UNIVERSITY OF SOUTHAMPTON

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING

Communications, Signal Processing and Control

Control of Atomic Force Microscopes

by

Umar A. A. Khan

Thesis for the degree of Doctor of Philosophy

November 2014

UNIVERSITY OF SOUTHAMPTON

ABSTRACT

FACULTY OF PHYSICAL SCIENCES AND ENGINEERING
Communications, Signal Processing and Control

Doctor of Philosophy

CONTROL OF ATOMIC FORCE MICROSCOPES

by Umar A. A. Khan

Atomic force microscopes or AFMs are instruments which use a mechanical probe to scan a sample and estimate surface topography with nanometer accuracy. The term *atomic force* originates from the fact that the imaging process relies upon the existence of the inter-atomic interaction force between the mechanical probe and sample surface.

These instruments have established themselves as a vital cutting edge tool for investigation of matter at the nanometer scale. Their widespread usage is due not only to their superior resolution but also because they can operate in any medium namely air, liquid and vacuum. Another major advantage is that, unlike their predecessor instruments AFMs do not require their samples to be conductive.This fact alone has enabled in situ imaging of biological samples with unprecedented resolution and without sample alteration. Other instruments like scanning electron microscopes (SEMs) can also view biological samples, however they require the samples to be prepared and dried. While some sample structure may be preserved, AFMs have no such limitation.

Despite the fact that AFMs offer all these advantages, the usage of a mechanical probe for image generation causes them to be inherently reliant upon a feedback control loop. This is because, the probe motion must be controlled in a suitable manner to avoid letting its motion dynamics distort the sample image. In addition, since the mechanical probe must be sequentially moved over the sample point by point, the imaging times are long and range from a few seconds to in excess of ten minutes.

Given that feedback control is an integral part of AFM operation, the end users are forced to manually tune Proportaional-Integral (PI) controllers which are used in most commercial AFMs. Since the vast majority of scientists using AFMs do not necessarily possess a knowledge of feedback control, they do this tuning though a manual trial and error procedure which consumes valuable research time. Although the control systems community has taken considerable interest in AFM control, the methods suggested often require high order controllers and are tested for a specific experimental set up.

The primary objective of this research is therefore to develop a novel automated controller synthesis mechanism which has the potential of being used in a diverse range of AFM setups. The method of choice for this research is Multiple Model Adaptive Control (MMAC). The motivation for this decision as well as experimental verification is provided in detail in this thesis. Given the wide commercial usage of PI controllers, the same are used as a starting point for this work. The applicability of the method suggested is however by no means restricted to them, and in the future can be extended to incorporate more sophisticated controllers, for instance robust controllers.

The second objective of this research is to investigate two novel methods which have the potential of substantially reducing the AFM imaging time. The first one suggests coarser scan trajectories to save time, and then estimates the sample image using a relatively new signal processing method called *Compressive Sensing*. The second method suggested uses the AFM's mechanical probe in a novel manner that can also substantially reduce imaging time.

Declaration of Authorship

I, Umar A. A. Khan , declare that the thesis entitled *Control of Atomic Force Microscopes* and the work presented in the thesis are both my own, and have been generated by me as the result of my own original research. I confirm that:

- this work was done wholly or mainly while in candidature for a research degree at this University;
- where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated;
- where I have consulted the published work of others, this is always clearly attributed;
- where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work;
- I have acknowledged all main sources of help;
- where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself;
- parts of this work have been published as: Khan et al. (2013a), Khan et al. (2013c), Khan et al. (2013b)

Signed:.....

Date:.....

Contents

Nomenclature

V_x	X piezo voltage
V_y	Y piezo voltage
V_z	Z piezo voltage
v_T	Normalised PSD voltage for transverse bending
v_L	Normalised PSD voltage for lateral bending
v_A	Quadrant A PSD voltage
v_B	Quadrant B PSD voltage
v_C	Quadrant C PSD voltage
v_D	Quadrant D PSD voltage
h	Controller signal
v_{SP}	Set-point voltage
A_{SP}	Amplitude set-point
A_o	Cantilever free vibration amplitude
f_D	Dither piezo drive signal
d_{TM}	Tapping mode deflection signal
A	Amplitude signal
H	Hammaker constant
R	Tip radius
d	Tip sample separation
a_o	Inter-atomic separation
E_{tip}	Tip elastic modulus
E_{sample}	Sample elastic modulus
ν_{tip}	Tip Poisson ratio
ν_{sample}	Sample Poisson ratio
E^*	Effective modulus of elasticity
F_{ts}	Tip-sample interaction
F_{nc}	Non-conservative component of the tip-sample interaction force
F_b	Force acting on cantilever beam due to base movement
F	Total force acting on cantilever beam
F_i	Total force corresponding to the i^{th} eigenmode
ζ	Sample viscosity
X_w	W coordinate frame X axis

Y_w	W coordinate frame Y axis
X_u	U coordinate frame X axis
Y_u	U coordinate frame Y axis
q	Tip deflection in coordinate frame U
Z	Z piezo height in coordinate frame W
y	Dither piezo height in coordinate frame W
Y	Total cantilever base height in coordinate frame W
Z_i	Z piezo height for the i^{th} eigenmode
y_i	Dither piezo height for the i^{th} eigenmode
Y_i	Total cantilever base height for the i^{th} eigenmode
q_i	Tip deflection corresponding to the i^{th} eigenmode
m_i	Cantilever mass corresponding to the i^{th} eigenmode
k_i	Cantilever stiffness corresponding to the i^{th} eigenmode
c_i	Damping corresponding to the i^{th} eigenmode
ω_i	Resonance frequency corresponding to i^{th} eigenmode
T	Sample topography height
$\hat{T}$	Estimated sample topography height
k_Z	Calibration constant for Z piezo voltage
A_{base}	Cantilever base vibration amplitude
ω	Cantilever base vibration frequency
E	Cantilever beam modulus of elasticity
I	Cantilever beam moment of inertia
L	Cantilever beam length
w	Beam displacement in frame W
u	Beam displacement in frame U
ρ_c	Cantilever beam mass per unit length
γ	Cantilever beam hydrodynamic damping coefficient
ϕ_i	i^{th} eigenmode
$\bar{\phi}_i$	Normalized i^{th} eigenmode
m_{tip}	Tip mass
$\bar{m}_{tip}$	Normalized tip mass
K	Cantilever transfer function gain
Q	Cantilever transfer function quality factor
ω_n	Cantilever transfer function resonance frequency
ω_p	Cantilever frequency response peak frequency
ω_L	Cantilever frequency response lower half power frequency
ω_H	Cantilever frequency response upper half power frequency
A_{max}	Cantilever frequency response peak amplitude
M_p	Peak amplitude of normalised second order transfer function
L_{scan}	Scan length in simulated AFM experiment
T_{scan}	Scan duration in simulated AFM experiment

T_h	Sample feature height in simulated AFM experiment
K_p	PI controller proportional gain
K_i	PI controller integrator gain
e	AFM control loop error for simulated and experimental results
$\mathcal{S}$	Signal space
$\mathcal{T}_t$	Truncation operator
$\mathcal{R}_{\sigma,t}$	Restriction operator
$\mathcal{U}$	Plant input space
$\mathcal{V}$	Plant output space
$\mathcal{U}_e$	Extended plant input space
$\mathcal{V}_e$	Extended plant output space
u_o, y_o	Plant input and output disturbances
u_1, y_1	True plant input and output
u_2, y_2	Measured input and output
$\mathcal{P}$	Plant
$\mathcal{C}$	Controller
$\mathcal{P}_{p^*}$	True plant
$\mathcal{P}_p$	p^{th} plant in the plant set
A_p, B_p, C_p, D_p	State space matrices corresponding to p^{th} plant in plant set
Σ	Kalman estimator covariance matrix
r_p	Residual corresponding to the p^{th} plant
$\hat{x}$	System state as estimated by the Kalman estimator
$\widetilde{y}$	System output as estimated by the Kalman estimator
Δ	EMMSAC system delay
q_f	Switching signal
D_j^i	Data subset corresponding to the i^{th} sample and j^{th} trail controller
G_j^i	Least residual plant corresponding to the i^{th} sample and j^{th} trail controller
C_k	Best controller corresponding to the k^{th} sample
χ_j^i	Performance metric for the i^{th} sample using the j^{th} trial controller
$\mathbb{D}$	Complete data set containing all data subsets
$\mathbb{G}$	Complete plant set containing all least residual plants
$\mathbb{R}$	Residual matrix
$\mathbb{G}$	Mean residual matrix
Φ_p	Mapping from data subset to least residual plant
Φ_c	Mapping from all data sub sets in a cluster to the best performance controller
Φ_q	Mapping from the true plant's measured input and output to the switching signal
p_i	Weight for the i^{th} basis vector
s	Sparse signal
$\hat{s}$	Estimated sparse signal
ψ_i	i^{th} basis vector
ϕ	Measurement matrix

ψ	Transformation matrix
Ξ	Image data matrix
Ξ^D	Discretized image data matrix
$\mathcal{Q}$	Mapping from the image data matrix to the discretized image data matrix
M	Cantilever FEM global mass matrix
C	Cantilever FEM global damping matrix
K	Cantilever FEM global stiffness matrix
M_i	Mass matrix for the i^{th} cantilever FEM element
C_i	Damping matrix for the i^{th} cantilever FEM element
K_i	Stiffness matrix for the i^{th} cantilever FEM element
$\vec{X}$	State vector for the cantilever FEM model
$\vec{F}$	Forcing vector for the cantilever FEM model
y_i, θ_i	Displacement and deflection of the cantilever at the i^{th} node
f_i, τ_i	Force and torque acting on the i^{th} tip
N_T	Number of tips on the cantilever beam for full beam AFM
N_S	Number of laser spots for full beam AFM
N_E	Number of elements in the cantilever FEM model
$\vec{e}$	Error vector for full beam AFM
$\vec{\theta}_{sp}$	Set-point deflection vector
$\vec{\theta}_{sp}$	True deflection vector
Z_c	Control signal for full beam AFM

List of Figures

List of Tables

Acknowledgements

I would like to thank my supervisor Dr. Mark French for his continued support throughout the course of this research. His insight and ability to determine the best directions of work have been an asset. In addition I would like to thank my co-supervisor Dr. Harold Chong for allowing me to access to his lab, and letting me use his equipment in previously untested ways. Lastly this work would not have been possible without the support of my parents and family, who played a pivotal role in getting me through the demanding periods of this research.

During the time I spent here, I had to explain what is atomic force microscopy to many people many times over. Over the years this act became second nature to me until I wrote the following mild introduction, which shall now precede the more formal one in my thesis.

Once upon a time, there lived a man called Binning,
He worked long and hard, but never really made a shilling,

Then came a day, when his luck really shone,
And no more could he, about his fate moan,

There in his mind dawned, the principle of **A F**antastic **M**achine,
The likes of which until then, had never before been seen,

So he put it together, and called it the **A F M**,
And it made the physicists happy, the whole lot of them,

Thus came to him fame, fortune, and unbounded glory,
And so began, the never ending AFM story **...**

Chapter 1

Introduction

Atomic Force Microscopes or AFMs are instruments which can observe microscopic features on a material surface, with a resolution in nanometers. The term *atomic force* originates from the fact that the imaging process relies upon the existence of the inter-atomic interaction force between the mechanical probe and sample surface.

AFMs generate images by moving a mechanical probe over the sample surface. The mechanical probe in this case is a small rectangular beam approximately 200 μm in length. The beam is fixed at one end and free at the other, and is often referred to as a *Cantilever*. The free end has a pyramid shaped structure with a sharp apex (or *tip*) that faces the sample surface. As the cantilever tip moves or *flies* a few nanometers above the sample, it *feels* the sample surface through the inter-atomic interaction force between tip and sample. For instance, if the tip moves over a *bump* in the sample surface, the tip deflects due to a change in the tip-sample interaction force. A controller measures this deflection though an optical motion detection mechanism, and moves the cantilever base up. When the bump has passed, the controller which is constantly monitoring tip deflection pushes the cantilever base down to its original height. The control signal, once calibrated serves as the estimate of the sample topography in the direction of motion. The entire sample is *rastered* in this manner *line* by *line* and the calibrated control signal is saved. The control signal for all the lines are then *stacked* to form the sample image. The precise meanings of the terms rastered, line and stacked will be explained in detail in the next chapter which focusses on modelling the AFM dynamics.

Given the ability of AFMs to operate in any medium on samples composed of any material, they have become a valuable tool for a wide cross-section of the scientific community. The purpose of this chapter is two-fold. Firstly to elucidate the factors behind the wide usage of this instrument. Secondly, to briefly address the limitations of the current AFMs and how these can be alleviated using control theory methods.

1.1 Why Atomic Force Microscopy?

Ever since the development of the first AFM in 1986 by Binnig and Quate (1986) these microscopes been utilized in a diverse range of applications. This popularity of AFMs is partly due to the fact that they were built in response to a limitation of their predecessor the Scanning Tunnelling Microscope (STM). These devices functioned by passing a *tunnelling* current through the sample being viewed and estimated the surface topography by regulating its value. The limitation of this approach was that it worked for conducting samples only. In contrast to this approach, the AFM utilizes the interaction forces present between materials in close proximity. This eliminates the dependence on the conductivity of the sample being viewed. Furthermore, AFMs can be operated in air, fluid and vacuum which greatly widens their applicability. Finally the ability to detect features in the nanometer range is also significant. A comparison of AFMs to three other forms of microscopy Transmission Electron Microscopy (TEM), Scanning Electron Microscopy(SEM) and Optical Microscopy is given in Table 1.1.

	AFM	**TEM**	**SEM**	**Optical**
Max Resolution	Atomic	Atomic	1's nm	100's nm
Typical Cost($\times 10^3$ USD)	100-200	500 or higher	200-400	10-50
Imaging Environment	air, fluid, vacuum	vacuum	vacuum	air, fluid
Sample preparation	Easy	Difficult	Easy	Easy

Table 1.1: Comparison of AFM and other microscopy techniques (Agilent).

While optical microscopes are the least costly, they can operate in any medium and their sample preparation is also easy, they have the lowest resolution. Scanning and transmission electron microscopes are costly and the sample needs to be placed in vacuum.

The fact that AFMs are capable of atomic resolution was first illustrated by the developer of the instrument in his contribution titled *Atomic Resolution with Atomic Force Microscope* (Binnig et al. (1987)). The author was able to generate images of a graphite surface which revealed the lattice structure with a resolution of 2.5Å. A very recent example of the resolving power of AFMs was demonstrated through the work of IBM researchers in Zurich (Gross et al. (2009)). The authors were able to image a single Pentacene molecule. As mentioned in the same contribution the significance of this development lies in the fact that while STMs can image atomic sized features of sample surfaces, the resolving of individual atoms in a molecule is still a challenge. The images obtained are provided below.

