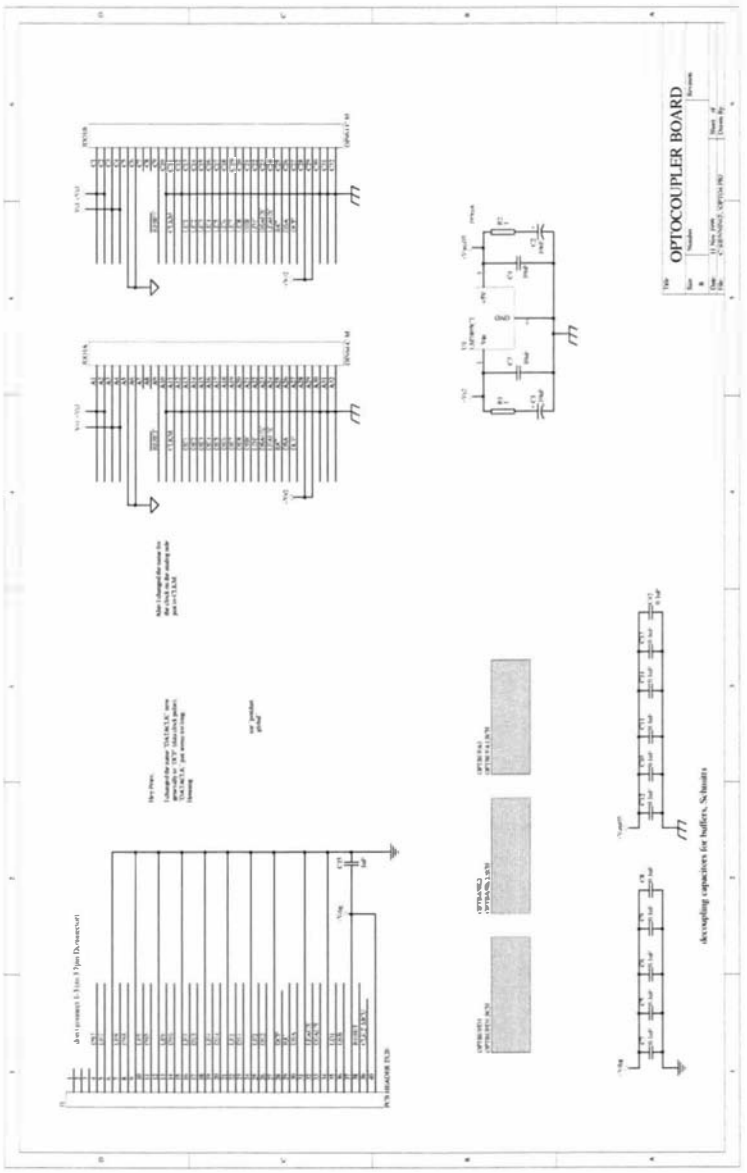

SHCP GENERATOR 12 BIT/LOAD COMMAND
555

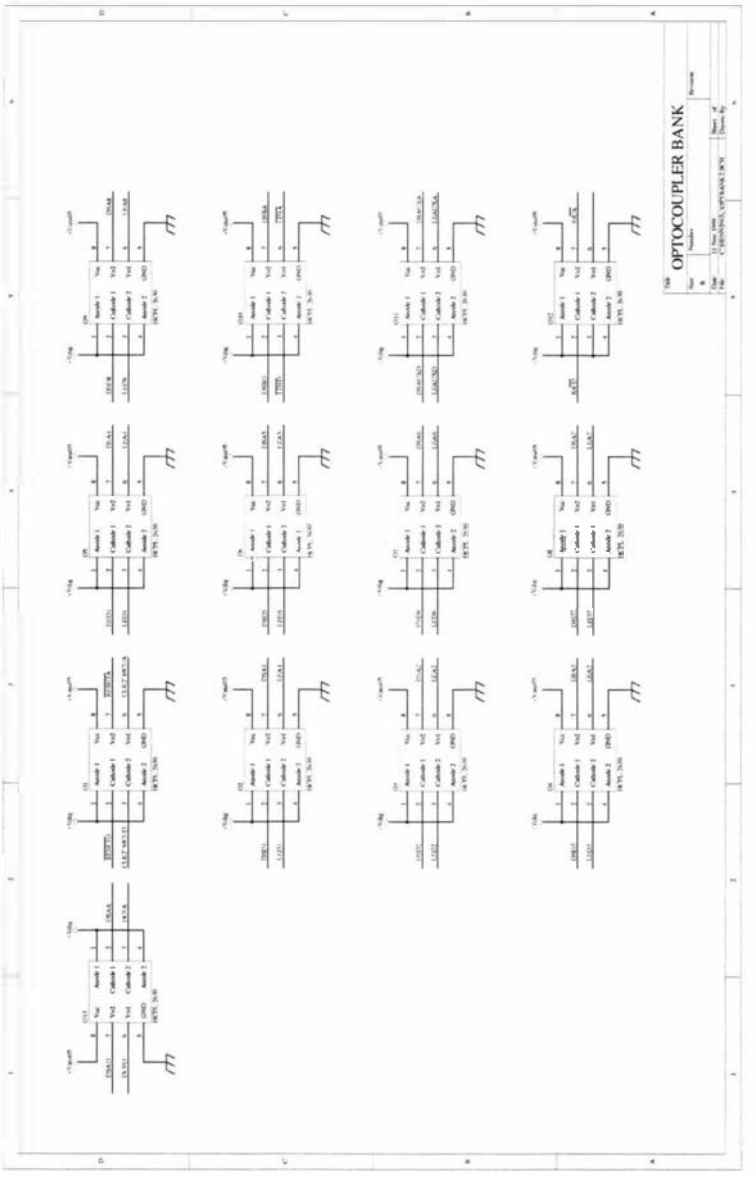

PARALLEL DATA BUFFER/ 2 * 8 BIT / DSP SIDE

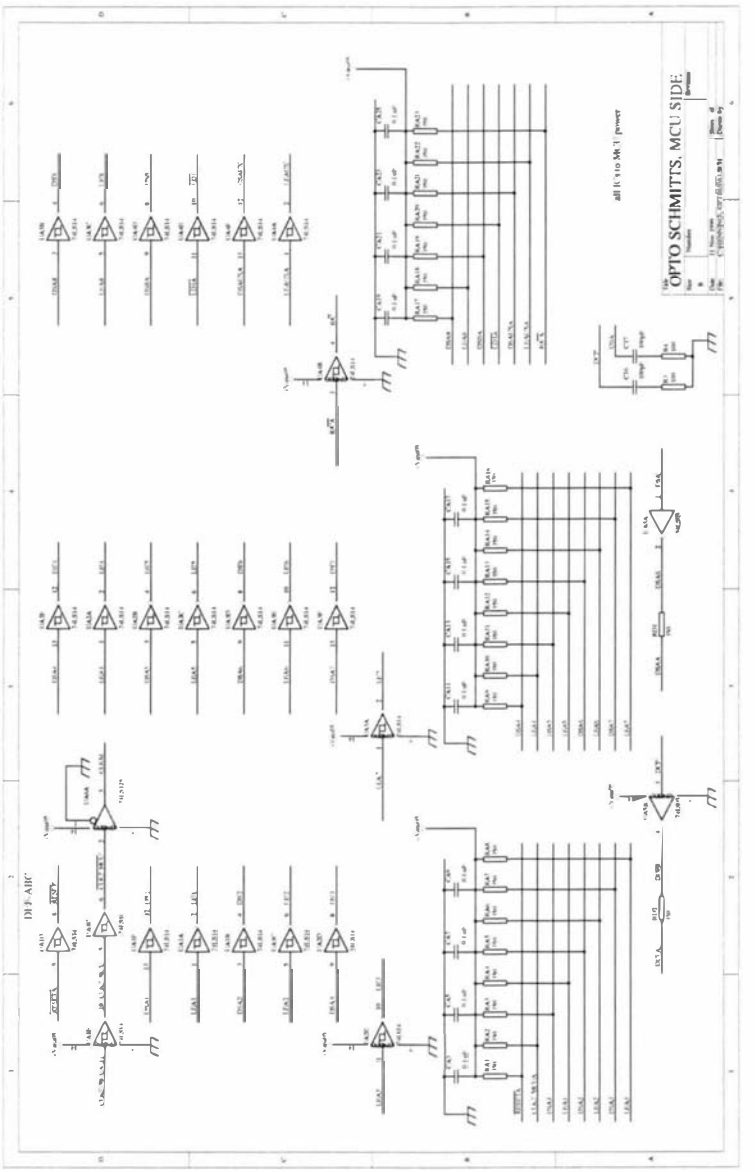

MCU BACKPLANE

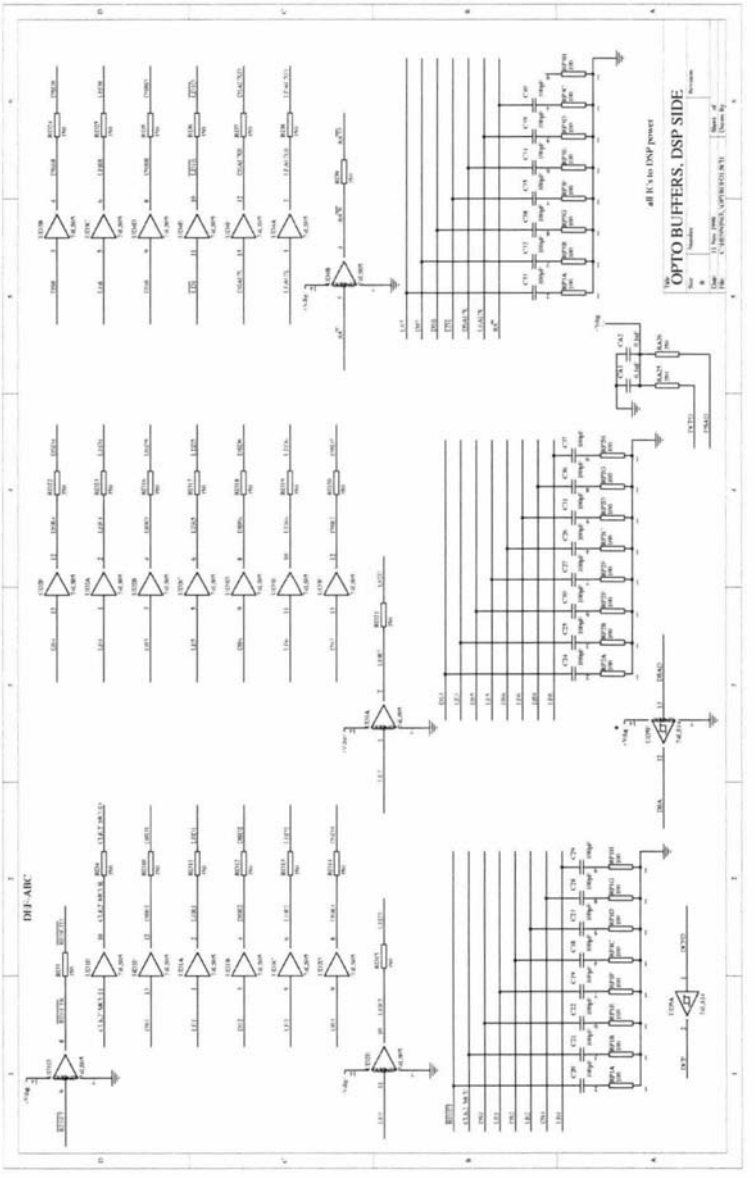

DIN Connectors

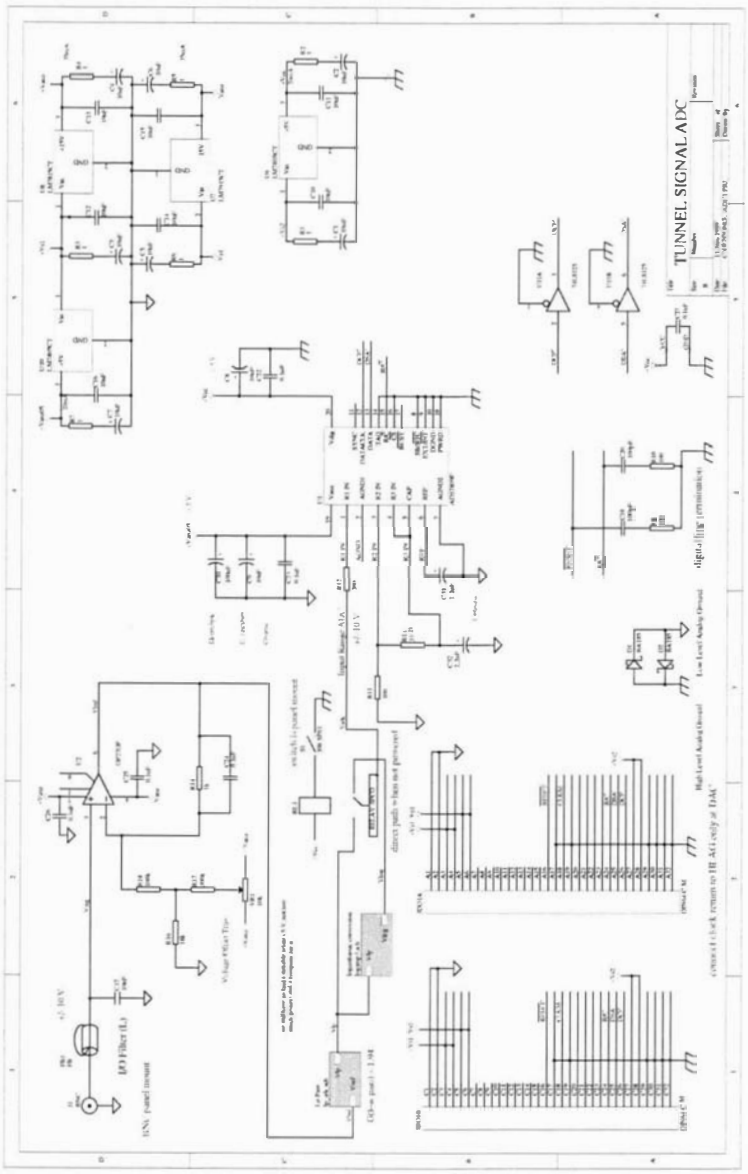

HV Connectors

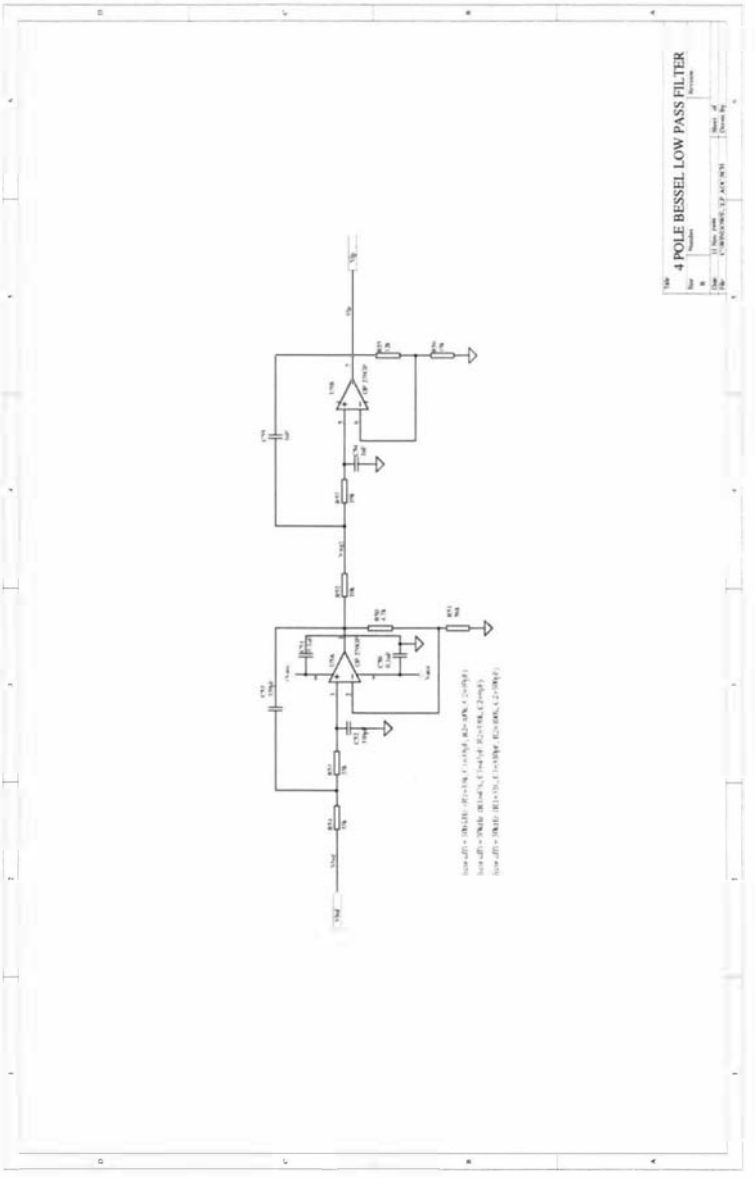

OPTOCOUPLER BOARD
decoupling capacitors for buffers, Schmitts

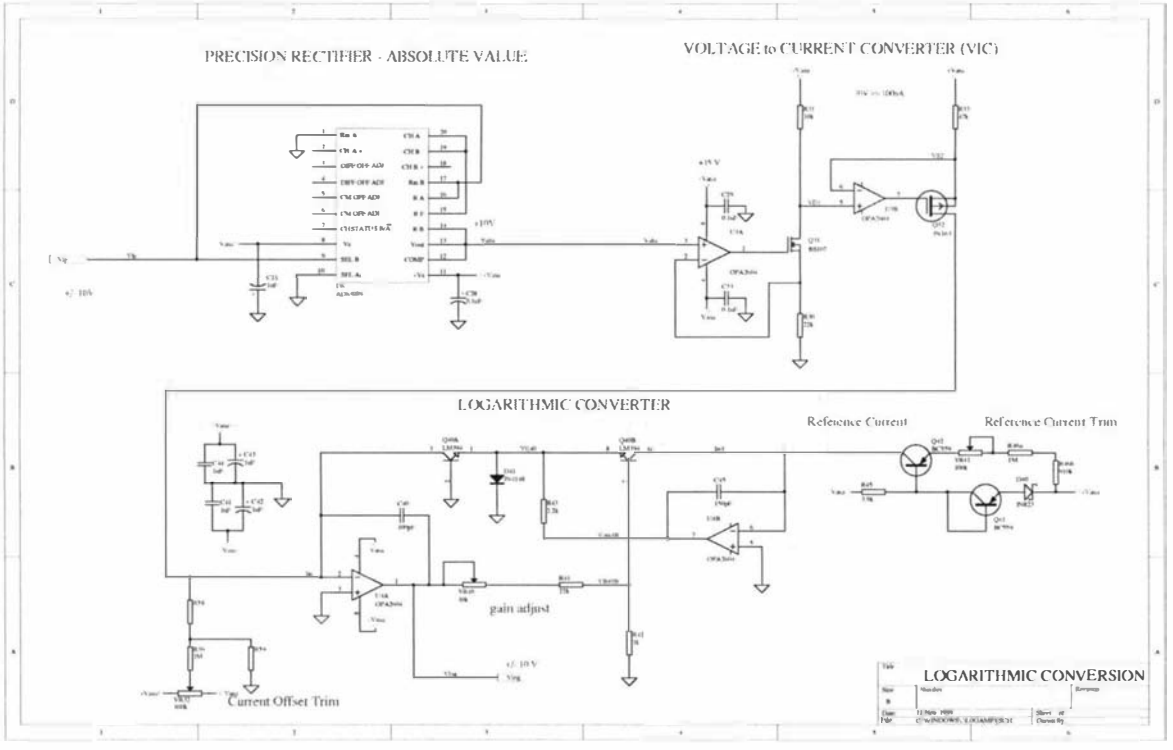

OPTOCOUPLER BANK