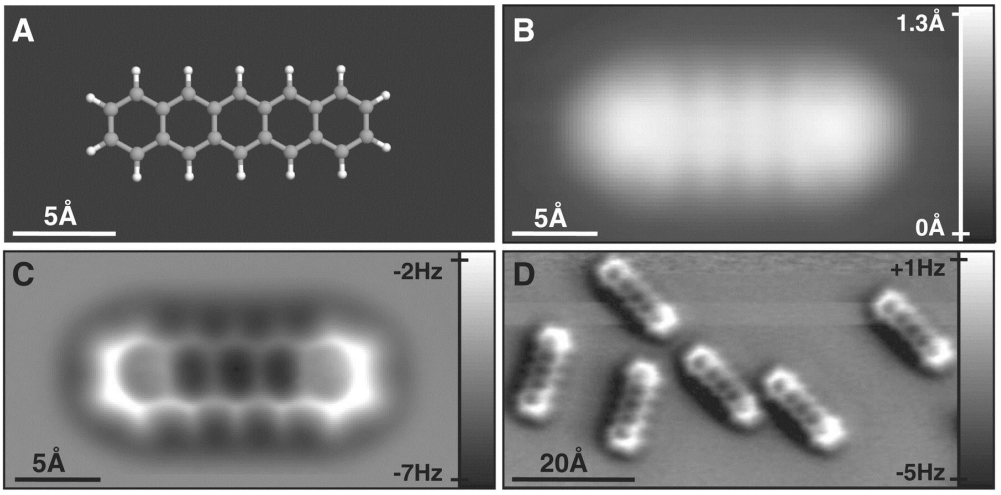

Figure 1.1: (a) *Ball-and-stick* model of the Pentacene molecule. (b) STM image (c and d) AFM images (Reprinted with permission from The American Association for the Advancement of Science.)

The top left image provides the *Ball-and-stick* model of the Pentacene molecule. Here the larger grey balls represent Carbon atoms, whereas the smaller white ones illustrate the Hydrogen atoms. The sticks illustrate the covalent bonds. The top right image is generated using a STM and clearly cannot resolve the structure of the molecule. The two bottom images are generate using an AFM. The particular mode used here is the *Non Contact Frequency Modulation* mode. This means that instead of keeping the cantilever static, it is vibrated at a fixed frequency and moved over the sample surface. During this movement, the frequency shifts in cantilever vibration are recorded, and these are then used for image generation.

AFMs have found substantial usage for the study of biological materials. An example is provided through the work of (Seelert et al. (2000)) as published in Nature.

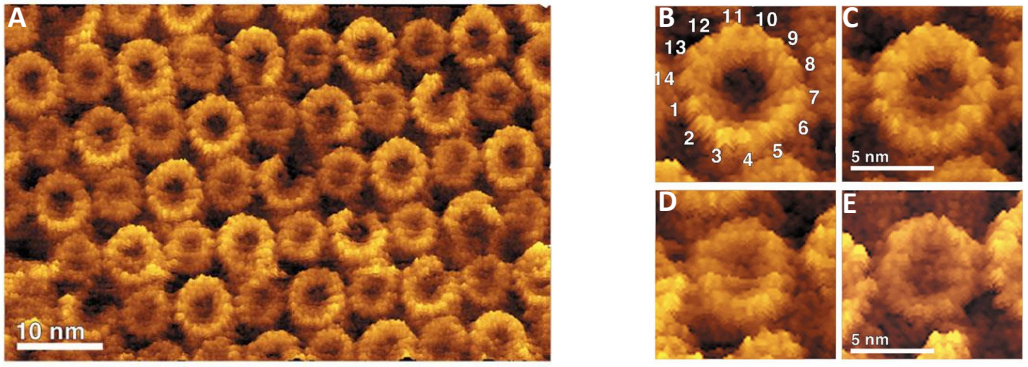

Figure 1.2: a) Narrow and wide ring structures. (b and c) Wide ring structures showing 14 subunits. (d and e) Narrow ring structures. (Reprinted with permission from the Nature Publishing Group.)

Figure 1.2 (a) illustrates the wide and narrow ring proton turbine structures in the two surfaces of an Adenosine Triphosphate (ATP) synthase. As mentioned in the same contribution, ATP synthases are enzymes which can catalyse the synthesis and breakdown of ATP.

Figures 1.2 (b) shows the wide ring structures. It was previously believed that the proton turbine has 12 subunits however this study using an AFM revealed that the number is 14. This is significant since any insight into the proton turbine structure has implications into understanding the corresponding molecular mechanisms and the efficiency of energy conversion. The bottom Figures 1.2 (c) illustrates the narrow ring structure. The topography variation range in these images is 2nm.

AFMs have also been used in the study of the activity of single molecules e.g RNA polymerase (Guthold et al. (1994), Kasas et al. (1997)), the motion of molecular motors such as Myosin V (Kodera et al. (2010)), cancerous cells (Iyer et al. (2009)) and nanomechanical analysis of cells (Yang et al. (2011)). Although these applications are significant, relatively lesser progress has been made to view these processes in real time as they happen. As a result many dynamic processes are still unobserved (Schitter and Rost (2008)).

Another field that has benefited from the development in AFMs is material science. AFMs are being used for the study of material forces e.g, adhesion force (Dos et al. (2010)), alteration of surfaces as in the case of the etching process (Morita et al. (2010)), interface between materials as in the case of nanocontacts (Grodzicki et al. (2009)) and also in the study of friction, lubrication and wear (Martnez-Martnez et al. (2009)).

In addition to this, the atomic force microscopy principle is not just restricted to viewing surface topography or measurement of other material properties, it can also be used for the purpose of nanomanipulation as mentioned in Yang and Jagannathan (2006) and lithography Dauksher et al. (2004). Finally, the concept is being investigated at IBM for the development of data storage devices which can store up to $1Tb/in^2$. This research is being conducted under the IBM Millipede project. The name originates from the fact that instead of having a single cantilever, the data storage device operates an array of cantilevers in parallel.

The purpose of this brief discussion is to elucidate the fact that the AFM device itself and the concept as well is pertinent to the research being conducted by a wide cross-section of researchers. Any automation that makes its usage easier or enhances its performance especially in terms of speed is certain to have far reaching consequences.

1.2 Performance Limitations in Current AFMs

While AFMs have a few advantages over other instruments, they have two key performance limitations namely,

• They require a substantial amount of time for generating an image.

• Furthermore, their usage demands that the end users have a basic understanding of feedback control theory.

Time limitations occur due to the fundamental nature of AFM operation, which requires a mechanical probe to scan the sample surface point by point in a sequential manner. Typical imaging times range from a few seconds for a small scan area to tens of minutes for larger area. Sample scan areas range from 100×100 nm to 20×20 μm.

As will be explained in the next chapter, the probe motion needs to be controlled using a feedback control mechanism. Commercial AFMs rely upon the rudimentary proportional integral (PI) controllers for correct functioning. These controllers must be tuned by the end users of these devices. Given the fact that AFMs find usage in a diverse range of fields from microbiology, material science to nano interrogation, the end users need to develop some understanding of control principles in order to use the device effectively.

It is clear that if the control tuning procedure can be automated, the usage of AFMs will become easier and hence these microscopes will become accessible to a wider group of scientists. The primary objective of this research therefore is to automate this tuning process. Although this thesis focuses on controller tuning alone, the concept investigated here can also be applied to more sophisticated control mechanisms for instance robust and adaptive controllers.

In addition to focusing on the automation of the AFM controller synthesis, this research will also investigate the second major problem facing AFMs today i.e, speed of operation. Current AFMs fall far short of the demands for operation speeds placed upon them, particularly by molecular biologists. This group of scientists is interested in viewing the interaction of biological molecules in real time, for instance DNA motors as illustrated in Kodera et al. (2010). This in turn requires that AFMs be able to view samples at frame rates as high as 30 frames per second, however currently this speed is mostly limited to close to a minute per single frame (Schitter and Rost (2008)). The second objective of this research is therefore to investigate means that can reduce the imaging times. Two promising methods that will be investigated for speed improvement during the course of this research are *Compressive Sensing* and a novel approach *Full Beam* atomic force microscopy. Furthermore it will also be explained how the automated tuning is vital for the compressive sensing approach.

This research has been reported through the following contributions,

• Khan, U., Chong, H., French, M. (2013) Automated controller tuning for Atomic Force Microscopes using Estimation Based Multiple Model Switched Adaptive Control. In: *IEEE Conference on Decision and Control*, 10 December 2013.

• Khan, U., Chong, H., French, M. (2013) Real time controller implementation for an atomic force microscope using a Digital Signal Processor. In: *IEEE International Conference on Applied Electronics*, 10 September 2013.

• Khan, U., Chong, H., French, M. (2013) Full beam Atomic Force Microscopy. In: *IEEE International Conference on Automation and Computing*, 14 September 2013.

The next section provides an outline of the work done in this dissertation.

1.3 Thesis Objectives and Outline

In the context of the information presented in the previous sections this thesis has two distinct objectives given below,

• The first objective is to develop an automated controller tuning mechanism that can remove the burden of manual gain tuning from the end user. This mechanism should be able to *retrofit around* existing commercial AFMs easily with minimal hardware intervention.

• The second objective of the thesis is to investigate novel ideas which have the potential of reducing imaging times in an AFM.

The remaining part of this section provides the thesis outline.

Chapter 2 discusses in detail the operating principles of a commonly used atomic force microscopy set up and generates the necessary models for controller synthesis. It also provides a literature review of the contributions made so far to address the limitations of AFMs, identifies potential limitations of these contributions and provides justification for the method of choice for this research. As it will be explained in the chapter, an *Estimation Based Multiple Model Switched Adaptive Control* (EMMSAC) method provides a promising solution to the control problems associated with atomic force microscopy.

Chapter 3 introduces the EMMSAC method, applies the same for a simulated atomic force microscopy experiment using the model developed in Chapter 2 and provides the simulation results.

Chapter 4 illustrates the experimental results obtained using a commercial AFM. In these experiments the EMMSAC algorithm was used to generate sample images with no manual tuning involved.

Chapter 5 explains two novel methods for reduction in AFM imaging times. The first one suggests *sparse* trajectories for scanning the sample and then using *Compressive Sensing* for estimating the sample topography. Compressive sensing is a relatively new development in the field of signal processing which attempts to reconstruct signals which have been sampled at sub Nyquist rates. The exact manner in which this is achieved, as well as how it is relevant for AFMs will be explained in detail in the chapter. The second method referred to as *Full Beam* atomic force microscopy, uses the AFM probe in a novel manner to scan the sample faster.

Chapter 6 concludes the work done in this research and highlights directions for possible future work.

Chapter 2

AFM Fundamentals

This chapter outlines the functioning of the atomic force microscope (AFM). This includes a description of the AFM set-up followed by modelling of the tip sample interaction, optical lever, description of the piezo-actuators, the optical detection system, simulation results of the AFM control loop and finally a literature review.

The motivation behind understanding and modelling AFM dynamics is that it is necessary for automated controller generation. In practice AFM users must work with a diverse range of AFM experimental set-ups, and each time they must manually tune the controller using a trial and error procedure. However if the AFM dynamics for a sufficiently large set of experimental set-ups is known a priori, it is possible to determine the *closest* model and suggest an appropriate controller. This can substantially reduce the time needed for controller tuning and make AFM usage easier. The model closeness is measured using deterministic Kalman filter residuals and is described in detail in Chapter 3.

2.1 AFM Setup

The set-up of an AFM as illustrated in Figure 2.1 consists of a cantilever, a piezo-actuator, an optical detection system, a controller and the sample. The typical cantilever has a length of 100-200 μm and is usually made of silicon oxide or silicon nitride (Sebastian (2004)). The cantilever has a sharp pyramid shaped tip which is meant to probe the surface. The other side of the cantilever i.e. the base is mounted on a piezo-actuator that can translate vertically This piezo-actuator is referred to as the Z piezo-actuator. The sample lies on top of a fixed base. The cantilever tip is then brought in close proximity of the sample, usually within approximately 100nm and is then forced to raster over the sample laterally. The rastering movement is made possible by the XY piezo-actuator block, as illustrated in the same figure. For this purpose, V_X and

V_Y are selected to be a triangular and ramp signals as explained in Section 2.4. Here V_X and V_Y are voltage input signals for the XY pizeo actuator block. As the rastering occurs the cantilever tip deflects due to the interaction force between the tip and the sample. Detailed explanations about the interaction force are provided in Section 2.2. The deflections are detected by an optical detection system which consists of a laser beam incident upon the cantilever tip and a split photo diode that detects the reflected laser. The split photo diode is a four quadrant sensor. The reflected laser spot must remain at the centre of the four quadrants. The voltages of the top two and bottom two quadrants are added. The difference between the two sums normalized by the sum of all the four voltages generates the final output of the photo diode, which is referred to as v_T. The controller measures the photo-diode voltage v_T and regulates the height of the cantilever base so that this voltage remains at a reference value v_{sp}. This in turn ensures that the distance between the tip and the sample is regulated. The controller can change the height of the cantilever base by altering the input voltage of the Z piezo-amplifier h. This is then amplified to V_Z and sent to the piezo-actuator. Finally, the sample topography is obtained by recording the control voltage h applied to the Z piezo-amplifier and scaling it with the actuator's calibration constant in nanometers/volts. This mode of AFM operation is referred to as *Contact Mode Atomic Force Microscopy*.

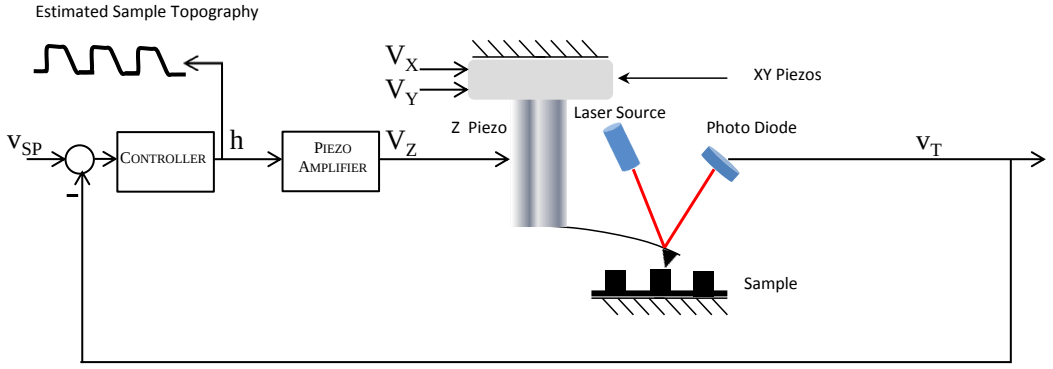

Figure 2.1: The AFM experimental setup (Contact Mode).

Although this form of atomic force microscopy is easy to understand, it has the disadvantage of greater contact between the tip and the sample. This is so because it wears the tip and can also damage soft biological samples. As a consequence an alternate form was developed and is known as *Dynamic Mode Atomic Force Microscopy*. Some authors refer to the same mode as *Tapping Mode* AFM since cantilever tip *taps* the sample surface. As illustrated in Figure 2.2 the only difference is that the cantilever base is excited by an additional piezo-actuator, namely the *dither piezo-actuator*.

The dither piezo-actuator is provided a sinusoidal voltage signal f_D at a frequency close to the first resonance frequency of the cantilever. This causes the cantilever tip to vibrate. Then as previously done, the vibrating cantilever tip is rastered over the sample. This time, owing to changes in the tip-sample interaction force due to variation

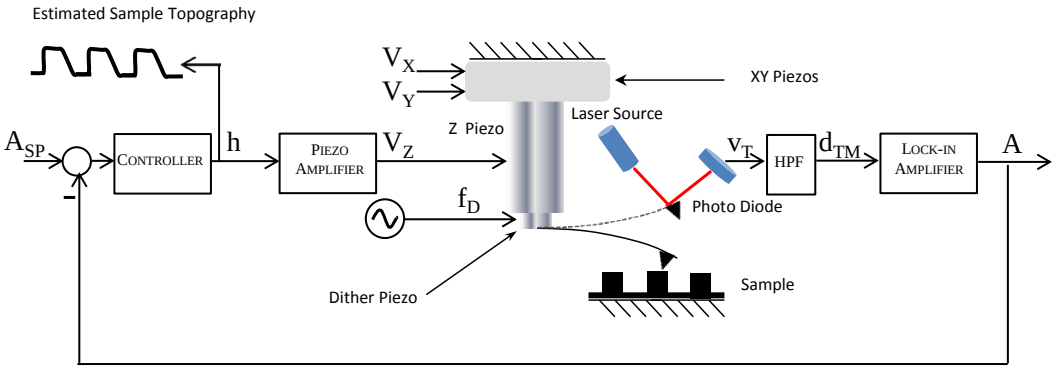

Figure 2.2: The AFM experimental setup (Dynamic Mode).

in sample topography the amplitude of tip vibration is altered. The purpose of the controller now is to regulate the height of the sample to keep the amplitude of tip vibration constant. This is achieved by extracting the amplitude A of the photo-diode signal d_{TM} using a lock-in amplifier and comparing it against a set point amplitude A_{sp}. The subscripts TM stand for tapping mode. The d_{TM} signal is a high pass filtered version of the photo-diode's voltage signal v_T. The high pass filtering extracts the high frequency cantilever vibration component of v_T. This dissertation is based only on this later case i.e, dynamic atomic force microscopy. The set up illustrated in Figure 2.2 is implemented using the CryoView 2000 AFM from Nanonics Imaging® available in the Nanogroup at the University of Southampton (illustrated in Figure 2.3 below).