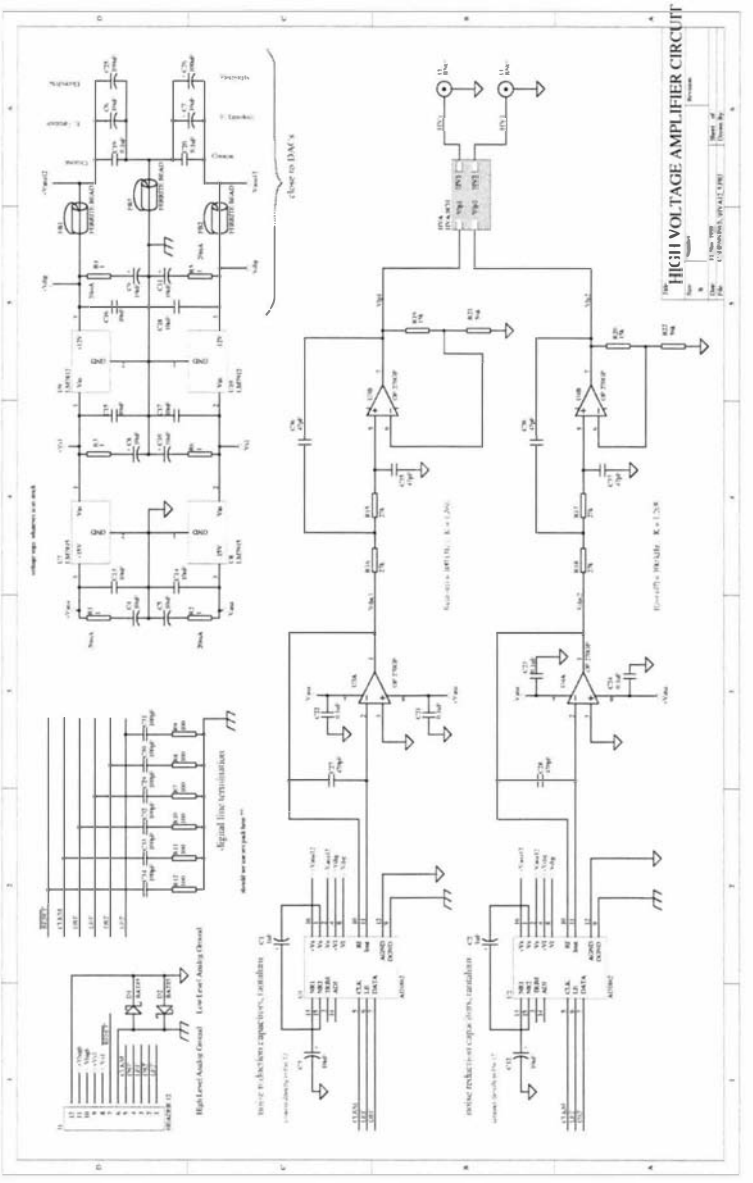

OPTO SCHMITTS, MCU SIDE

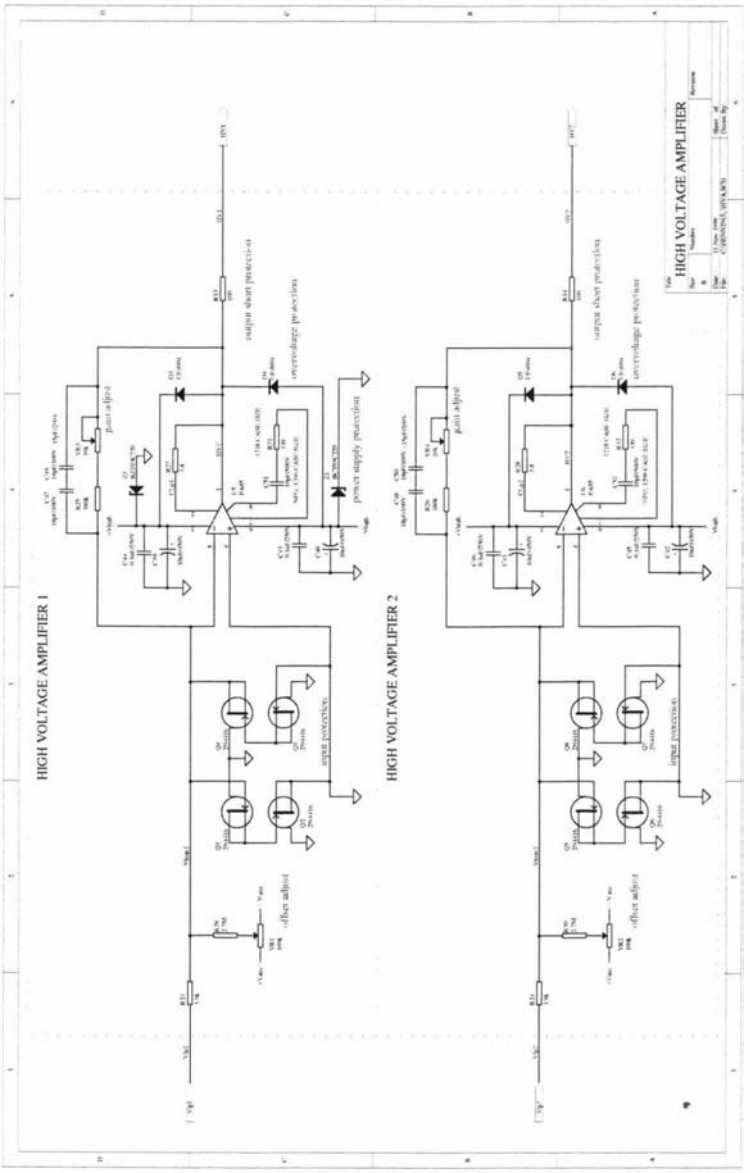

OPTO BUFFERS, DSP SIDE
all IC's to DSP power
DEF ABC

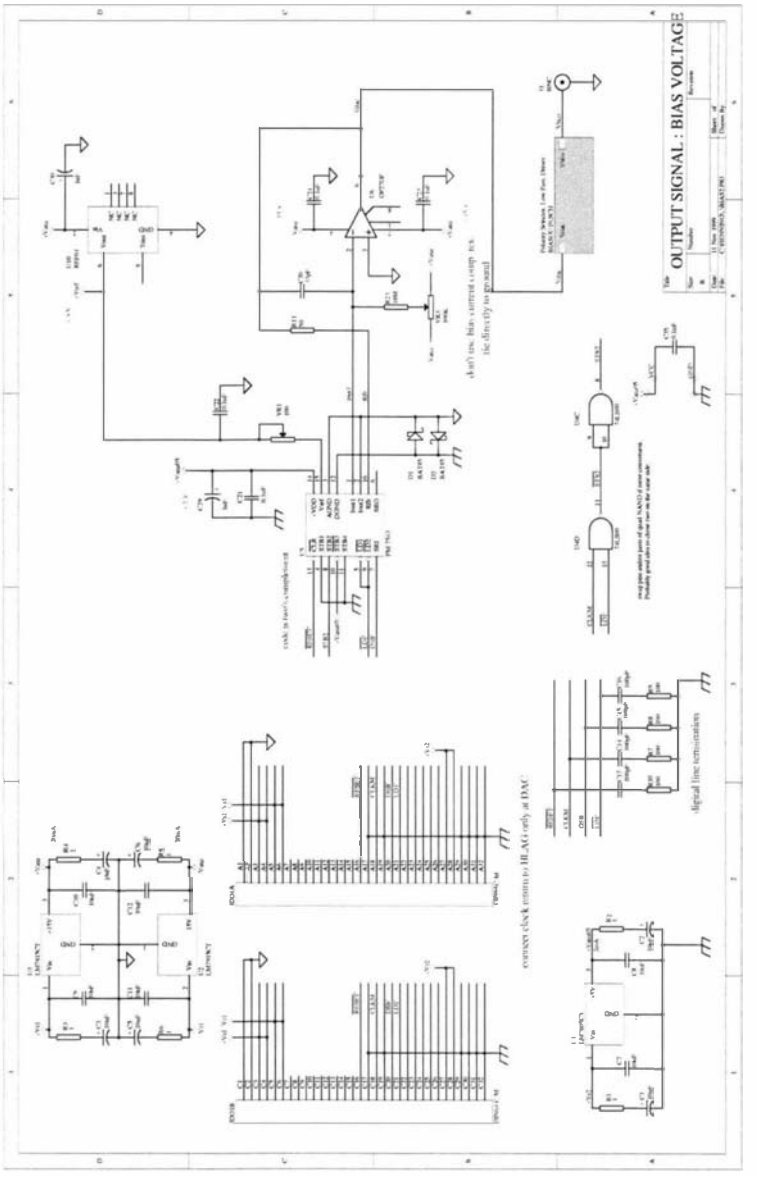

TUNNEL SIGNAL ADC

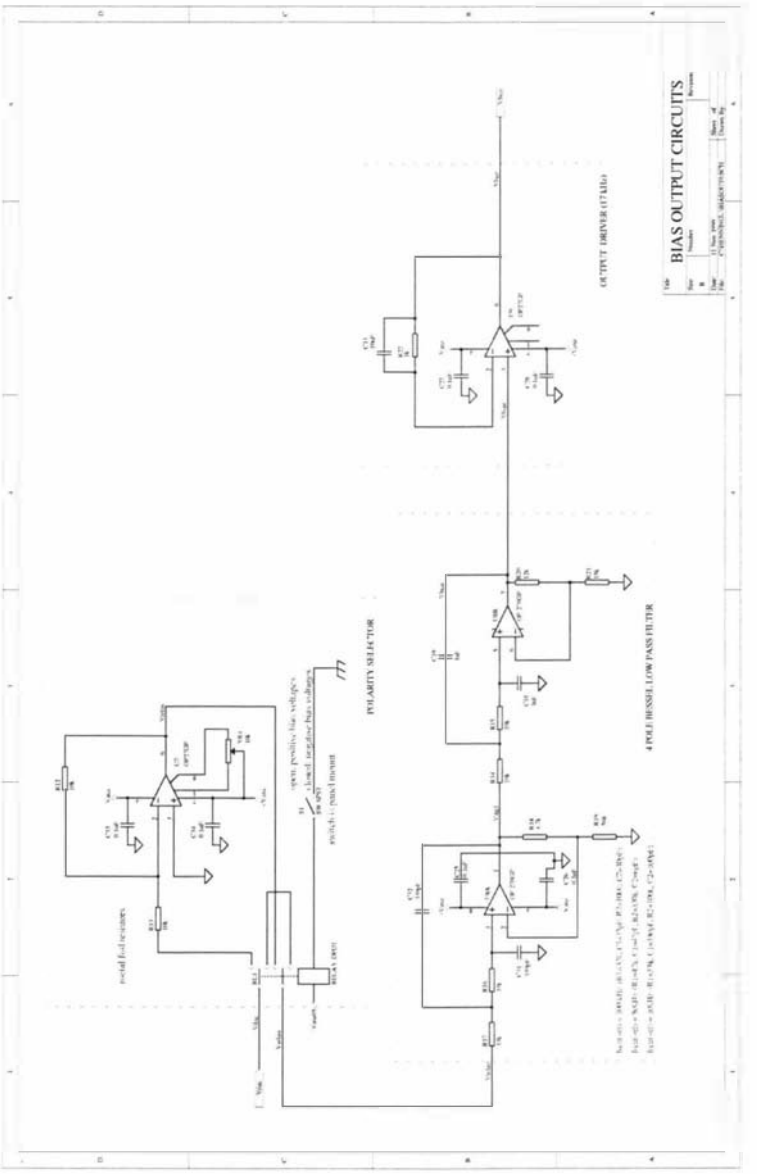

4 POLE BESSEL LOW PASS FILTER


PRECISION RECTIFIER - ABSOLUTE VALUE
VOLTAGE to CURRENT CONVERTER (VIC)
LOGARITHMIC CONVERTER
Reference Current
Reference Current Trim
gain adjust
Current Offset Trim
LOGARITHMIC CONVERSION