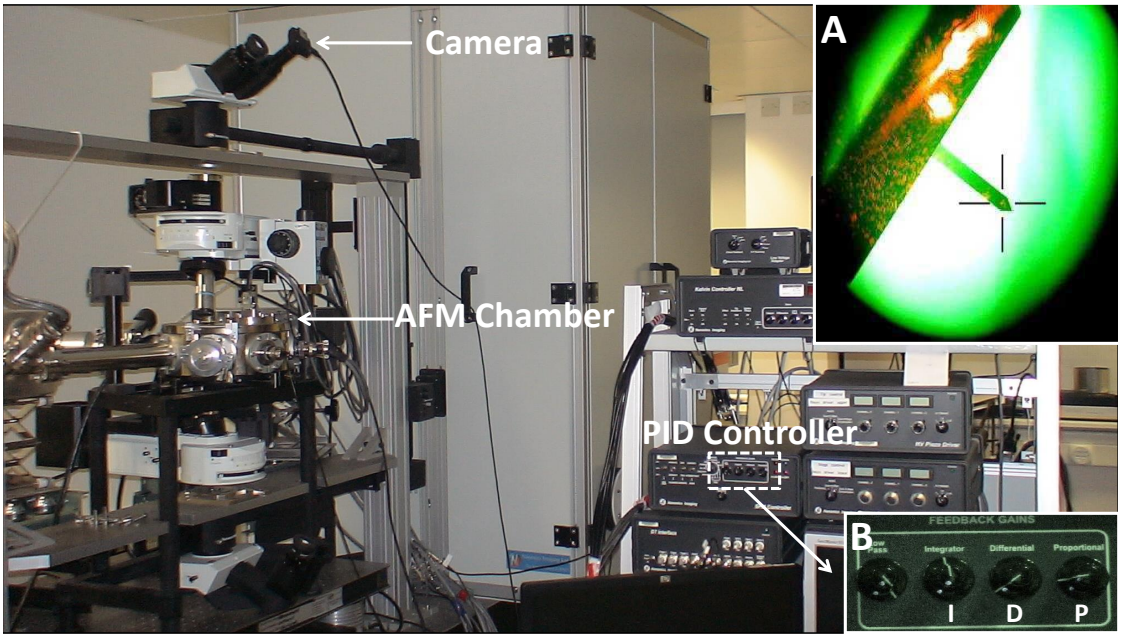

Figure 2.3: The Nanonics CV2000 AFM. Inset A:Cantilever as seen from the camera. Inset B: PID controller.

The figure illustrates the Nanonics CV 2000 AFM. The AFM chamber is a *pot* like ensemble with a hatch on top. The hatch contains the X, Y, Z, and dither piezo-actuators. It also has a window through which a laser beam can pass. In addition it also allows for viewing the cantilever and the laser spot using the camera at the top of the instrument. Inset A illustrates the view typically seen by the camera during AFM operation. As can be seen, the laser spot is close to the base of the cantilever. The user must then move the spot manually to the tip of the cantilever indicated by the cross hairs. The right side of the figure illustrates the analog electronics hardware provided by the manufacturer. Inset B illustrates the PID control panel.

As can be seen, these can only be tuned manually. This led to the necessity of making a customized hardware using a Digital Signal Processor (DSP), which can be interfaced directly with an automated tuning algorithm running on a desktop.

The remaining sections 2.2-2.5 of this chapter describe in detail the components of atomic force microscopy outlined here. Section 2.6 provides simulation results for the closed loop AFM system and the last section provides a literature review of the existing control methods used for AFM control.

2.2 Tip Sample Interaction

The tip sample interaction force is the interaction force between the cantilever tip and the sample surface. It consists of long range attractive and short range repulsive forces as mentioned in Garcia and Paulo (1999). This force has a conservative component that depends only on tip sample separation d and a non-conservative one that depends on the tip velocity $\dot{d}$ (Melcher et al. (2008b)). The force is dominated by van der Walls force (Israelachvili (1985)) between two atoms if the tip sample separation is greater than inter-atomic distance for the sample a_o. If the separation is less than a_o the interaction is determined by the adhesion force given by the Dejarguin-Muller-Toporov(DMT) theory. The conservative component of the interaction force can thus be given as,

$$F_{ts} = \begin{cases} -\frac{HR}{6d^2} & \text{if } d > a_o; \\ -\frac{HR}{6a_o^2} + \frac{4}{3}E^*\sqrt{R}(a_o - d)^{\frac{3}{2}} & \text{if } d \leq a_o. \end{cases} \tag{2.1}$$

where R is the tip radius, H is the Hammaker constant and E^* is the effective elastic modulus of the sample. One possibility of modelling the non-conservative component of the interaction force is to use the Kelvin-Vigot viscoelastic contact damping model as suggested by Melcher et al. (2008b). Using this model the non-conservative component can be given as,

$$F_{nc}(d, \dot{d}) = -\zeta \dot{d} \sqrt{R(a_o - d)} \tag{2.2}$$

Here ζ is the sample viscosity. The current simulation neglects the non-conservative component and its effect remains to be explored.

The following Figure 2.4 illustrates the conservative component of the tip-sample interaction force for Hammaker constant $H = 7.1 \times 10^{-20} J$, tip radius $R = 10nm$, intermolecular distance $a_o = 0.16nm$ and effective elasticity modulus $E = 1.3GPa$.

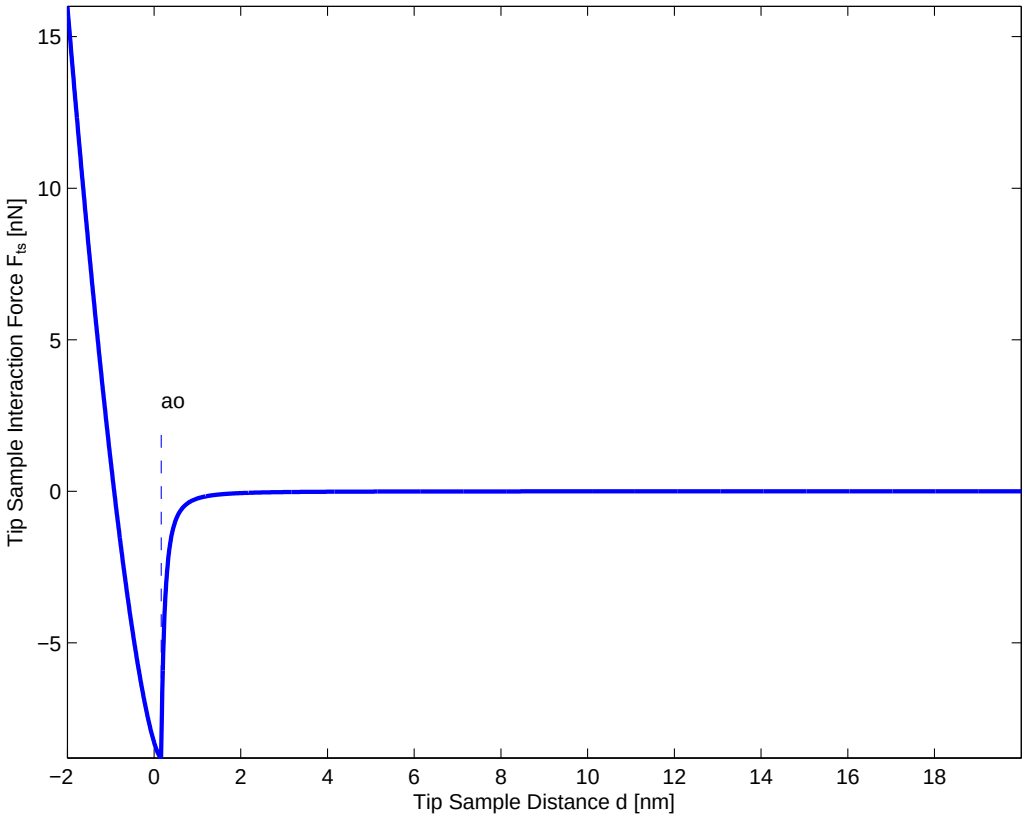

Figure 2.4: Tip Sample Interaction Force.

The region where the tip-sample separation is greater than a_o is referred to as the *attractive* region. The other one is known as the *repulsive* region. In the former the tip is attracted towards the sample surface, whereas in the later tip starts penetrating the sample surface and is repulsed.

The effective elasticity modulus E^* can be determined from the elasticity moduli of the tip E_{tip} and sample E_{sample} and the Poisson ratios for tip, ν_{tip} and sample, ν_{sample} using the relation (Melcher et al. (2008b)),

$$E^* = \left[\frac{1 - \nu_{tip}^2}{E_{tip}} + \frac{1 - \nu_{sample}^2}{E_{sample}} \right]^{-1} \tag{2.3}$$

2.3 Optical Lever Model

The purpose of this section is to derive a model for the optical lever dynamics (cantilever), which will subsequently be used for application of control principles. It concludes by experimental identification of the actual cantilever used in this research, and a comparison between the true and estimated system response.

The optical lever can be modelled as a cantilever beam which is excited at the base (the fixed end) and the free end has a tip mass attached. The tip end is nearly atomically sharp i.e, there is only one atom at the apex and is effected by the tip sample interaction forces described in section 2.2. The cantilever is an infinite degree of freedom system (DOF) where a single DOF corresponds to an eigenmode of the beam.

The derivation starts by representing the cantilever dynamics using the Euler Bernoulli beam equation and then uses Galerkin discretization to obtain a spring mass damper system that is equivalent to the first eigenmode. Two assumptions are made. Firstly, it is assumed that cantilever is vibrating in air. Secondly, it is assumed that the base is excited with a sinusoid at a frequency close to the resonance of the first eigenmode.

Figure 2.5 illustrates the cantilever beam beam being modelled and the equivalent spring mass damper system.

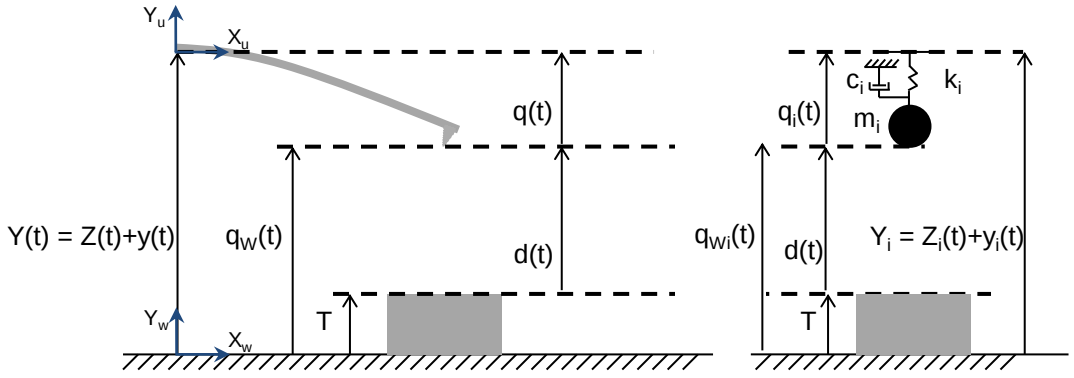

Figure 2.5: Euler Bernoulli beam and spring mass system equivalence.

Here $Z(t)$ and $y(t)$ are the movements of the cantilever base due to the action of the Z and dither piezo-actuators respectively at time t. The actual movement $Y(t)$ is then simply the sum of the two. These quantities are represented in the global coordinate frame W. Since the dither piezo is actuated by a sinusoidal signal f_D, the resulting movement is also sinusoidal and can be written as follows,

$$y(t) = A_{base} \cos \omega t \tag{2.4}$$

where ω is the excitation frequency in rads^{-1} and A_{base} is the amplitude of the cantilever base movement due to the action of the dither piezo. The height of the sample surface feature is denoted by T and the tip sample separation by $d(t)$. For the purpose deriving the cantilever dynamics a separate coordinate frame U is attached to the base of the cantilever. In this frame the tip deflection is denoted by $q(t)$. The tip height in the frame W is referred to as $q_W(t)$ where,

$$q_W(t) = T + d(t) \tag{2.5}$$

Lastly, m_i, k_i, c_i, q_i, q_{Wi} and Y_i represent the mass, spring constant, damping constant, tip deflection, tip height and base movement for equivalent spring-mass-damper system corresponding to the i^{th} eigenmode.

The derivation is carried out in three steps as done by Melcher et al. (2011). First the base excitation force is converted into an equivalent force that acts uniformly on the entire length of the beam. In the second step the Euler Bernoulli equation is solved for the case with no damping, base excitation and tip sample interaction to obtain the eigenmodes of the beam. Finally the Galerkin discretization procedure is applied to get an equivalent spring mass system that explains the dynamics of a single eigenmode of the beam. The remaining part of this section provides the derivation.

The dynamics of the cantilever using the Euler Bernoulli equation can be written as,

$$EI\frac{\partial^4 w(x,t)}{\partial x^4} + \rho_c \frac{\partial^2 w(x,t)}{\partial t^2} + \gamma \frac{\partial w(x,t)}{\partial t} = F_{ts}(d)\delta(x-L) \tag{2.6}$$

where x is the axial coordinate, E, I, L and ρ_c are the beams elastic modulus, moment of inertia, length and mass per unit length respectively, γ is the hydrodynamic damping coefficient and $w(x,t)$ represents the beam deflection as a function of axial displacement and time. F_{ts} is the tip sample interaction force that acts at the tip of the beam as indicated by the Dirac Delta operator $\delta(x-L)$.

The beam deflection $w(x,t)$ in the frame attached to the base W has three contributing factors, i.e $w(x,t) = Z(t) + y(t) + u(x,t)$. Here $u(x,t)$ is the beam deflection in the frame attached to the cantilever U, and $q(t) = u(L,t)$ is the tip deflection. Writing $u(x,t)$ in terms of the three components, bringing all terms with $\dot{y}, \ddot{y}, \dot{Z}, \ddot{Z}$ to the right and assuming $\left|\dot{Z}(t)\right| < \left|\dot{y}(t)\right|$ and $\left|\ddot{Z}(t)\right| < \left|\ddot{y}(t)\right|$, we get,

$$EI\frac{\partial^4 u(x,t)}{\partial x^4} + \rho_c\frac{\partial^2 u(x,t)}{\partial t^2} + \gamma\frac{\partial u(x,t)}{\partial t} = F_{ts}(d,\dot{d})\delta(x-L) - \rho_c\ddot{y}(t) - \gamma\dot{y}(t) \tag{2.7}$$

Now letting,

$$F_b(x,t) = -\rho_c\ddot{y}(t) - \gamma\dot{y}(t) \tag{2.8}$$

where F_b is the base excitation force, the equation becomes,

$$EI\frac{\partial^4 u(x,t)}{\partial x^4} + \rho_c\frac{\partial^2 u(x,t)}{\partial t^2} + \gamma\frac{\partial u(x,t)}{\partial t} = F_{ts}(d,\dot{d})\delta(x-L) + F_b \tag{2.9}$$

Before the Galerkin discretization procedure can be applied, the Euler Bernoulli equation needs to be solved for the undamped and unforced case to derive the eigenmodes. For this case the Euler Bernoulli equation becomes,

$$EI\frac{\partial^4 u(x,t)}{\partial x^4} + \rho_c\frac{\partial^2 u(x,t)}{\partial t^2} = 0 \tag{2.10}$$

The solution to this equation can be obtained using the separation of variables approach. The reason for doing this is that vibrating the beam always exhibits linear combinations of unique profiles. These profiles do not change shape, only their amplitudes change with time (Meirovitch (2001)). These profiles are the eigenmodes of the beam. There can be multiple eigenmodes and for each one there exists one corresponding solution. The solution corresponding to the i^{th} eigenmode can therefore be written as,

$$u_i(x,t) = \phi_i(x)q_i(t) \tag{2.11}$$

where $\phi_i(x)$ is the i^{th} eigenmode and $q_i(t)$ is its amplitude. Substituting this solution into (2.10) we get for each value of i,

$$-EIq_i(t)\frac{d^4\phi_i(x)}{dx^4} = \rho_c\phi_i(x)\frac{d^2 q_i(t)}{dt^2} \tag{2.12}$$

Now using the standard argument in the method of separation of variables (Meirovitch (2001)) both sides of the equality must be equal to a constant. Since all the quantities appearing in equation are real therefore the constant must also be real. Let the constant be ω_i^2, then it can be shown

$$-\frac{EI}{\rho_c\phi_i(x)}\frac{d^4\phi_i(x)}{dx^4} = \frac{1}{q_i(t)}\frac{d^2 q_i(t)}{dt^2} = \omega_i^2 \tag{2.13}$$

and therefore we get the following two equations,

$$\frac{d^2 q_i(t)}{dt^2} + \omega_i^2 q_i(t) = 0 \tag{2.14}$$

and

$$\frac{d^4 \phi_i(x)}{dx^4} - \left(\frac{\alpha_i}{L}\right)^4 \phi_i(x) = 0;\ \alpha_i^4 := \frac{\omega_i^2 \rho_c L^4}{EI} \tag{2.15}$$

The first (2.14) has a solution of the form,

$$q_i(t) = c_1 \sin \omega_i t + c_2 \cos \omega_i t \tag{2.16}$$

Here c_1 and c_2 are constants that can be determined from the initial position profile $u(x,0)$ and velocity profile $\dot{u}(x,0)$. The second (2.15) has a solution of the form,

$$\phi_i(x) = A \sin\left(\frac{\alpha_i x}{L}\right) + B \cos\left(\frac{\alpha_i x}{L}\right) + C \sinh\left(\frac{\alpha_i x}{L}\right) + D \cosh\left(\frac{\alpha_i x}{L}\right) \tag{2.17}$$

The constants A,B,C and D can be found by applying the boundary conditions. For the current problem there are four conditions, two for the fixed end and two for the free end. The boundary conditions for the fixed end are given by,

$$u(0,t) = 0 \tag{2.18}$$

$$\frac{d}{dx} u(x,t)\Big|_{x=0} = 0 \tag{2.19}$$

The first condition implies that the fixed end does not move in the co-ordinate frame attached to the base, and the second one relates to the fact that the beam has no slope with respect to the x-axis at the fixed end. The boundary conditions for the free end are as follows,

$$\frac{d^2}{dx^2} u(x,t)\Big|_{x=L} = 0 \tag{2.20}$$

$$EI \frac{d^3}{dx^3} u(x,t)\Big|_{x=L} = m_{tip} \ddot{u}(L,t) \tag{2.21}$$

where m_{tip} is the mass of the cantilever tip.

Now applying the solution $u_i(x,t) = \phi_i(x) q_i(t)$ into the boundary conditions we get,

$$
\begin{aligned}
u(0,t) &= 0 \implies \phi_i(0)q_i(t) = 0 \implies \phi_i(0) = 0 && \text{(2.22a)}\\
\frac{\partial}{\partial x}u(x,t)\Big|_{x=0} &= 0 \implies \left[\frac{d}{dx}\phi_i(x)\right] q_i(t) = 0 \implies \frac{d}{dx}\phi_i(x)\Big|_{x=0} = 0 && \text{(2.22b)}\\
\frac{\partial^2}{\partial x^2}u(x,t)\Big|_{x=L} &= 0 \implies \left[\frac{d^2}{dx^2}\phi_i(x)\right] q_i(t) = 0 \implies \frac{d^2}{dx^2}\phi_i(x)\Big|_{x=L} = 0 && \text{(2.22c)}\\
\frac{\partial^3}{\partial x^3}u(x,t)\Big|_{x=L} &= m_{tip}\ddot{u}(L,t) \implies EI\frac{d^3}{dx^3}\phi_i(x) + \omega_i^2 m_{tip}\phi_i(L) = 0 && \text{(2.22d)}
\end{aligned}
$$

The last relation can be obtained as follows,

$$
\begin{aligned}
EI\frac{\partial^3}{\partial x^3}u(x,t)\Big|_{x=L} &= m_{tip}\ddot{u}(L,t)\\
EI\frac{d^3}{dx^3}\phi_i(x)\Big|_{x=L} q_i(t) &= m_{tip}\phi_i(L)\ddot{q}_i(t)
\end{aligned}
\qquad (2.23)
$$

Since $\ddot{q}_i(t) = -\omega_i^2 q_i(t)$ as mentioned in (2.14),

Therefore,

$$EI\frac{d^3}{dx^3}\phi_i(x)\Big|_{x=L} + \omega_i^2 m_{tip}\phi_i(L) = 0 \qquad (2.24)$$

Now applying the boundary condition in (2.22a) into (2.17) we obtain,

$$\phi_i(0) = B + D = 0 \qquad (2.25)$$

therefore,

$$D = -B \qquad (2.26)$$

The boundary condition in (2.22b) results in the following expression,

$$\frac{d\phi_i(x)}{dx}\Big|_{x=0} = \frac{\alpha_i}{L}\left(A\cos\frac{\alpha_i x}{L} - B\sin\frac{\alpha_i x}{L} + C\cosh\frac{\alpha_i x}{L} + D\sinh\frac{\alpha_i x}{L}\right) = 0 \qquad (2.27)$$

after setting $x = 0$ we get,

$$C = -A \qquad (2.28)$$

Now using the boundary condition in (2.22c) and eliminating C and D we get,

$$\begin{aligned} \frac{d^2}{dx^2}\phi(x)\Big|_{x=L} = \frac{\alpha_i^2}{L^2}\left(-A\sin\alpha_i - B\cos\alpha_i - A\sinh\alpha_i - B\cosh\alpha_i\right) &= 0 \\ \implies A(\sin\alpha_i + \sinh\alpha_i) + B(\cos\alpha_i + \cosh\alpha_i) &= 0 \end{aligned} \tag{2.29}$$

Repeating the same procedure for the last boundary condition (2.22d), eliminating C and D and substituting $m_{tip} = \bar{m}_{tip}\rho_c L$ we get,

$$\begin{aligned} &[(\cos\alpha_i + \cosh\alpha_i) - \alpha\bar{m}_{tip}(\sin\alpha_i - \sinh\alpha_i)]\,A + \quad \ldots \\ &[(-\sin\alpha_i + \sinh\alpha_i) - \alpha\bar{m}_{tip}(\cos\alpha - \cosh\alpha_i)]\,B = 0 \end{aligned} \tag{2.30}$$

Here ρ_c is the linear density of the cantilever beam, and $\bar{m}_{tip}$ is the tip mass normalized with respect to the beam mass.

Combining the last two equations in matrix form we get,

$$\begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{21} & \psi_{22} \end{bmatrix} \begin{bmatrix} A \\ B \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{2.31}$$

where

$$\begin{aligned} \psi_{11} &= \sin\alpha_i + \sinh\alpha_i \\ \psi_{12} &= \cos\alpha_i + \cosh\alpha_i \\ \psi_{21} &= (\cos\alpha_i + \cosh\alpha_i) - \alpha_i\bar{m}_{tip}(\sin\alpha_i - \sinh\alpha_i) \\ \psi_{22} &= (-\sin\alpha_i + \sinh\alpha_i) - \alpha_i\bar{m}_{tip}(\cos\alpha_i - \cosh\alpha_i) \end{aligned} \tag{2.32}$$

For a non-trivial solution it is clear that the solution $\begin{bmatrix} A \\ B \end{bmatrix}$ lies in the null space of the 2×2 matrix, which therefore must be rank deficient and its determinant must be zero. The simplified determinant is set equal to zero to give the following equation,

$$\cos\alpha_i\cosh\alpha_i + 1 + \bar{m}_{tip}\alpha_i(\cos\alpha_i\sinh\alpha_i - \sin\alpha_i\cosh\alpha_i) = 0 \tag{2.33}$$

This relation is known as the *dispersion relationship*. The derivation for this relation can be found in Appendix A. The values of α_i that solve this relation are used to determine resonance frequencies of the eigenmodes according to the relation,

$$\omega_i^2 = \frac{EI\alpha_i^4}{\rho_c L^4} \tag{2.34}$$

Since the system is rank deficient only one of the constants can be determined, the other one is chosen arbitrarily. Choosing $A = 1$ and solving for B we get,

$$B = -\frac{\sin\alpha_i + \sinh\alpha_i}{\cos\alpha_i + \cosh\alpha_i} \tag{2.35}$$

Substituting all the constants in the general expression for an eigenmode (2.17) we get,

$$\phi_i(x) = \sin\frac{\alpha_i x}{L} - \sinh\frac{\alpha_i x}{L} - \left[\frac{\sin\alpha_i + \sinh\alpha_i}{\cos\alpha_i + \cosh\alpha_i}\right]\left[\cos\frac{\alpha_i x}{L} - \cosh\frac{\alpha_i x}{L}\right] \tag{2.36}$$

Since the cantilever is a linear time invariant system, by using the superposition property of linear systems the sum of all solutions corresponding to each eigenmode is also a solution and can be written as,

$$u(x,t) = \sum_{i=0}^{\infty} \phi_i(x) q_i(t) \tag{2.37}$$

Equation (2.15) represents a differential eigenvalue problem of the form $A\phi(x) = \lambda\phi(x)$, where $A = \frac{d^4}{dx^4}$ is a differential operator, $\lambda = \left(\frac{\alpha_i}{L}\right)^4$ are the eigenvalues. The eigenfunctions that solve this differential eigenvalue problem span the entire space of solutions $u(x,t) = \sum_{i=0}^{\infty} \phi_i(x) q_i(t)$. Therefore the solution obtained using the eigenfunctions is the complete solution.

This is the solution that will be used for the remaining part of this derivation.

The solution to the original equation (2.6) of the Euler Bernoulli beam for the forced and damped case can now be given using the fact that the normalized eigenfunctions $\bar{\phi}(x)$ can be used as the basis for the Galerkin discretization procedure. The discretization is performed by substituting the solution $u(x,t) = \sum_{i=1}^{\infty} \bar{\phi}_i(x) q_i(t)$ into (2.6) and then finding the residual and setting it equal to zero to solve for $q_i(t)$ as follows,

$$\int_0^L \bar{\phi}_i(x) \left(\rho_c \sum_{j=1}^{\infty} \bar{\phi}_j(x)\ddot{q}_j(t) + \gamma\bar{\phi}_i(x)\dot{q}_i(t) + EI\sum_{j=1}^{\infty} \frac{d^4}{dx^4}\bar{\phi}_j(x) q_j(t) - F(x,t) \right) dx = 0 \tag{2.38}$$

where,

$$\bar{\phi}_i(x) = \frac{\phi_i(x)}{\phi_i(L_c)} \tag{2.39}$$

Moving the summation outside the integral we get,

$$\left(\rho_c \sum_{j=1}^{2} \int_0^L \bar{\phi}_i \bar{\phi}_j dx\right) \ddot{q}_j + \left(\gamma \sum_{j=1}^{2} \int_0^L \bar{\phi}_i \bar{\phi}_j dx\right) \dot{q}_j + \left(EI \sum_{j=1}^{2} \int_0^L \bar{\phi}_i \bar{\phi}_j'''' dx\right) q_j = \int_0^L \bar{\phi}_i F dx \tag{2.40}$$

For the purpose of conserving space the arguments of the eigenfunctions (x), the forcing function (x,t) are dropped and derivatives with respect to x are indicated with prime. Equation (2.40) will now be solved for the case where $i = 1, 2$ and $j = 1, 2$. The same solution can be extended for the infinite dimensional case as well. Substituting the values for i and j we get the following matrix equation,

$$\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} + \begin{bmatrix} k_{11} & k_{12} \\ k_{21} & k_{22} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \tag{2.41}$$

where,

$$\begin{aligned}
m_{11} &= \rho_c \int_0^L \bar{\phi}_1^2 dx, \quad m_{12} = m_{21} = \rho_c \int_0^L \bar{\phi}_1 \bar{\phi}_2 dx, \quad m_{22} = \rho_c \int_0^L \bar{\phi}_2^2 dx \\
c_{11} &= \gamma \int_0^L \bar{\phi}_1^2 dx, \quad c_{12} = c_{21} = \gamma \int_0^L \bar{\phi}_1 \bar{\phi}_2 dx, \quad c_{22} = \gamma \int_0^L \bar{\phi}_2^2 dx \\
k_{11} &= EI \int_0^L \bar{\phi}_1 \bar{\phi}_1'''' dx, \quad k_{22} = EI \int_0^L \bar{\phi}_2 \bar{\phi}_2'''' dx \\
k_{12} &= EI \int_0^L \bar{\phi}_1 \bar{\phi}_2'''' dx, \quad k_{21} = EI \int_0^L \bar{\phi}_2 \bar{\phi}_1'''' dx \\
F_1 &= \int_0^L \bar{\phi}_1 F dx, \quad F_2 = \int_0^L \bar{\phi}_2 F dx
\end{aligned} \tag{2.42}$$

Now using the following relationship,

$$EI \int_0^L \bar{\phi}_i \bar{\phi}_j'''' dx = EI \int_0^L \bar{\phi}_i'' \bar{\phi}_j'' dx - \omega_j^2 m_{tip} \tag{2.43}$$

The terms k_{ij} can be written as,

$$
\begin{aligned}
k_{11} &= EI\int_0^L (\bar{\phi}_1'')^2 dx - \omega_1^2 m_{tip}, \quad k_{22} = EI\int_0^L (\bar{\phi}_2)^2 dx - \omega_2^2 m_{tip} \\
k_{12} &= EI\int_0^L \bar{\phi}_1''\bar{\phi}_2'' dx - \omega_2^2 m_{tip}, \quad k_{21} = EI\int_0^L \bar{\phi}_1''\bar{\phi}_2'' dx - \omega_1^2 m_{tip}
\end{aligned}
$$

The detailed derivation for (2.43) can be found in A.2. Substituting k_{11}, k_{22}, k_{12} and k_{21} into (2.41) we get,

$$
\begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} + \begin{bmatrix} \tilde{k}_{11} & \tilde{k}_{12} \\ \tilde{k}_{21} & \tilde{k}_{22} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} - m_{tip} \begin{bmatrix} \omega_1^2 & \omega_2^2 \\ \omega_1^2 & \omega_2^2 \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \tag{2.44}
$$