HIGH VOLTAGE AMPLIFIER CIRCUIT


HIGH VOLTAGE AMPLIFIER 1
HIGH VOLTAGE AMPLIFIER 2
HIGH VOLTAGE AMPLIFIER
output short protection
overvoltage protection
power supply protection
input protection
offset adjust
gain adjust


OUTPUT SIGNAL : BIAS VOLTAGE
digital line terminators


BIAS OUTPUT CIRCUITS
OUTPUT DRIVER (74HC)
POLARITY SELECTOR
4 POLE BESSEL LOW PASS FILTER

References

[1] David W. Abraham, C. C. Williams, and H. K. Wickramasinghe. Noise reduction technique for scanning tunneling microscopy. *Applied Physics Letters*, 53(16):1503–1505, 1988.

[2] M. Aguilar, P. J. Pascual, and A. Santisteban. Scanning tunneling microscope automation. *IBM Journal for Research and Development*, 30(5):525–532, 1986.

[3] Masato Aketagawa, Koji Takada, Keiko Kobayashi, Nobuhito Takeshima, Masayoshi Noro, and Yoshinori Nakayama. Length measurement using a regular crystalline lattice and a dual tunnelling unit scanning tunelling microscope in a thermo-stabilized cell. *Measurement Science and Technology*, 9:1076–1081, 1998.

[4] John Alexander. Re: problem (ii) with STM piezo tube scanner. Email to Scanning Probe Microscope List, October 30 1997. Available Email: spm@di.com.

[5] E. I. Altman, D. P. DiLella, J. Ibe, K. Lee, and R. J. Colton. Data acquisition and control system for molecules and atom-resolved tunneling spectroscopy. *Review of Scientific Instruments*, 64(5):1239–1243, 1993.

[6] Analog Devices, Norwood, Michigan 02062-9106, USA. *Amplifier Reference Manual*, 1991.

[7] Analog Devices, Norwood, Michigan 02062-9106, USA. *ADSP-21020, Ez-Lab Evaluation Board Manual*, 1992.

[8] Analog Devices, Norwood, Michigan 02062-9106, USA. *Data Converter Reference Manual*, 1992.

[9] Analog Devices, Norwood, Michigan 02062-9106, USA. *32/40-Bit IEEE Floating-Point DSP Microprocessor ADSP-21020*, 1993.

[10] Analog Devices, Norwood, Michigan 02062-9106, USA. *ADSP-21020/21010 User's Manual*, 1993.

[11] Liu Anwei, Hu Xiaotang, Liu Wenhui, and Ji Guijun. An improved control technique for the electrochemical fabrication of scanning tunneling microscopy microtips. *Review of Scientific Instruments*, 68(10):3811–3813, 1997.

[12] Apex Microtechnology Corporation, Tucson, Arizona, USA. *Apex Hybrid & IC Handbook*, 1991.

[13] Neil Ashcroft and David Mermin. *Solid State Physics*. Saunders College Publishing, Philadelphia, 1976.

[14] Karl J. Åström and Tore Hägglund. *PID Controllers: Theory, Design and Tuning*. Instrument Society of America, second edition, 1995.

[15] Karl J. Åström and Björn Wittenmark. *Computer Controlled Systems*. Information and System Science Series. Prentice-Hall, 1984. A third edition is available on the market.

[16] P. W. Atkins. *Physical Chemistry*. Oxford University Press, third edition, 1988.

[17] David M. Auslander, Yasundo Takahashi, and Michael J. Rabins. *Introducing Systems and Control*. McGraw-Hill, 1974.

[18] C. Barchesi, A. Cricenti, R. Generosi, C. Giammichele, M. Luce, and M. Rinaldi. A flexible implementation of scanning probe microscopy utilizing a multifunction system to a pc-pentium controller. *Review of Scientific Instruments*, 68(10):3799–3802, 1997.

[19] John R. Barnes. *Electronic System Design - Interference and Noise Control Techniques*. Prentice-Hall, 1987.

[20] David R. Baselt, Steven M. Clark, Michael G. Youngquist, Charles F. Spence, and John D. Baldeschwieler. Digital signal processor control of scanned probe microscopes. *Review of Scientific Instruments*, 64(7):1874–1882, 1993.

[21] Andrew Bateman and Warren Yates. *Digital Signal Processing Design*. Computer Science Press, New York, 1989.

[22] Stefan Behler, Mark Rose, Frank Ogletree, and Miquel Salmeron. Method to characterize the vibrational response of a beetle type scanning tunneling microscope. *Review of Scientific Instruments*, 68(1):124–128, 1997.

[23] Stefan Behler, Mark K. Rose, James C. Dunphy, D. Frank Ogletree, Miquel Salmeron, and Claude Chapelier. Scanning tunneling microscope with continuous flow cryostat sample cooling. *Review of Scientific Instruments*, 68(6):2479–2485, 1997.

[24] Stuart Bennett. *Real-Time Computer Control: An Introduction*. Systems and Control Engineering. Prentice Hall, 1988.

[25] K. Besocke. An easily operable scanning tunneling microscope. *Surface Science*, 181:145–153, 1987.

[26] John Billingsley. *Controlling with Computers*. McGraw-Hill, Maidenhead, England, 1989.

[27] G. Binnig and H. Rohrer. Scanning tunneling microsocpy. *IBM Journal for Research and Development*, 30(4):355–369, 1986.

[28] G. Binnig and H. Rohrer. Scanning tunneling microsocpy – from birth to adolescence. *Reviews of Modern Physics*, 56:615–625, 1987.

[29] Gerd Binnig and D. P. E. Smith. Single-tube three-dimensional scanner for scanning tunneling microscopy. *Review of Scientific Instrumentation*, 57(8), 1986.

[30] H. Birk, K. Oostveen, and C. Schöneberger. Preamplifier for electric-current noise measurements at low temperature. *Review of Scientific Instruments*, 67(8):2977–2980, 1996.

[31] Ralph E. Blake. *Basic Vibration Theory*, chapter 2. In Harris [91], fourth edition, 1996.

[32] Ilan A. Blech. Properties of materials. In Donald Christiansen, editor, *Electronics Engineer's Handbook*, chapter 9. McGraw-Hill, New York, fourth edition, 1997.

[33] H. Bourque and R. M. Leblanc. Electrochemical fabrication of scanning tunneling microscopy tips without an electronic shut-off control. *Review of Scientific Instruments*, 66(3):2695–2697, 1995.

[34] John Brignell and Neil White. *Intelligent Sensor Systems*. Sensors. Institute of Physics Publishing, Bristol and Philadelphia, second edition, 1996.

[35] I. N. Bronstein and K. A. Semendjajew. *Taschenbuch der Mathematik*. Verlag Nauka, Moskau, 23rd edition, 1987. In German.

[36] P. J. Bryant, S. Kim, and R. Yang. Technique for shaping tunneling microscope tips. *Review of Scientific Instruments*, 58(6):1115, 1987.

[37] P. A. Buksh. Very low current amplifier for computerized data acquisition systems in noisy environments. *Review of Scientific Instruments*, 63(1):846–849, 1992.

[38] Burleigh Instruments, Inc., Burleigh Park, Fishers, New York 14453-0755, U.S.A. *Inchworm Motor System*.

[39] Burr-Brown Corporation, Tucson, Arizona, USA. *Applications Handbook*, 1994.

[40] Burr-Brown Corporation, Tucson, Arizona, USA. *Burr-Brown IC Data Book–Data Conversion Products*, 1994. ADS7809 ADC.

[41] Carl Zeiss Jena GmbH, Zeiss Gruppe, 07740 Jena, Germany. *Zeiss Beetle UHV SPM*.

[42] James R. Carstens. *Electrical Sensors and Transducers*. Prentice Hall, 1993.

[43] James Carvill. *Mechanical Engineer's Data Handbook*. Butterworth Heinemann Ltd, Oxford, 1993.

[44] Julian Chen. *Introduction to Scanning Tunneling Microscopy*. Oxford Series In Optical And Imaging Sciences. Oxford University Press, 1993.

[45] C. J. Chesmond, P. A. Wilson, and M. R. Lepla. *Advanced Control System Technology*. Edward Arnold, London, 1991.

[46] S. M. Clark, D. R. Baselt, C. F. Spence, M. G. Youngquist, and J. D. Baldeschwieler. Hardware for digitally controlled scanned probe microscopes. *Review of Scientific Instruments*, 63(10):4296–4307, 1992.

[47] Phil Collins. Re: problem (ii) with STM piezo tube scanner. Email to Scanning Probe Microscope List, October 29 1997. Available Email: spm@di.com.

[48] J. H. Coombs. Properties of vacuum tunneling currents: Anomalous barrier heights. *IBM Journal for Research and Development*, 30(5):455–459, 1986.

[49] Bruce D. Craig, editor. *Handbook of Corrosion Data*. ASM International, Metals Park, OH, USA, 1989.

[50] Charles E. Crede and Jerome E. Ruzicka. *Theory of Vibration Isolation*, chapter 30. In Harris [91], fourth edition, 1996.

[51] A. Cricenti, E. Paparazzo, M. A. Scarselli, L. Moretto, and S. Selci. Preparation and characterization of tungsten tips for scanning tunneling microscopy. *Review of Scientific Instruments*, 65(5):1558–1560, 1994.

[52] R. D. Cutkosky. Versatile scan generator and data collector for scanning tunneling microscopes. *Review of Scientific Instruments*, 61(3):960–964, 1990.

[53] James W. Dally, William F. Riley, and Kenneth G. McConnell. *Instrumentation for Engineering Measurements*. John Wiley & Sons, second edition, 1993.

[54] D K Das-Gupta. *Piezoelectric and Pyroelectric Materials*, chapter 3. In Petty et al. [151], 1995.