Here $\tilde{k}_{ij}$ is the first part of k_{ij} i.e the one obtained from the integration. From (2.14) we know that $\omega_j^2 q_j = -\ddot{q}_j$, therefore (2.44) becomes,

$$
\begin{bmatrix} m_{11}+m_{tip} & m_{12}+m_{tip} \\ m_{21}+m_{tip} & m_{22}+m_{tip} \end{bmatrix} \begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} + \begin{bmatrix} \tilde{k}_{11} & \tilde{k}_{12} \\ \tilde{k}_{21} & \tilde{k}_{22} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \tag{2.45}
$$

Now using the orthogonality and companion orthogonality conditions for the eigenmodes (Meirovitch (2001)) given below,

$$
\begin{aligned}
\int_0^L \rho_c \bar{\phi}_i \bar{\phi}_j dx &= -m_{tip}, \quad i,j = 1,2,\ldots; i \neq j && \text{(2.46a)} \\
EI\int_0^L \bar{\phi}_i'' \bar{\phi}_j'' dx &= 0, \qquad\quad i,j = 1,2,\ldots; i \neq j && \text{(2.46b)}
\end{aligned}
$$

Using these,

$$
\begin{aligned}
m_{12} &= m_{21} = -m_{tip} + m_{tip} = 0 && \text{(2.47)} \\
c_{12} &= c_{21} = \frac{-\gamma m_{tip}}{\rho_c} && \text{(2.48)} \\
\tilde{k}_{12} &= \tilde{k}_{21} = 0 && \text{(2.49)}
\end{aligned}
$$

Equation 2.45 thus becomes,

$$
\begin{bmatrix} m_{11}+m_{tip} & 0 \\ 0 & m_{22}+m_{tip} \end{bmatrix} \begin{bmatrix} \ddot{q}_1 \\ \ddot{q}_2 \end{bmatrix} + \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} \dot{q}_1 \\ \dot{q}_2 \end{bmatrix} + \begin{bmatrix} \tilde{k}_{11} & 0 \\ 0 & \tilde{k}_{22} \end{bmatrix} \begin{bmatrix} q_1 \\ q_2 \end{bmatrix} = \begin{bmatrix} F_1 \\ F_2 \end{bmatrix} \tag{2.50}
$$

For operation in air, the first eigenmode is generally sufficient to model the cantilever dynamics as mentioned in Melcher et al. (2008a). Therefore only the first eigenmode is used for the current research, which reduces the system of differential equations (2.50) to a single differential equation given below,

$$m_1\ddot{q_1} + c_1\dot{q_1} + k_1 q_1 = F_1 \tag{2.51}$$

where $m_1 = m_{11} + m_{tip}, c_1 = c_{11}$ and $k_1 = \tilde{k}_{11}$

The solution of this equation $q_1(t)$ gives the deflection of the cantilever tip $u(L,t)$. The force F_1 has two components, namely the tip-sample interaction force F_{ts} and force acting on the cantilever base due to the action of piezo-actuators F_b. Equation 2.51 thus becomes,

$$m_1\ddot{q_1} + c_1\dot{q_1} + k_1 q_1 = F_{ts}(d) + F_b \tag{2.52}$$

Since the cantilever behaviour can be approximated by a spring mass damper system, the force on the cantilever base F_b can be given as,

$$F_b = k_1(Y_1) \tag{2.53}$$
$$F_b = k_1(y_1 + Z_1) \tag{2.54}$$

where Y_1 is the *equivalent* cantilever base movement corresponding to the first eigenmode. y_1 and Z_1 are the components corresponding to the movement due to the dither piezo and the Z piezo as illustrated in Figure 2.5. Furthermore Melcher et al. (2007) shows that m_1, c_1, k_1 and Y_1 can be approximated as

$$\begin{aligned} m_1 &= \frac{m_c}{4} \\ c_1 &= \frac{m_1\omega_1}{Q_1} \\ k_1 &= 1.03k_c \\ Y_1 &= B_1 Y \end{aligned} \tag{2.55}$$

Here m_c, k_c and Q_1 are the cantilever mass, stiffness and quality. B_1 is referred to as a *modal parameter* and approximately equals 1.5 (Prakash et al. (2009)). The base excitation force can thus be further simplified as,

$$F_b = k_1(B_1 y + B_1 Z) \tag{2.56}$$
$$F_b = k_1 B_1 (y + Z) \tag{2.57}$$
$$F_b = k_1 B_1 (A_{base} \cos \omega t + Z) \tag{2.58}$$

where A_{base} and ω are the amplitude and frequency of the sinusoidal drive signal.

The final block diagram of the AFM system with $y(t)$, Z(t) and sample topography T(t) as input and tip height $q_W(t) = q_{W1}(t)$ as output can be drawn as illustrated in Figure 2.6. It must be noted that this block diagram represents the AFM system in the frame W. The topography signal is now written as a function of time. This is because, as the vibrating cantilever rasters above the sample surface the sample surface topography varies with time.

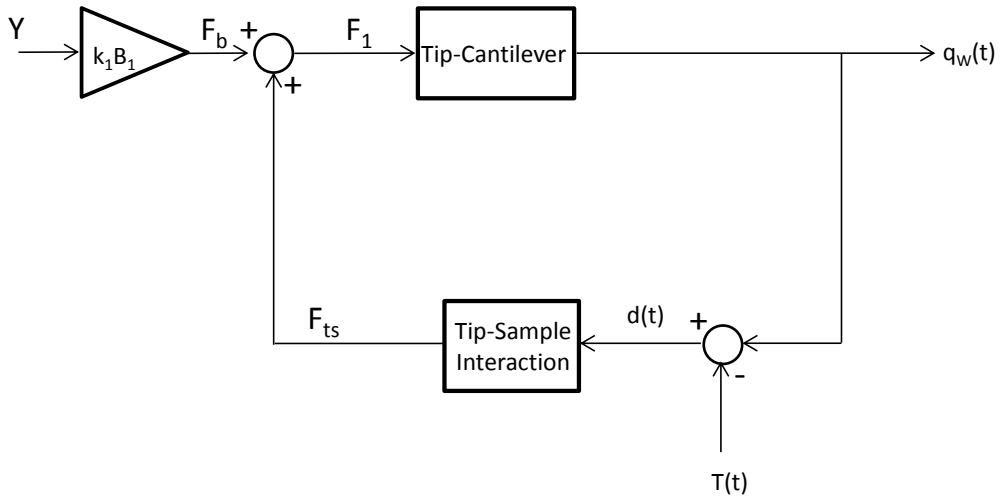

Figure 2.6: AFM block diagram.

2.3.1 Experimental Determination of Cantilever Transfer Function

The cantilever transfer function is determined simply by providing sinusoidal excitation signals to the dither piezo attached with the cantilever base and measuring the corresponding sinusoidal tip deflection. In this case 300 sinusoidal signals with frequencies ranging from 60 kHz to 70 kHz are used as input. Each signal has an amplitude of 0.6 V and lasts for 1 ms. The tip deflection d_{TM} is then passed through a lock in amplifier which extracts the amplitude $A(t)$ for each signal. All experiments in this section were performed on the CryoView 2000 AFM from Nanonics Imaging. Figure 2.7 illustrates the frequency responses of the cantilever when it is *far* from the surface. The term far

implies that the cantilever tip is at a sufficient distance from the tip so that the tip sample force is negligible.

This frequency response is next used to determine the transfer function for the cantilever. Since it is already known that the response of the cantilever can be modelled by a second order under damped LTI system with no zeros, the form of each transfer function is given by

$$G_c(s) = \frac{K\omega_n^2}{s^2 + \frac{\omega_n}{Q}s + \omega_n^2} \tag{2.59}$$

where K is the system gain, ω_n is the resonance frequency and Q is the quality factor. The quality factor is determined easily by finding the maximum amplitude A_{max}, the frequency at which the amplitude is maximized ω_p and the two frequencies ω_L, ω_H where the amplitude is $\frac{A_{max}}{\sqrt{2}}$. The quality factor is then simply $Q = \frac{\omega_p}{\omega_H - \omega_L}$. While this relation for quality factor holds true for a series RLC circuit that has an s term in the numerator of its transfer function. However since the current system has high Q (low damping) the relationship is still valid.

Next, the resonance frequency is determined by the relationships given below,

$$\begin{aligned} \omega_n &= \frac{\omega_p}{\sqrt{1-2\zeta^2}} \\ \zeta &= \frac{1}{2Q} \end{aligned} \tag{2.60}$$

Finally the system gain K is determined using the following relations,

$$\begin{aligned} M_p &= \frac{1}{2\zeta\sqrt{1-\zeta^2}} \\ K &= \frac{A_{max}}{M_p} \end{aligned} \tag{2.61}$$

The finally identified parameters are provided below in table 2.1.

Table 2.1: Identified cantilever transfer function parameters

T.F	ω_n [kHz]	Q	K
G_c	65.497	199.9	0.0421

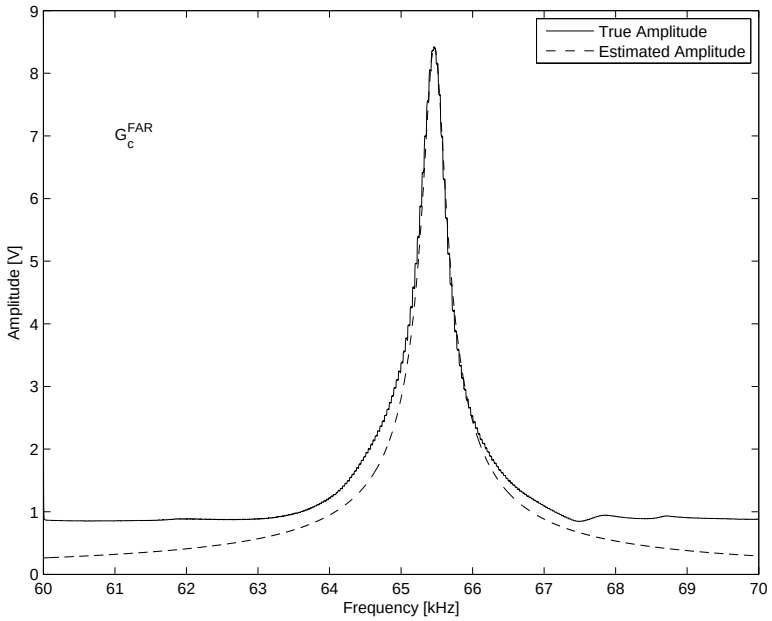

Figure 2.7: Cantilever frequency response *far* from the surface.

Figure 2.7 also plots the simulated response of the identified transfer function. It is clear that there is good agreement between the true and simulated cantilever behaviour.

Although the estimated amplitude does not match the true amplitude at greater separation from the resonance frequency, this difference makes no difference for the current application. This holds true because the cantilever is always excited very close to its first resonance frequency.

2.4 Piezo-Actuators

Piezo-actuators are made of materials which change shape in response to application of electric potential. This property of these materials is used to make piezo-actuators, which can generate linear motion upon application of a voltage signal. Since the operation of the AFM requires movement in three translational degrees of freedom, the piezo-actuators must be used in specially designed constructions to enable the necessary movement.

One possibility is to use a single tube piezo-actuator developed by Veeco (2006) (illustrated in Figure 2.8). This actuator can bend sideways to provide the two lateral degrees of freedom along the x y axis and can also expand and contract to provide movement along the z axis. These movements are achieved by applying suitable voltages to inbuilt electrode pairs, one for each degree of freedom. The Multimode AFM available at the University of Southampton also employs an AS-130 J tube scanner with an x y range of $125\mu m \times 125\mu m$ and a vertical range of $5\mu m$ Veeco (2006). The primary disadvantage

of this actuator is the coupling between the x y and the z directions. As the tube bends sideways z axis value is also altered. There are two ways in which this problem can be resolved. The first method is to use the knowledge of the geometry of the actuator and correct the acquired image through image processing (Abramovitch et al. (2007)).

The second method as illustrated in Figure 2.8 is to use an alternate actuation configuration in which movement along each axis is achieved by a separate actuator. Two of the actuators are placed inside a *nano-positioning stage* (also referred to as the x y *scanner*) for actuation along the x y axis and the z-actuation is done with a separate actuator.

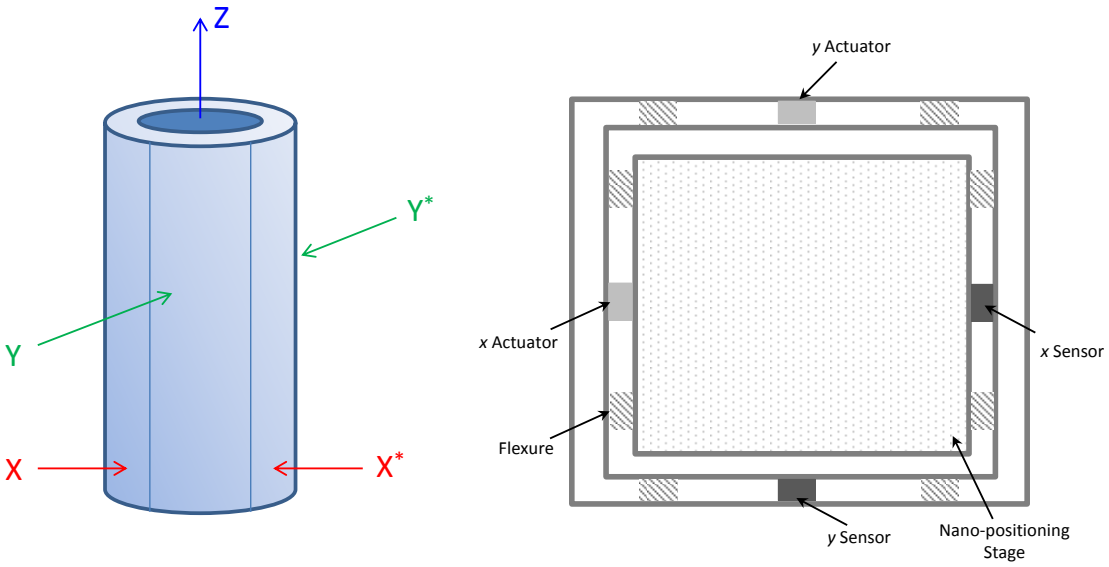

Figure 2.8: Piezoactuators i) tube actuator (left) (Veeco (2006)) ii) Nanopositioning stage (right)(Abramovitch et al. (2007))

The rastering of the sample is always done by providing a triangular waveform to the x axis actuator and a ramp to the other one. If a nano-positioning stage is being used the inner actuator is chosen to be the x axis since it must move a smaller mass and therefore is better suited for the triangular movement which has higher frequency content. This rastering pattern is illustrated in Figure 2.9. If a tube actuator is used rastering can be done by providing the same waveforms to the x and y electrodes respectively.

Each two consecutive scan lines are referred to as the *trace* and *retrace* lines. A trace line scans the sample from the bottom to top (or left to right) and a retrace scans in the opposite direction.