[55] Joseph Davis, editor. *Concise Metal Engineering Data Book*. ASM International, Material Park, OH, USA, 1997.

[56] K. Dickmann, F. Demming, and J. Jersch. New etching procedure for silver scanning tunneling microscopy tips. *Review of Scientific Instruments*, 67(3):845–846, 1996.

[57] D. P. DiLella, J. H. Wandass, and R. J. Colton. Control systems for scanning tunneling microscopes with tube scanners. *Review of Scientific Instruments*, 60(6):997–1002, 1989.

[58] N. John DiNardo. *Nanoscale Characterization of Surfaces and Interfaces*. VCH Verlagsgesellschaft, Weinheim, Germany, 1994.

[59] Kai Dirscherl. Personal commmunication.

[60] Richard C. Dorf, editor. *The Engineering Handbook*. CRC Press, 1996.

[61] Klaus Dransfeld and Paul Kienle. *Physik II Elektrodynamik*. Oldenbourg Verlag, Munich, second edition, 1986. In German.

[62] U. Dürig, J. K. Gimzewski, and D. W. Pohl. Experimental observation of forces acting during scanning tunneling microscopy. *Physical Review Letters*, 57(19):2403–2406, 1986.

[63] U. Dürig, O. Züger, and D. W. Pohl. Observation of metallic adhesion using the scanning tunneling microscope. *Physical Review Letters*, 65(3):349–352, 1990.

[64] Elsevier, Oxford. *Handbook of Industrial Materials*, second edition, 1992.

[65] L. M. Eng, F. Eng, Ch. Seuret, A. Kündig, and P. Günter. Inexpensive, reliable control electronics for stick-slip motion in air and ultrahigh vacuum. *Review of Scientific Instruments*, 67(2):401–405, 1996.

[66] Engineering Plastics Ltd., Palmerston North. *Metric & Imperial O Rings*, 1996.

[67] R. Fainchtein and P. R. Zarriello. A computer-controlled technique for electrochemical STM tip fabrication. *Ultramicroscopy*, 42-44:1533–1537, 1992.

[68] J. H. Ferris, J. G. Kushmerick, J. A. Johnson, M. G. Yoshikawa Youngquist, R. B. Kessinger, H. F. Kingsbury, and P. S. Weiss. Design, operation, and housing of an ultrastable, low temperature ultrahigh vacuum scanning tunneling microscope. *Review of Scientific Instruments*, 69(7):2691–2695, 1998.

[69] D. Grant Fisher. Process control: An overview and personal perspective. *The Canadian Journal of Chemical Engineering*, 69:5–26, 1991.

[70] Mircea Fotino. Tip sharpening by normal and reverse electrochemical etching. *Review of Scientific Instruments*, 64(1):159–167, 1993.

[71] Glenn A. Fried, X. D. Wang, and K. W. Hipps. Gold-coated tungsten tips for scanning tunneling microscopy. *Review of Scientific Instruments*, 64(6):1495–1501, 1993.

[72] Bernard Friedland. *Advanced Control System Design*. Prentice Hall, London, 1996.

[73] J. Frohn, J. F. Wolf, K. Besocke, and M. Teske. Coarse tip distance adjustment and positioner for a scanning tunneling microscope. *Review of Scientific Instruments*, 60(6):1200–1201, 1989.

[74] B. Gauthier-Manuel and L. Garnier. Design for a 10 picometer magnetic actuator. *Review of Scientific Instruments*, 68(6):2486–2489, 1997.

[75] J. K. Gimzewski and R. Möller. Transition from the tunneling regime to point contact studied using scanning tunneling microscopy. *Physical Review B*, 36(2):1284–1287, 1987.

[76] W. Göhde, J. Tittel, Th. Basché, C. Bräuchle, U. C. Fischer, and H. Fuchs. A low-temperature scanning confocal and near-field optical microscope. *Review of Scientific Instruments*, 68(6):2466–2474, 1997.

[77] A. A. Gorbunov and B. Wolf. The use of silver tips in scanning tunneling microscopy. *Review of Scientific Instruments*, 64(8):2393–2394, 1993.

[78] S. Grafström, J. Kowalski, and R. Neumann. Design and detailed analysis of a scanning tunneling microscope. *Measurement Science and Technology*, 1:139–146, 1990.

[79] H. Granicher. *Piezoelectricity*, volume 13, pages 592–597. McGraw-Hill, eighth edition, 1997.

[80] B. A. Gregory. *An Introduction to Electrical Instrumentation and Measurement Systems.* MacMillan Press, Houndsmill, Basingstoke, Hampshire, second edition, 1981.

[81] Gerhard Groos. Re: problem with STM piezo tube scanner. Email to H.Klank, October 27 1997.

[82] B. Hacker, A. Hillebrand, T. Hartmann, and R. Guckenberger. Preparation and characterization of tips for scanning tunneling microscopy of biological specimens. *Ultramicroscopy*, 42-44:1514–1518, 1992.

[83] V. M. Hallmark, S. Chiang, J. F. Rabolt, J. D. Swalen, and R. J. Wilson. Observation of atomic corrugation on Au(111) by scanning tunneling microscopy. *Physical Review Letters*, 59(25):2879–2882, 1987.

[84] Claus Hamann and Michael Hietschold. *Raster-Tunnel-Mikroskopie.* Akademie Verlag, Berlin, 1991. In German.

[85] R. J. Hamers. Atomic-resolution surface spectroscopy with the scanning tunneling microscope. *Annual Reviews of Physical Chemistry*, 40:531–559, 1989.

[86] R. J. Hamers, R. M. Tromp, and J. E. Demuth. Surface electronic structure of Si(111)-(7 × 7) resolved in real space. *Physical Review Letters*, 56(18):1972–1975, 1986.

[87] Günther Hämmerlin and Karl-Heinz Hoffmann. *Numerische Mathematik.* Springer-Verlag, 1989. In German.

[88] A. Hammiche, Yu Wei, I. H. Wilson, and R. P. Webb. The surrey STM: Construction, development, and evaluation of a scanning tunneling microscope. *Review of Scientific Instruments*, 62(12):3010–3021, 1991.

[89] Charles A. Harper. *Handbook of Plastics, Elastomers and Composites.* McGraw-Hill, second edition, 1992.

[90] Charles A. Harper and Martin B. Miller. *Electronic Packaging, Microelectronics and Interconnection Dictionary.* McGraw-Hill, New York, 1993.

[91] Cyril Harris, editor. *Shock and Vibration Handbook.* McGraw-Hill, New York, fourth edition, 1996.

[92] R. Heer, C. Eder, J. Smoliner, and E. Gornik. Floating electrometer for scanning tunneling microcope applications in the femtoampere range. *Review of Scientific Instruments*, 68(12):4488–4491, 1997.

[93] P. Heuell, S. Cuzdi, M. A. Kulakov, and B. Bullemer. An ultra-high resolution control unit for a scanning tunnelling microscope. *Thin Solid Films*, 264:217–222, 1995.

[94] P. Heuell, M. A. Kulakov, and B. Bullemer. An adaptive scan generator for a scanning tunneling microscope. *Review of Scientific Instruments*, 65(1):89–92, 1994.

[95] Peter Heuell. *Entwicklung eines prozeßrechnergestützten Raster-Tunnelmikroskops und Morphologie-Bestimmung des Halbleiters Siliziumkarbid.* PhD thesis, Universität der Bundeswehr, München, 1994. In German.

[96] Melissa Hines. Re: problem with STM piezo tube scanner. Email to H.Klank, October 23 1997.

[97] K. W. Hipps. Re: problem with STM piezo tube scanner. Email to H.Klank, October 24 1997.

[98] Lisa A. Hockett and Stephen E. Creager. A convenient method for removing surface oxides from tungsten STM tips. *Review of Scientific Instruments*, 64(1):263–264, 1993.

[99] A. J. Hoeven, E. J. van Loenen, P. J. G. M. van Hooft, and K. Oostveen. A multiprocessor data acquisition and analysis system for scanning tunneling microscopy. *Review of Scientific Instruments*, 61(6):1668–1673, 1990.

[100] Paul Horowitz and Winfield Hill. *The Art of Electronics.* Cambridge University Press, Cambridge, second edition, 1989.

[101] Paul Horowitz and Winfield Hill. *Table 5.2: VCVS Low-Pass Filters*, chapter 5, page 274. In Horowitz [100], second edition, 1989.

[102] Harry Hoster. Personal commmunication.

[103] Steven M. Hues, Charles F. Draper, Kenneth P. Lee, and Richard J. Colton. Effect of PZT and PMN actuator hysteresis and creep on nanoindentation measurements using force microscopy. *Review of Scientific Instruments*, 65(5):1561–1565, 1994.

[104] Mark Hunter. The happy world of tips. Master's thesis, Massey University, Palmerston North, NZ, to be submitted 2000.

[105] J.P. Ibe, P. P. Bey Jr., S. L. Brandow, R. A. Brizzolara, N. A. Burnham, D. P. DiLella, K. P. Lee, C. R. K. Marrian, and R. J. Colton. On the electrochemical etching of tips for scanning tunneling microscopy. *Journal of Vacuum Science and Technology A*, 8(4):3570–3575, 1990.

[106] Sang il Park and C. F. Quate. Digital filtering of scanning tunneling microscope images. *Journal of Applied Physics*, 62(1):312–314, 1987.

[107] J. Jersch, F. Demming, I. Fedotov, and K. Dickmann. Wide-band low-noise tunnel current measurements in laser assisted experiment. *Review of Scientific Instruments*, 70(7):3173–3176, 1999.

[108] Jan F. Jørgensen. *Scanning Probe Image Processor for MS Windows 95/98/NT Version 0.577.* Danish Institute for Fundamental Metrology, August 1998.

[109] Hiroshi Kaizuka. Application of capacitor insertion method to scanning tunneling microscopy. *Review of Scientific Instruments*, 60(10):3119–3122, 1989.

[110] John H. Karl. *An Introduction to Digital Signal Processing.* Academic Press, San Diego, 1989.

[111] G Kaye and T Laby. *Tables of Physical and Chemical Constants.* Longman Group Limited, Burnt Mill, Harlow, England, sixteenth edition, 1995.

[112] Keithley Instruments, 28775 Aurora Rd., Cleveland, OH 44139, USA. *Low Level Measurements*, fourth edition, 1993.

[113] Brian W. Kernighan and Dennis M. Ritchie. *The C Programming Language.* Prentice Hall, second edition, 1988.

[114] Ki Hyun Kim, Sang-Kee Eah, Byoungho Lee, Chang Ho, and Cho Wonho Jhe. A novel digital feedback scheme of shear-force control in near-field scanning optical microscopy. *Review of Scientific Instruments*, 68(7):2783–2786, 1997.

[115] Klaas B. Klaassen. *Electronic Measurement and Instrumentation.* Cambridge University Press, 1996.

[116] M. Klein and G. Schwitzgebel. An improved lamellae drop-off technique for sharp tip preparation in scanning tunneling microcopy. *Review of Scientific Instruments*, 68(8):3099–3103, 1997.

[117] S. Kleindiek and K. H. Herrmann. A miniaturized scanning tunneling microscope with large operation range. *Review of Scientific Instruments*, 64(3):692–693, 1993.

[118] Sh. Kogan. *Electronic Noise and Fluctuations in Solids.* Cambridge University Press, 1996.

[119] Jacqueline Kroschwitz and Mary Howe-Grant, editors. *Kirk-Othmer Encyclopedia of Chemical Technology*, volume 4. John Wiley & Sons, New York, fourth edition, 1992.

[120] Horst Kuchling. *Taschenbuch der Physik.* Harri Deutsch, Thun, 1986. In German.

[121] Y. Kuk and P. J. Silverman. Scanning tunneling microscope instrumentation. *Review of Scientific Instruments*, 60(2):165–180, 1989.

[122] Erik Lægsgaard. Problem (ii) with STM piezo tube scanner. Email to H.Klank, October 29 1997.

[123] Graham J. Legett. *Scanning Tunnelling Microscopy and Atomic Force Microscopy*, chapter 9, pages 393–449. In Vickerman [189], 1997.

[124] J R Leigh. *Applied Control Theory.* Peter Peregrinus, Stevenage, 1982.

[125] J R Leigh. *Control Theory - A Guided Tour.* Peter Peregrinus, Stevenage, 1992.

[126] L. Libioulle, Y. Houbion, and J.-M. Gilles. Very sharp gold and platinum tips to modify gold surfaces in scanning tunneling microscopy. *Journal of Vacuum Science and Technology B*, 13(3):1325–1331, 1995.

[127] L. Libioulle, Y. Houbion, and J.-M. Gilles. Very sharp platinum tips for scanning tunneling microscopy. *Review of Scientific Instruments*, 66(1):97–100, 1995.

[128] J. W. Lyding, S. Skala, J. S. Hubacek, R. Brockenbrough, and G. Gammie. Variable-temperature scanning tunneling microscope. *Review of Scientific Instruments*, 59(9):1897–1902, 1988.

[129] K. Maeda, S. Sugita, H. Kurita, M. Uota, S. Uchida, M. Hinomaru, and Y. Mera. Spatial variation of $1/f$ current noise in scanning tunneling microscopes. *Journal of Vacuum Science and Technology B*, 12(3):2140–2143, 1994.

[130] G. Mariotto, M. D'Angelo, and I. V. Shvets. Dynamic behavior of a piezowalker, inertial and frictional configurations. *Review of Scientific Instruments*, 70(9):3651–3655, 1999.

[131] Othmar Marti. STM piezo drift (reply). Email to Scanning Probe Microscope List, July 1997. Available Email: spm-digest@di.com.

[132] Allan J. Melmed. The art and science and other aspects of making sharp tips. *Journal of Vacuum Science and Technology B*, 9(2):601–608, 1991.

[133] Gerhard Meyer. A simple low-temperature ultrahigh-vacuum scanning tunneling microscope capable of atomic manipulation. *Review of Scientific Instruments*, 67(8):2960–2965, 1996.

[134] B. A. Morgan and G. W. Stupian. Digital feedback control loops for scanning tunneling microscopes. *Review of Scientific Instruments*, 62(12):3112–3113, 1991.

[135] Ralph Morrison. *Grounding and Shielding Techniques in Instrumentation.* Wiley, 1986.

[136] P. T. Moseley and A. J. Crocker. *Sensor Materials.* Sensors. Institute of Physics Publishing, Bristol and Philadelphia, 1996.

[137] P. T. Moseley and A. J. Crocker. *Table 5.5: Piezoelectric Parameters for PZT*, chapter 5, page 131. In *Sensors* [136], 1996.

[138] F. Müller, A.-D. Müller, O. Meissner, A. Heilmann, and M. Hietschold. Enhanced local surface conductivity measurements by scanning tunneling microscopy. *Review of Scientific Instruments*, 68(8):3104–3107, 1997.

[139] Raúl C. Munoz, Paolo Villagra, Germán Kremer, Luis Moraga, and Guillermo Vidal. Control circuit for a scanning tunneling microscope. *Review of Scientific Instruments*, 69(9):3259–3267, 1998.

[140] L. A. Nagahara, T. Thundat, and S. M. Lindsay. Preparation and characterization of STM tips for electrochemical studies. *Review of Scientific Instruments*, 60(10):3128–3130, 1989.

[141] C. Y. Nakakura, V. M. Phanse, G. Zheng, G. Bannon, E. I. Altman, and K. P. Lee. A high-speed variable-temperature ultrahigh vacuum scanning tunneling microscope. *Review of Scientific Instruments*, 69(9):3251–3258, 1998.

[142] A. J. Nam, A. Teren, T. A. Lusby, and A. J. Melmed. Benign making of sharp tips for STM and FIM: Pt, Ir, Au, Pd, and Rh. *Journal of Vacuum Science and Technology B*, 13(4):1556–1559, 1995.

[143] Peter G. Nelson. Technical background. Technical report, Technical Manufacturing Corporation, 15 Centennial Drive, Peabody, Massachusetts 01960, USA, 1999. Internet: www.techmfg.com.

[144] L. Olesen, M. Brandbyge, M. R. Sørensen, K. W. Jacobsen, E. Lægsgaard, I. Stensgaard, and F. Besenbacher. Apparent barrier height in scanning tunneling microscopy revisited. *Physical Review Letters*, 76(9):1485–1488, 1996.

[145] H. Olin. Design of a scanning probe microcope. *Measurement Science and Technology*, 5:976–984, 1994.

[146] A. I. Oliva, A. Romero G., J. L. Peña, E. Anguiano, and M. Aguilar. Electrochemical preparation of tungsten tips for a scanning tunneling microscope. *Review of Scientific Instruments*, 67(5):1917–1921, 1996.