Although the rastering pattern just mentioned may appear trivial to execute however the speed at which this can be done is severely limited by the presence of two factors namely, i) vibrations ii) modelling uncertainty. The first problem occurs due the presence of resonant modes of vibrations in the piezo-actuators. The resonance frequency of the

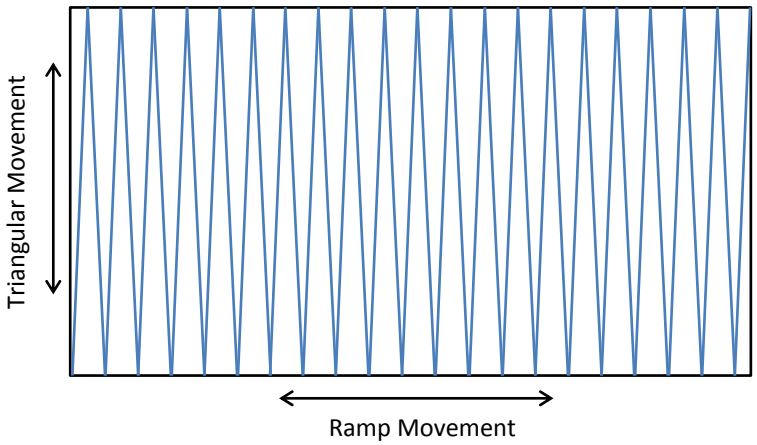

Figure 2.9: Raster pattern employed in AFMs

first vibration mode is typically in the range of 200Hz to a few tens of kHz (Devasia et al. (2007)). If the excitation signal contains a frequency that is close to the resonance frequency, the actuator begins to oscillate and as a result the raster pattern of movement is distorted. This in turn distorts the image that is generated by the AFM. An example of this distortion is provided through experimental data by Devasia et al. (2007) and is illustrated in Figure 2.10.

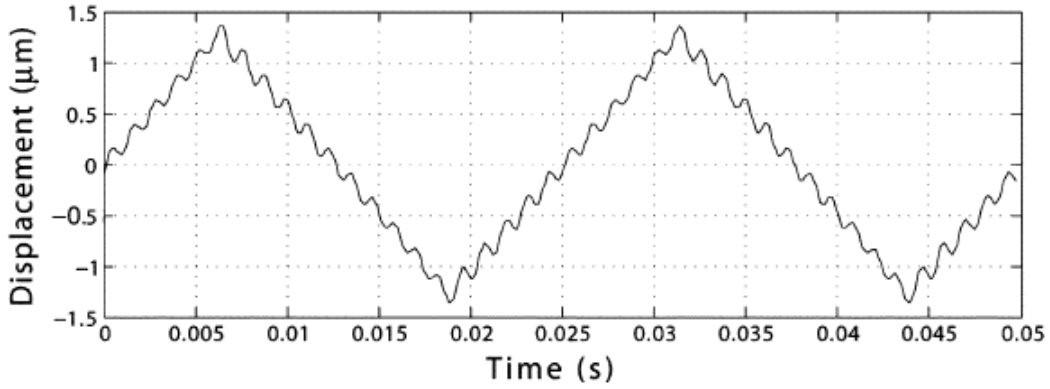

Figure 2.10: Distortion in ratering pattern as given in Devasia et al. (2007)

As can be seen clearly in addition to the triangular movement the actuator oscillates as well. In this case the actuator was driven with a triangular wave form with a frequency of 40 Hz. The first resonant frequency of the actuator is 850Hz. The distortions in this case are due to the 21^{st} and 23^{rd} harmonics of the triangular waveform which are close to the first resonance frequency.

Ideally, resonance in any plant may be cancelled by designing an appropriate controller, however this approach is limited by the second problem associated with piezo-actuators i.e modelling uncertainty (Abramovitch et al. (2007)). It is difficult to develop *a priori*

accurate models of a piezo-actuator (Devasia et al. (2007)). In addition to the uncertainty, the actuator dynamics change over time due to ageing effects which therefore requires that current AFM users must calibrate the piezo-actuators every month as mentioned in Veeco (2006). Due to these reasons the use of robust, adaptive and learning control methods has been widely investigated over the past decade.

Another factor that makes the use of such control methods difficult is the limitation on the maximum sample rate imposed by current hardware. Controllers can be designed which compensate even for the higher harmonics of the actuators, but as a rule of thumb according to Abramovitch et al. (2007) the sample rate must be 10-20 times greater than the highest dynamic of interest. The same author also also gives an example that if a controller needs to include a 300kHz resonance the sample rate must be 3-6 MHz. This imposes severe limitations on the complexity of the control algorithm.

The piezo-actuator stage used for the current research is the *3D FlatScan*^{TM} from Nanonics Imaging. The piezo stage consists of two piezo-actuators each for actuation in the lateral X and Y directions, namely X^+, X^- and the Y^+,Y^-. The pair of actuators for each direction are provided voltages with opposite polarity. Thus if one actuator expands the other contracts, causing movement in the direction of the expanding actuator. The same applies to actuation in the Z direction with three pairs of piezo-actuators. Figure 2.11 (left) illustrates the piezo-actuators. Figure 2.11 (right) illustrates the 3D view of the same stage and shows the position of the cantilever.

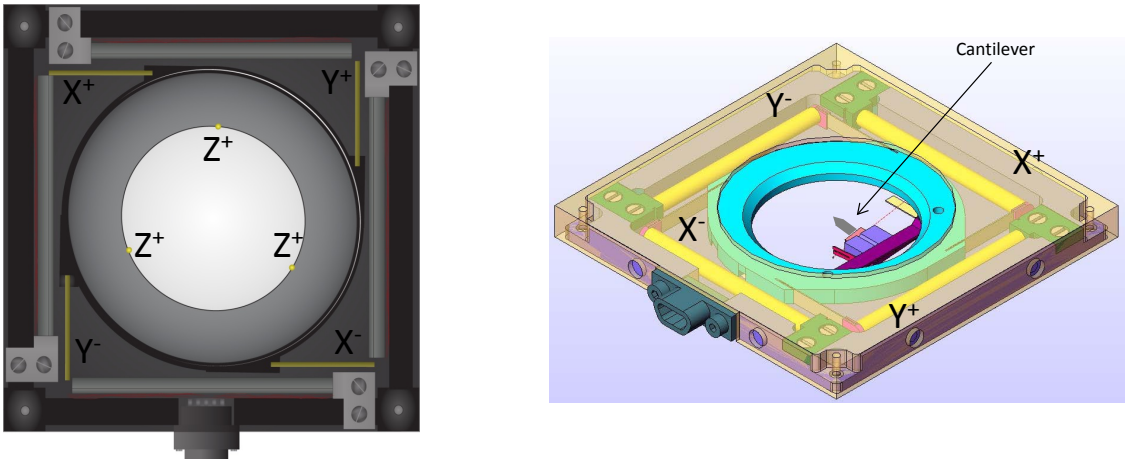

Figure 2.11: Nanonics piezo-actuator stage (Nanonics Imaging). Top view(left) 3D view(right)

This piezo-stage has no sensors for position measurement and is therefore run in open loop. As a consequence the piezo movement is calculated by multiplying the piezo input voltage with a calibration constant.

2.5 Optical Detection System

The optical detection system consists of a laser source and a four quadrant photo detector. As illustrated in the Figure 2.12 the purpose of having four quadrants namely A, B, C and D is to enable the detection of not just the transverse vibration of the beam but also lateral bending movement. The transverse vibration is detected by summing the top two and bottom two quadrant voltages and finding their difference. The difference is then divided by the sum of the voltages from all the quadrants to eliminate the effect of variation in laser intensity. The final deflection is obtained in the form of a normalized signal v_T as follows,

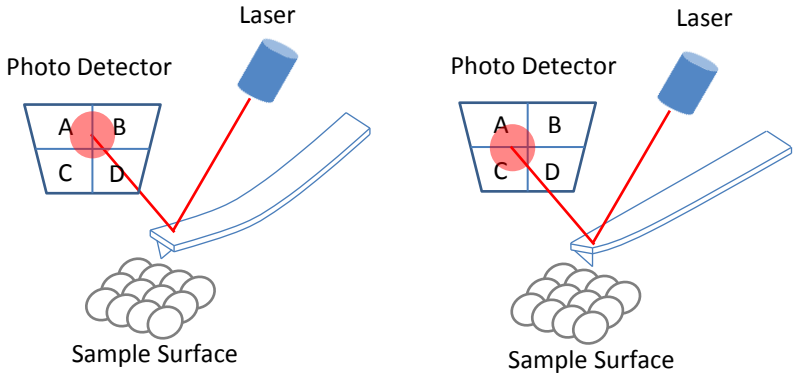

Figure 2.12: The AFM optical system for detection of transverse bending (left) and lateral bending (right).

$$v_T = \frac{(v_A + v_B) - (v_C + v_D)}{v_A + v_B + v_C + v_D} \tag{2.62}$$

The normalized signal for the detection of lateral bending can be obtained using,

$$v_L = \frac{(v_A + v_C) - (v_B + v_D)}{v_A + v_B + v_C + v_D} \tag{2.63}$$

The dynamic mode atomic force microscopy uses only the transverse bending of the cantilever. The lateral bending of the cantilever beam has been investigated for the determination of the surface friction of a surface (Caron et al. (2004)).

2.6 The AFM Control Loop

The complete AFM control loop will now be illustrated. The purpose of the controller is to regulate the amplitude of the cantilever vibration. If the topography of the sample

being rastered beneath the vibrating tip changes, this in turn changes the tip sample interaction force. As a consequence the amplitude of vibration is altered. The purpose of the controller is to regulate the height of the vibrating tip by changing the piezo-actuator input voltage h so that the amplitude of vibration remains the same. Finally the estimated topography $\hat{T} = k_Z \times h$, where k_Z is the calibration constant for the piezo-actuator with units of $\frac{nm}{V}$. This concept is illustrated in Figure 2.13.

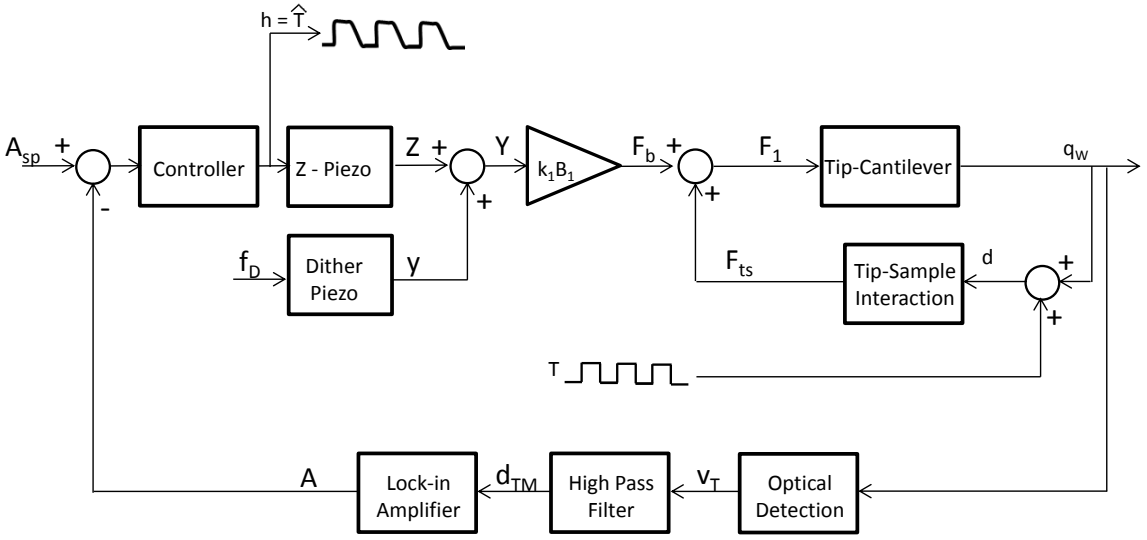

Figure 2.13: The AFM control loop.

The control loop illustrated in the figure has been chosen to generate simulation results for a commonly used AFM experimental set up, i.e Tapping mode atomic force microscopy. The parameters used for the simulation are illustrated in (Table 2.2). Here A_o is the amplitude with which the cantilever will vibrate when it is far enough from the surface so that the tip sample interaction force has no effect. The cantilever quality factor Q and resonance frequency $f_n = \frac{\omega_n}{2\pi}$ are determined using the system identification experiment explained in the previous section. These two parameters along with cantilever length, width and fluid density and viscosity of air are then used to determine the cantilever stiffness. This is done by using an analytical method referred to as the *Sader* method as mentioned in Sader et al. (1999).

The additional high pass filter is used because only the vibration amplitude of the cantilever deflection signal is of interest for the AFM operation. This is implemented using a first order high pass filter with a corner frequency equal to a tenth of the cantilever resonance.

The excitation signal is a sinusoid with amplitude A_{base} and frequency f. The excitation amplitude is chosen so that the amplitude of free vibration A_o equals 30nm. The frequency of the excitation signal is selected be equal to the cantilever resonance frequency so that the amplitude of vibration A exhibits maximum sensitivity to changes in

sample surface height. The sampling frequency is selected to be 5 MSPS (mega samples per second), which is more than 75 times the resonance frequency. Thus the sampling time T_s is $0.2\mu s$. This excessive sampling frequency is motivated purely through empirical observation and other contributions reporting similar simulation results for instance Melcher et al. (2008a). A proportional integrator (PI) controller is used in the loop. This is motivated by the fact that a large number of commercial AFMs use the same controller and it has been found to be sufficient for sample topography estimation. The exact values of the controller gains K_p and K_i are selected using an exhaustive search procedure described in Section 3.2.2. It is assumed that the cantilever material is Silicon and the sample is made of fused silica SiO_2. The material properties for these substances are obtained from Stark et al. (2004). This control loop was simulated for a sample with a step profile with a feature height T_h of 120 nm. The set point A_{sp} for the controller is set at $0.8A_o = 24nm$. Generally the set point amplitude is selected to be *slightly* below A_o. If the set point is selected to be a small fraction of A_o, this will cause the cantilever tip to be closer to the surface. In a physical AFM system this can cause sample damage if the sample is soft, or wearing of the tip if the sample is hard.

Two assumptions are made in the simulation, i) the piezo-actuator and optical detector dynamics are neglected i.e, they have unity transfer functions ii) the optical measurement noise is neglected. The piezo-actuator dynamics can be neglected because in the current AFM the scan rate is very low, i.e only 1 scan line per second or 0.5 Hz. Furthermore, this model is only applicable for the case where both the lateral and vertical movements are accomplished by the piezo-actuator attached with the cantilever base. This set up is one of the two common AFM set ups used in practice. In the second one, the sample is placed on a piezo stage. The stage then moves the sample in the lateral and vertical directions. The cantilever is attached to a separate dither piezo. Therefore, in one set up the cantilever is moved, whereas in the other one the sample is moved. The model used in this research can be easily modified for either case.

Figure 2.14 illustrates the true and estimated topography. The estimated topography is simply the controller voltage V_Z. Since the current implementation neglects piezo-actuator dynamics the calibration constant is $1\frac{nm}{V}$. As it is clear that the control signal can provide an approximation of the true surface, however the accuracy depends on the tuning of the controller.

The next Figure 2.15 shows the true and the set point amplitudes. As can be seen the true amplitude oscillates around the desired level however substantial deviations occur when the topography abruptly changes.

Parameter	**Symbol**	**Value**
Cantilever		
Amplitude of free vibration (nm)	A_o	30
Cantilever stiffness(N/m)	k	40
Cantilever oscillation mode	i	1
Quality factor	Q	180
Resonance frequency(kHz)	f_n	65
Cantilever transfer function gain	K	1
Modal parameter	B_i	1.5
Scan		
Excitation amplitude(nm)	A_{base}	0.10
Excitation frequency(kHz)	f	65
Sampling frequency(μs)	T_s	0.2
Scan time (s)	T_{scan}	0.05
Scan length (nm)	L_{scan}	10
Proportional gain	K_p	0.04
Integral gain	K_i	0.007
Tip-Sample interaction		
Hamaker constant(J)	H	0.2×10^{-18}
Tip radius(nm)	R	20
Elastic modulus (tip) (GPa)	E_{tip}	130
Elastic modulus (sample) (GPa)	E_{sample}	1.2
Poison ratio (tip)	ν_{tip}	0.3
Poison ratio (sample)	ν_{sample}	0.3
Inter-atomic Separation (nm)	a_o	0.3
Sample topography		
Feature profile		step
Feature height (nm)	T_h	120

Table 2.2: AFM simulation parameters.