[147] A. I. Oliva, Victor Sosa, R. de Coss, Raquel Sosa, N. López Salazar, and J. L. Peña. Vibration isolation analysis for a scanning tunneling microscope. *Review of Scientific Instruments*, 63(6):3326–3329, 1992.

[148] J. J. Paggel and M Förster. An artefact in scanning tunneling microscopy images of highly oriented pyrolytic graphite. *Applied Physics A*, 59:419–425, 1994.

[149] Bruno Paillard, Ran Tang, and Paul Rowntree. Digital linearization and cancellation of capacitive coupling for a scanning tunneling microscope. *Review of Scientific Instruments*, 69(4):1770–1780, 1998.

[150] Sang-Il Park and C. F. Quate. Tunneling microscopy of graphite in air. *Applied Physics Letters*, 48(2):112–114, 1986.

[151] Michael Petty, Martin Bryce, and David Bloor, editors. *An Introduction to Molecular Electronics*. Oxford University Press, 1995.

[152] Philips, Eindhoven, The Netherlands. *Piezoelectric Ceramics, Speciality Ferrites*, 1997. Data Handbook MA03.

[153] Physik Instrumente, Waldbronn, Germany. *Tutorial on Piezoelectric Actuators*. Internet: http://www.physikinstrumente.com/tutorial/index.html.

[154] R. Piner and R. Reifenberger. Computer control of the tunnel barrier width for the scanning tunneling microscope. *Review of Scientific Instruments*, 60(10):3123–3127, 1989.

[155] D. W. Pohl and R. Möller. "Tracking" tunneling microscopy. *Review of Scientific Instruments*, 59(6):840–842, 1988.

[156] Dieter W. Pohl. Some design criteria in STM. *IBM Journal of Research and Development*, 30(4):417–427, 1986.

[157] T. L. Porter. Scanning tunneling microscope data acquisition and control in Visual Basic. *Review of Scientific Instruments*, 64(12):3530–3533, 1993.

[158] Precision Rubber Products (Canada) Ltd., NZ Distributor: Precision Bearings Limited, 2 Botha Rd, Penrose, Auckland. *Green O-rings with DuPont Viton*, 1996?

[159] J. P. Rabe. *Scanning Tunnelling Microscopy*, chapter 12, pages 261–278. In Petty et al. [151], 1995.

[160] Geetha Ramaswamy. Piezo troubleshooting. Email to H.Klank, October 30 1997.

[161] Kenneth L. Ratzlaff. *Introduction to Computer-Assisted Experimentation.* Wiley Interscience, 1987.

[162] Francis H. Raven. *System Compensation*, chapter 155. In Dorf [60], 1996.

[163] R. S. Robinson, T. H. Kimsey, and R. Kimsey. A digital integrator and scan generator coupled with dynamic scanning for scanning tunneling microscopy. *Review of Scientific Instruments*, 62(7):1772–1775, 1991.

[164] Z. Y. Rong, A. Chang, and E. L. Wolf. A flexible implementation of scanning tunneling spectroscopy utilizing c-language on the personal computer. *Review of Scientific Instruments*, 63(7):3646–3651, 1992.

[165] RS Components, Auckland, New Zealand. *RS Catalogue*, 1998.

[166] Juan José Sáenz and Ricardo García. Near field emission scanning tunneling microscopy. *Applied Physics Letters*, 65(23):3022–3024, 1994.

[167] J. Scheuring, W. Clauss, and D. P. Kern. Self-optimizing and adaptive digital signal processor based algorithms in scanning tunneling microscopy. *Review of Scientific Instruments*, 69(12):4191–4194, 1998.

[168] M. Schmid and P. Varga. Analysis of vibration-isolating systems for scanning tunnelig microscopes. *Ultramicroscopy*, 42-44:1610–1615, 1992.

[169] Peter H. Schroer and Jordan Becker. Computer automation for scanning tunneling microscopy. *IBM Journal for Research and Development*, 30(5):543–552, 1986.

[170] Adel S. Sedra and Kenneth C. Smith. *Microelectronic Circuits.* Oxford Series in Electrical Engineering. Oxford University Press, fourth edition, 1998.

[171] Robert Simpson. *Introductory Electronics For Scientists And Engineers.* Allyn and Bacon, Boston, second edition, 1987.

[172] Richard Smith. Re: problem with STM piezo tube scanner. Email to H.Klank, October 23 1997.

[173] Walter F. Smith. Re: problem with STM piezo tube scanner. Email to H.Klank, October 24 1997.

[174] J. M. Soler, A. M. Baro, and N. García. Interatomic forces in scanning tunneling microscopy: Giant corrugations of the graphite surface. *Physical Review Letters*, 57(4):444–447, 1986.

[175] Staveley Sensors Inc, 91 Prestige Park Circle, East Hartford, CT 06108, USA. *EBL Piezoceramic Tube Characteristics*, 1994.

[176] William F. Stokey. *Vibration of Systems Having Distributed Mass and Elasticity*, chapter 7. In Harris [91], fourth edition, 1996.

[177] E. Stoll and O. Marti. Restoration of scanning-tunneling-microscope data blurred by limited resolution, and hampered by $1/f$ noise. *Surface Science*, 181:222–229, 1987.

[178] J Tapson and J Greene. The resonant behavior of segmented piezoceramic tubes. *Review of Scientific Instruments*, 68(7):2797–2799, 1997.

[179] M. E. Taylor. Dynamics of piezoelectric tube scanners for scanning probe microscopy. *Review of Scientific Instruments*, 64(1):154–158, 1993.

[180] Technical Manufacturing Corporation, 15 Centennial Drive, Peabody, Massachusetts 01960, USA. *Micro-g Vibration Isolation Tables.*

[181] J. Tersoff and D. R. Hamann. Theory and application for the scanning tunneling microscope. *Physical Review Letters*, 50(25):1998–2001, 1983.

[182] J. Tersoff and N. D. Lang. Tip-dependent corrugation of graphite in scanning tunneling microscopy. *Physical Review Letters*, 65(9):1132–1135, 1990.

[183] T. Tiedje, J. Varon, H. Deckman, and J. Stokes. Tip contamination effects in ambient pressure scanning tunneling microscopy imaging of graphite. *Journal of Vacuum Science and Technology A*, 6(2):372–375, 1988.

[184] Anthony Truscott. Re: problem (ii) with STM piezo tube scanner. Email to H.Klank, October 29 1997.

[185] Kenji Uchino. *Piezoelectric Actuators and Ultrasonic Motors.* Electronic Materials: Science and Technology. Kluwer Academic Publishers, Boston, 1997.

[186] Vance J. VanDoren. The Smith predictor: Process engineer's crystal ball. *Control Engineering*, pages 59–60, July 1996.

[187] Vance J. VanDoren. Basics of proportional-integral-derivative control. *Control Engineering*, pages 97–102, May 1998.

[188] Vance J. VanDoren. Ziegler-Nichols methods facilitate loop tuning. *Control Engineering*, pages 78–80, September 1998.

[189] John C. Vickerman, editor. *Surface Analysis – The Principal Techniques.* Wiley, 1997.

[190] S. Vieira. The behavior and calibration of some piezoelectric ceramics used in the STM. *IBM Journal for Research and Development*, 30(5):553–555, 1986.

[191] M. Wilms, M. Kruft, G. Bermes, and K. Wandelt. A new and sophisticated electrochemical scanning tunnelling microscope for the investigation of potentiodynamic processes. *Review of Scientific Instruments*, 70(9):3641–3650, 1999.

[192] M. Wilms, M. Schmidt, G. Bermes, and K. Wandelt. New and versatile ultrahigh vacuum scanning tunneling microscope for film growth experiments. *Review of Scientific Instruments*, 69(7):2696–2703, 1998.

[193] J. Wintterlin, J. Wiechers, H. Brune, T. Gritsch, H. Höfer, and R. J. Behm. Atomic-resolution imaging of closed-packed metal surfaces by scanning tunneling microscopy. *Physical Review Letters*, 62(1):59–62, 1989.

[194] T. M. H. Wong and M. E. Welland. A digital control sytem for scanning tunnelling microscopy and atomic force microscopy. *Measurement Science and Technology*, 4:270–280, 1993.

[195] Hugh Young. *University Physics.* Addison-Wesley Publishing Company, Reading, Massachusetts, eighth edition, 1992.

[196] Udo Zölzer. *Digital Audio Signal Processing.* John Wiley & Sons, 1997.