2.7 Literature Review

Feedback control mechanisms are integral for the functioning of atomic force microscopes. The most important function is the regulation of the vertical tip sample separation or the cantilever vibration amplitude. As a consequence, ever since the development of the first AFM in 1986, feedback mechanisms have been developed and investigated for this purpose. However initially, no controllers were used for regulating the piezo stages that raster the sample laterally beneath the cantilever. This was instead done in an open loop manner (Agarwal and Salapaka (2008)). With availability of greater computational power at the start of the last decade control theorists began applying modern control methods e.g, robust, model based inverse control and adaptive control for lateral control as well. The entire body of AFM control literature can be classified into two groups, *lateral* and *vertical* control. In the first case the controller is responsible for lateral rastering movement of the X and Y piezos. The second group focusses on

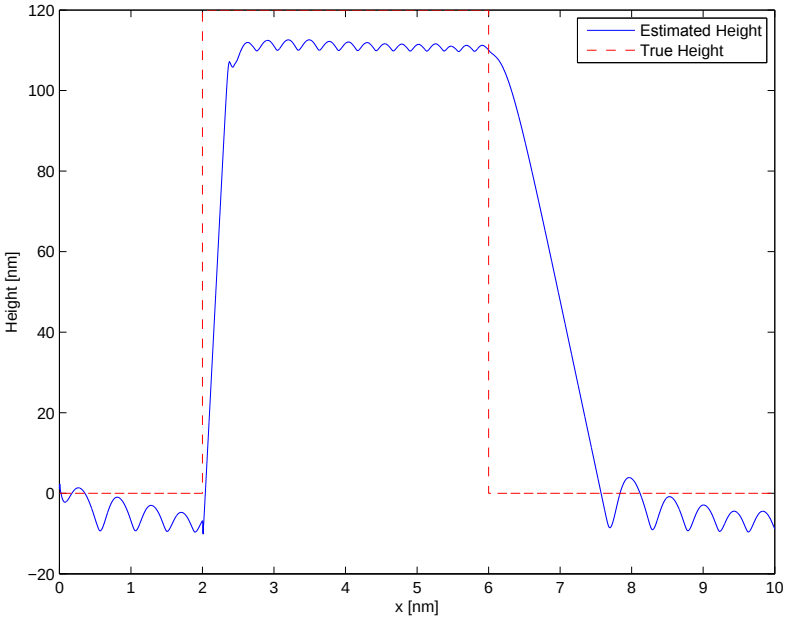

Figure 2.14: Comparison of true vs estimated topography.

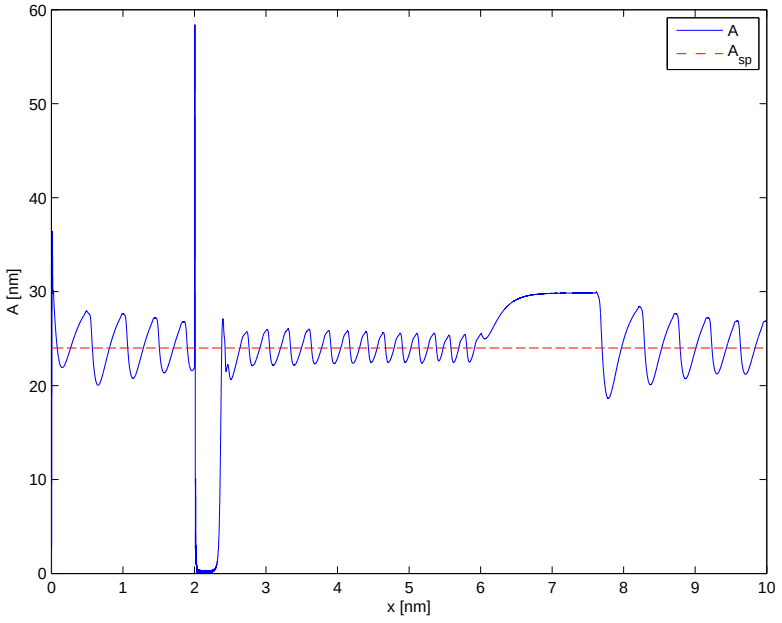

Figure 2.15: True amplitude (A) vs set point amplitude (A_{sp}).

control strategies developed for regulating the cantilever tip deflection or amplitude by controlling the Z piezo movement. The next section explains the performance measures used as benchmarks, followed by a description of the control strategies themselves in the remaining two sections.

2.7.1 Performance Measures

As mentioned in Chapter 1, the most important limitations of AFM performance are high scan time, and the difficulty of use. These result in two objectives for any AFM control design.

The first objective is to generate an image faster. This translates directly into the scan rate of the lateral X, Y piezos. In control theoretic terms this is the closed loop bandwidth of the X, Y piezo position control system. The higher the bandwidth, the greater will be the scan rate and the lesser the time needed to scan a sample.

The second objective is to develop an automatic controller generation mechanism which results in a controller that maximises the accuracy of the topography estimate. Although a number of methods have been suggested for controller synthesis, these contributions would need some a priori knowledge of control theory from the end user. In addition they would be difficult to retrofit around existing commercial AFMs. The accuracy can be expressed as the squared 2-norm of the estimation error $||T - \hat{T}||_2^2$, which should be minimized.

While this measure is ideal for measuring estimation accuracy, it is practically impossible to use. The reason is that it needs complete knowledge of the true topography T. An alternate measure is therefore generally used, which is the error in the quantity being regulated. For the case of contact mode, it is the squared 2-norm of the difference between the set point PSD (photo sensitive diode) voltage and the true PSD voltage, i.e $||e||_2^2 = ||v_{sp} - v_T||_2^2$.

For the case of dynamic mode it is the squared 2-norm of the difference between the set point amplitude A_{sp} and the actual amplitude A, i.e $||e||_2^2 = ||A_{sp} - A||_2^2$. While it is easily measured, the squared 2-norm of the error is not the best metric since a low value of this metric does not necessarily correspond to a better quality image. This is because it does not take into account the oscillatory nature of the sample image or the difference between the trace and retrace scan lines. Ideally there should be no controller induced oscillations in the control signal and the trace and retrace signals should be as similar as possible.

The next two sections explain the control strategies developed so far for lateral and vertical control. The last section concludes the literature review and highlights the aspects of atomic force microscope control that need to be investigated further and are the subject of this thesis.

2.7.2 Lateral Control

The lateral control strategies focus on actuation problems associated with piezo-actuators. As mentioned previously, there are two performance objectives namely tracking bandwidth and the positioning accuracy. The tracking bandwidth determines the highest frequency up until which the piezo-actuator can follow a reference positioning signal. Higher bandwidth implies faster scanning and lesser imaging times. The biggest difficulty in this context is the existence of resonant modes in the pizeo-actuator as mentioned in Section 2.4. If the exact transfer functions of the piezo-actuators are known, it would be possible to generate controllers that cancel these resonant modes. However since there is always an uncertainty in the piezo-actuator transfer functions this is difficult. Furthermore, the presence of *Creep*, *Hysteresis* and position sensor noise have a detrimental effect of positioning accuracy. The creep phenomenon corresponds to the gradual drift in piezo position from its steady-state position. This means that if a step input is provided the piezo will first quickly move towards a steady state position, but then instead of remaining there it will continue to drift away slowly.

One possible control design tool that provides a systematic method for achieving the above mentioned performance objectives is H_∞ control, which has been investigated by a number of researchers. This review provides one example (Lee (2010)) that explains how the H_∞ framework is applicable for lateral control. This contribution was selected because it provides a systematic procedure that is generic and is a good representation of of the work reported in similar contributions. Figure 2.16 illustrates the nano-positioning control loop as reported by the same author Lee (2010).

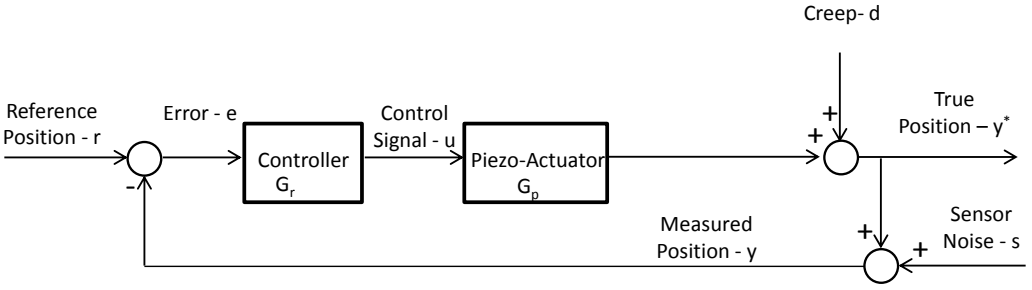

Figure 2.16: Nanopositioning control loop.

The author uses the plant G_p to model resonance and hysteresis, whereas creep is considered to be an external disturbance d. The plant model G_p is obtained by generating the frequency response of the piezo-actuator with the piezo input voltage as input and position sensor voltage as output. This modelling method has been followed in general in all contributions that suggest controllers for piezo-actuators. Here G_r is the suggested robust H_∞ controller.

The true position y^* is then measured but it is corrupted with sensor noise s and results in noisy measurement y. The true system error e^* can then be defined as,

$$\begin{aligned} e^* &= r - y^* \\ e^* &= r - (\frac{G_r G_p}{1 + G_r G_p} r - \frac{G_r G_p}{1 + G_r G_p} s + \frac{1}{1 + G_r G_p} d) \\ e^* &= r - (Tr - Ts + Sd) \end{aligned} \tag{2.64}$$

Where $S = \frac{1}{1+G_r G_p}$ is the *Sensitivity Function* and $T = \frac{G_r G_p}{1+G_r G_p}$ is the *Complementary Sensitivity Function*. Then using the fact that $T + S = I$ where I is the unity transfer function we get,

$$\begin{aligned} e^* &= r - Tr + Ts - Sd \\ e^* &= r(I - T) + Ts - Sd) \\ e^* &= Sr + Ts - Sd \\ e^* &= S(r - d) + Ts \end{aligned} \tag{2.65}$$

The control design objective then is clearly to minimize e^*. This then translates into designing controllers that minimize S in the frequency range corresponding to the reference signal r and creep disturbance d, and minimizes T in a frequency range corresponding to sensor noise s. Furthermore since the sensitivity function also equals $\frac{dT/T}{dG/G}$ the peak of the sensitivity function $||S||_\infty$ corresponds to the robustness of the closed loop system to modelling uncertainty. Therefore another objective of the design procedure is to minimize the peak of the sensitivity function. The author then describes in detail how a controller can be designed that meets these requirements.

The next immediate extension to using feedback controllers is to add a feed-forward controller as illustrated in the figure below,

As reported in Lee and Salapaka (2009), while the feedback controller provides robustness the feed-forward component allows the designer to overcome the algebraic constraints associated with designs that have only a feedback controller. The author in this contribution again uses the H_∞ frame work to design both feedback and feed-forward controllers which enable faster operation of the piezo-actuators. Both approaches are finally compared in the same contribution. While the feedback only control enables a closed loop bandwidth of 49.4 Hz, the addition of the feed-forward controller improves the bandwidth to 214.5 Hz.

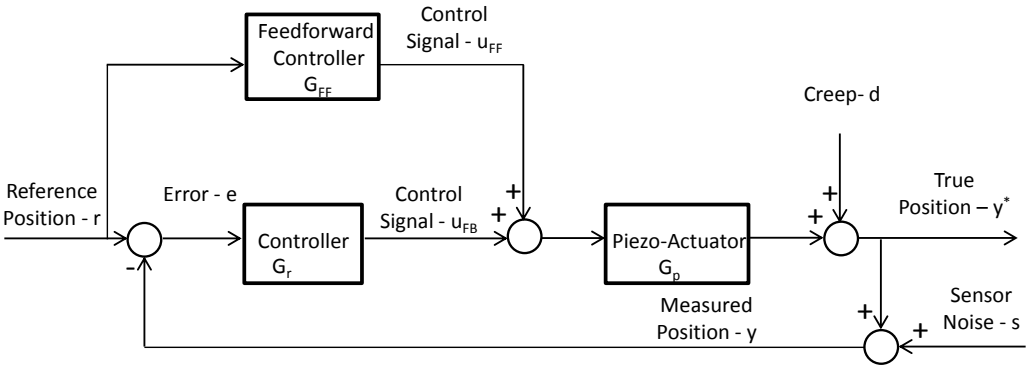

Figure 2.17: Nano-positioning control loop with feed forward controller.

An alternate approach to design the feed-forward controller is to use the inverse model of the piezo plant dynamics as reported in Clayton et al. (2009). The author reports scan rates of up to 445 Hz using this method. This however leads to a number of problems including those arising from model uncertainty, non-minimum phase plants and plants with an improper inverse. These have been discussed in detail in the same contribution.

The third major direction adopted by researchers is *Iterative Learning Control*. The major motivation behind this is the fact that in most AFM applications the scanning trajectory is periodic. According to Devasia et al. (2007) and Devasia et al. (2007) ILC controllers have been used to control both piezo-actuator hysteresis and vibration.

2.7.3 Vertical Control

This section provides a review of the various control methods that have been developed for regulating the height of the cantilever base while the sample is being rastered underneath the cantilever tip. The performance measure in this context is that the estimated topography $\hat{T}$ approximates the actual topography T as closely as possible. While it would be desirable to have a single number i.e, $||\hat{T} - T||_2$ to compare the effectiveness of the methods, most contributions report the results graphically. This is done by testing the proposed solution on a priori known samples (generally calibration grids), and plotting the estimated topography. Furthermore, numerical comparisons would be meaningful only if all the contributions imaged the same sample using the same experimental setup which is not the case.

Before the control design can proceed the AFM system must be modelled. The entire AFM operation in the vertical direction can be adequately described by modelling three components i.e, the piezo-actuator connected to the cantilever base, the cantilever and the tip-sample interaction. The piezo-actuator is modelled using the same procedure described in the previous section. The cantilever movement is nearly always modelled

as a linear combination of its eigenmodes. The tip-sample interaction is always modelled using either the Lennard-Jones potential or the Derjaguin-Muller-Toporov (DMT) model.

Once a model for the AFM system is obtained, a number of control methods can be applied for Topography estimation. These range from standard methods like H_∞ and robust adaptive control to elaborate ones which use Poincaré sections for controller generation. The remaining sub sections provide a review of the modelling procedures and controller synthesis methods reported for control of vertical dynamics in contact mode and tapping mode.

Contact Mode : The simplest method to model the vertical dynamics in contact mode is to determine the transfer function between the vertical piezo input voltage and the PSD output. An example of this method is provided by Sebastian (2004). The tip sample interaction is then modelled as a disturbance acting on the system. The author then uses an H_∞ norm minimization framework to design the controller. A similar modelling process is used by Fleming (2009) followed by the design and analysis of a simple integral controller.

An alternative method for making the non-linear AFM model usable for controller generation is to model only the first eigenmode of the cantilever as a spring mass damper system and linearise the non-linear tip sample interaction force around an operating point as has been done in Rifai and Toumi (2005). A slightly different method is used by Rifai and Toumi (2007) who first derives the transfer function between the piezo vertical input voltage and the PSD output using elementary bending theory Rifai (2007) and then identifies the transfer function coefficients using an experimental procedure that involves sinusoidal excitation of the scanner and approach and retract curves. The author then proceeds to design the controller using a robust adaptive method.

A similar procedure is used by Fujimoto and Oshima (2008a) where the author first constructs a spring mass damper system equivalent to the cantilever and the tip-sample interaction force and then determines the transfer function from the sum of the piezo height and topography to the PSD voltage. The system parameters are then identified using the corresponding frequency response. The same procedure is repeated in Fujimoto and Oshima (2008b) with the exception that a *pseudo disturbance* identification input is used to construct an ARX model. These models are then used for topography reconstruction using an observer.

Apart from these conventional approaches a different set of methods is followed by a few researchers and these are described briefly in the remaining part of this section. Krstic et al. (2006) suggests using a *Shear Beam* model for the cantilever instead of the commonly used *Euler Bernoulli* one. The author then suggests back stepping boundary control along with change of variables that converts the beam model into a wave equation. According to the contribution this transformation allows easy stability analysis

and also makes achievement of the desired performance easier. This is the only exception where the dynamics of the beam are not explained using a linear combination of its eigenmodes.

An entirely different approach is used by Li et al. (2009) where the cantilever is assumed to be clamped to the piezo-actuator on and *pinned* to the surface. Instead of being modelled as a *clamped free* beam, the cantilever is modelled as a *clamped pinned beam*. The surface is then modelled as a linear combination of sinusoids.

In addition to these contributions there are two more that propose models for the vertical dynamics in contact mode and discuss the dynamics exhaustively. These however do not use the models for control. Stark et al. (2004) proposes a general state space model for the cantilever dynamics, which contains two states for each mode. The tip-sample interaction force is linearised at two points, one for the repulsive and attractive regions each. The dynamics of the system are then discussed for both regions. Lastly Vazquez et al. (2006) presents a method which starts by modelling the cantilever beam using the Euler Bernoulli equation as usual, places the linearised tip-sample interaction force as a boundary condition for the free end and then determines the transfer function between the force per unit length of the cantilever and the cantilever slope at the free end.

Tapping Mode : The method employed for AFM vertical dynamics in tapping mode is similar to the one used for the contact mode. All the contributions start by taking into account the nonlinear tip-sample interaction. They then proceed to simplify the transient behaviour of the cantilever vibration amplitude using either a linearisation procedure or a saturation. The factor that makes these simplifications feasible is that as the vibrating cantilever approaches the sample surface, the vibration amplitude decreases in a *nearly* linear manner. Likewise the amplitude increases as the cantilever moves away from the surface.

The simplest approach has been used by Necipoglu et al. (2010). The author considers the transfer function between the sum of topography and piezo height to the vibration amplitude to be unity. The reason stated for this simplification is that the amplitude response is linear and its dynamics are much faster than the piezo stage. Next, a repetitive controller is designed. This controller makes use of the fact that two scan lines over the sample surface are nearly always similar.

In a similar contribution Zhang et al. (2010) the transfer function between piezo input voltage and the PSD deflection is first approximated using a frequency response method. A saturation non-linearity is then added to model the fact that the oscillation amplitude will be zero if the tip sample separation is zero, and amplitude will saturate if the tip sample separation is greater than a threshold. The threshold in this case equals the amplitude of free vibration. This model is then used to design a robust adaptive controller for regulating the cantilever height. A similar model has been used by Shiraishi and Fujimoto (2009) to design an observer based controller.

A more systematic approach is adopted by Lee (2010) who models the amplitude dynamics using the Krylov-Bogoliubov-Mitropolsky (KBM) procedure and then attempts to use a linearised version for estimating surface. However the author notes in the dissertation that the model does not perform well at high frequency and finally uses the DC gain for the linearised model for controller generation. Another systematic approach is used by Misra et al. (2008). This contribution starts by modelling the AFM dynamics using Poincaré sections, followed by linearisation and pole placement using a Luenberger observer.

Apart from these contributions there is one that proposes a control scheme that eliminates the requirement for any simplification of the amplitude dynamics. Yagasaki (2010) suggests setting the cantilever base velocity proportional to the instantaneous difference between the desired tip position and the true position. This, according to the author enables control of cantilever vibrations and also measurement of surface topography. Finally, there are two contributions that propose complete elimination of any modelling requirements. Yamasue and Hikihara (2009) suggests a *Time Delayed Feedback* in which the control signal is proportional to the change in tip velocity between two consecutive time steps. Jeong et al. (2006) suggests zero order predictor to regulate the tapping impulse strength every cycle.

2.7.4 Alternate Methods for Faster Imaging

In addition to the design of better control mechanisms, many researchers have investigated a number of alternate methods to increase the scan rate. The reduction in imaging times has significant consequences, specifically in the study of biological specimens like DNA. Conventional imaging times vary from a few seconds to a few minutes, however the DNA molecules *translocate* in excess of 10 times per second (Anderson (2007)). Translocations are mutations in which chromosome segments change position (Cornell University - Department of Animal Science). This means that while high resolution images can be obtained, the slow imaging speed still constitutes a bottle neck.

The first possibility that has been investigated is to design piezo-actuators that have a higher first resonance frequency. This in turn allows the rastering movement to be performed faster, thus reducing imaging time.

One such example is provided by Schitter et al. (2006). In this contribution the authors design a piezo actuation stage with a first resonance frequency above 22 kHz. The high resonance enables scan rates of up to 1030 lines per second. They achieve this by making the scanner as compact and rigid as possible. The trade off is that the scan size is limited to $13 \times 13 \mu m$.

In this context, a novel approach to improving the scan speed is to use a completely different actuator. Picco et al. (2007) suggests using a tuning fork for moving the sample

along one lateral degree of freedom. In this set up, the sample is placed on an arm of the tuning fork. The base of the fork is connected to a separate piezo-actuator. The tuning fork is vibrated and simultaneously, the second piezo-actuator moves the fork base. The cantilever is connected to a separate tube piezo-actuator and is positioned at a fixed location above the sample. Since the tuning fork has a high first resonance frequency up to 100 kHz, the movement along one lateral degree of freedom can be done at high speed. The second piezo-actuator connected to the fork base simply serves to drag the sample beneath the cantilever. The authors report they were able to scan an area of about $500 \times 500nm$ at 1000 frames per second.

The second possibility that researchers have investigated is to use specially designed scan trajectories to reduce imaging time. The problem with the conventionally used raster scan trajectories is that they attempt to visit every single point on the sample and in addition require abrupt direction changes at the extremities of the scan region.

Since the raster scan pattern is exhaustive, it consumes a significant amount of time to generate an image. An immediate solution is to use trajectories that minimize the number of points scanned and construct the image from this minimal information. One possibility is to use any coarse scan pattern feasible and estimate the sample image using *Compressive Sensing* (Song et al. (2011)). This relatively new method from the domain of signal processing can estimate a signal even when it has been sampled at sub Nyquist frequencies. The only constraint is that the signal should be either *sparse* itself or in any transformed domain e.g the DCT (Discrete Cosine Transform) domain. Thus if the sample has a periodic nature, i.e if it has repeating patterns, it is possible to drastically reduce the number of points scanned and still construct an image.

The exhaustive raster scan can also be avoided by using a priori information about the sample. For instance if the sample is *long and string like*, as in the case of DNA strands, a *curve tracking* scanning method can be devised so that the cantilever moves only over the strands. This possibility has been investigated by Anderson (2007) with potentially a 25 times reduction in imaging time. The approach however does require a priori sample knowledge and is suitable for specialized applications.

Lastly, the second problem associated with traditional raster scanning is the abrupt direction reversals at the scan area extremities. This is problematic because abrupt changes in commanded motion direction excites higher resonances in the piezo-actuators. In this context Fleming and Wills (2009) suggest optimal scan trajectories which have minimal frequency content near the resonance frequency of the piezo-actuator. In the time domain these trajectories appear visually very similar to the original raster trajectories, however they have *softer* direction reversals.

Even though non-raster trajectories reduce imaging time, they make controller tuning difficult. Conventional AFM controller gain tuning methods require the end user to observe the trace and retrace scans and tune the gains so that the two would be as

similar as possible. This is intuitive since generally the user has some a priori knowledge of the sample shape. In case of non-raster scanning it is clearly difficult to compare trace and retrace scans. In this context it will be clearly feasible to use an automated tuning mechanism.

The third possible solution for reduction in imaging time is to use the a priori information that is generated during the scan itself. The fact that is used here is that consecutive scan lines in any sample image are very often *similar*. This fact is exploited by Wu et al. (2009). The authors suggest an Iterative Learning Controller for controlling the Z actuator. Since the sample that they use has minimal change between two scan lines, they use the information in one scan line to reduce the number of iterative learning control iterations for the next scan line. The authors report an eight times speed up over commercial AFMs using this method.

2.7.5 Conclusion

While a considerable amount of research has been done for making AFMs faster and more accurate, the methods proposed so far have two major limitations. Firstly, each of these contributions develop their solutions for a particular cantilever sample pair and do not take into account that both cantilevers and samples are regularly changed. In addition AFMs being a versatile tool are operated in different operation modes and mediums (e.g air, liquid and vacuum). Therefore AFMs inherently have a *multiple model* nature. It is clearly difficult to develop a single robust controller or an inverse model based feed-forward controller etc. which can take into account the vast variations in an AFM experiment.

The second limitation is that the controller schemes suggested are often complex. This means that their implementation requires considerable effort. As an example the robust H_∞ suggested in some of the contributions often results in high order controllers. Most commercial AFMs in contrast, use very simple PI controllers for image generation. While the sophisticated high order controllers may offer speed advantages over the simple PI controller, it will be difficult for AFM end users to understand and implement the H_∞ controllers. This holds true because a large number of end users have little or no knowledge of controller synthesis and its implementation.

This problem has received limited attention in literature so far. One exception is the work done by Burns (2010). The author's research aims to make AFMs *user independent*, and suggests two novel methods for this end. The first one is to use *Active Cantilevers*. These cantilevers are thermally excited and their movement is sensed using piezo resistive material on the cantilever. This eliminates the requirement for an optical position detection method, which in turn removes the burden of positioning the laser spot on the cantilever tip and alignment of the photo diode. While this is not directly related to

controller design, this innovation does make the usage of AFMs easier. Next, the author suggests a *just in time* system identification method for piezo-actuator dynamics. In it the actuator dynamics are estimated from the sample image as it is being obtained. The model is then used to compensate for the scanner resonance using a feed-forward model inversion method. The key advantage of this strategy is that no sensors are required to measure piezo-actuator movement, and can thus be implemented on existing AFMs. The limitation however is that the resulting controller can have an order up to six.

While the methods suggested above are promising, it is the observation of this research that a multiple model adaptive control (MMAC) algorithm *fits* the atomic force microscopy control problem very well in general. A MMAC algorithm would be one which uses a bank of candidate plant models and switches in a controller corresponding to a plant which best represents the current AFM experiment. The plant set would consist of plants that represent selected combinations of samples, scanning probes, operating modes and mediums. Furthermore, given the common usage of PI controllers in commercial AFMs, it will be feasible to initiate the development of the MMAC algorithm with PI controllers as well.

The final implementation will only need to measure the AFM plant inputs and outputs, possibly the Z and dither piezo input voltages and the photo-diodes output. The algorithm will then determine the *closest* plant in the plant set and switch in the corresponding controller. The closeness of a candidate plant in the plant set to the true AFM plant is measured using Kalman filter residuals and is described in detail in the next chapter. The complete mechanism therefore imitates a human user, who constantly observes the image being generated and tunes the controller gains manually to maintain the image quality.

An added advantage of this method this that, while the PI controller generating the Z piezo voltage signal needs to be implemented in real-time, the MMAC algorithm does not need to be implemented in real- time. It can run in parallel on a desktop PC with sufficient processing power. It does not have to be real- time since the controller gains need to be tuned once for a particular AFM experiment. Once the right gains have been found they suffice for the generating the entire sample image. If the sample exhibits substantial material property variation in the region being imaged, the gains might need to be updated a limited number of times.

To conclude, the suggested multiple model control algorithm acts as a *recommender* system. If the AFM hardware permits, the recommended gains can be switched in automatically. If this is not possible the user can manually update the recommended gains. This algorithm can therefore readily be *wrapped around* most commercially available AFMs and can make the usage these instruments easier.

Lastly, there is no restriction to use this method only with PI controllers. In the future if AFM manufactures switch to more complex control designs, the same algorithm can

be augmented to incorporate robust controllers. This could lead to a reduction of the plant and controller sets, and reduction in computational requirements.

Given the significance of reduction in imaging time, this research also investigates two novel methods for faster AFM operation. The first one is to generate images using coarse scan trajectories and compressive sensing. The second one is to use the entire length of the cantilever beam for image generation, rather than just the tip. Both methods are described in detail in Chapter 5.

Chapter 3

Automated Controller Generation

Now that the AFM system dynamics are known, it needs to be determined exactly how the Estimation based Multiple Model Switched Adaptive Control (EMMSAC) method can be used for automating the controller parameter tuning. Section 3.1 describes the functioning of the EMMSAC algorithm, followed by Section 3.2 which describes the application of EMMSAC for the control of a simulated AFM. In an actual work environment such a mechanism can be used to suggest possible controller gains to the user thus reducing the time spent in trial and error tuning.

3.1 Estimation Based Multiple Model Switched Adaptive Control (EMMSAC)

The basic concept behind the EMMSAC control method is that it consists of two sets, a plant set and a controller set. The plant set consists of candidate plants, each of whom could be a potential approximation of the true plant which needs to be controlled. The controller set consists of one controller for each plant in the plant set. These controllers are selected using any suitable controller synthesis procedure for instance LQR, pole placement. The method used for generating controllers is an exhaustive search procedure described in detail in Section 3.2.2. Once the sets are known, the input output data from the true plant is recorded. This data is then used to select a plant from the plant set which is *closest* to the true plant. The closeness is measured by a *Closeness* function which is described in detail in Section 3.1.4. Once the closest plant has been selected, the corresponding controller is switched into the control loop.

The remaining parts of this section describe the EMMSAC algorithm.

3.1.1 Preliminaries

For the purpose of this thesis, all signals are defined in discrete time i.e all quantities are measured at fixed time intervals in the set of integers $\mathbb{Z}$. A natural extension is the concept of a signal space given below,

$$\mathcal{S} := map(\mathbb{Z}, \mathbb{R}^h)$$

where $h \in \mathbb{N}$ which is the set of natural numbers, and $\mathbb{R}$ is the set of real numbers.

Since it will at times be the case in that only the initial portion of a signal is of interest or only a window of finite length, two operators need to be defined namely the *Truncation Operator* and the *Restriction Operator*. These are defined below.

The truncation operator is defined as,

$$\mathcal{T}_t v(\tau) = \begin{cases} v(\tau) & if \quad 0 \leq \tau \leq t, t \in \mathbb{Z}, \quad v \in \mathcal{S} \\ 0 & otherwise \end{cases} \tag{3.1}$$

Thus the operator returns the signal values up to time $t \in \mathbb{Z}$ and is zero every where else.

The restriction operator $\mathcal{R}_{\sigma,t} : \mathcal{S} \to \mathbb{R}^{h(\sigma+1)}$ extracts a window of finite length from $\sigma \geq 0$ as follows,

$$\mathcal{R}_{\sigma,t} v := (v(t-\sigma), \ldots, v(t)), \quad v \in \mathcal{S} \tag{3.2}$$

thus the operator extracts a window of length σ from the signal v.

Although the objective is to control a system where all the signals will have a bounded 2-norm, this cannot be assumed a priori. Therefore two signal spaces are defined $\mathcal{V}$ and $\mathcal{V}_e$. The l_2 norm for a vector $v \in \mathbb{R}^n$ is defined as,

$$||v||_2 = \sqrt{v_1^2 + v_2^2 + \ldots v_n^2} \tag{3.3}$$

For $\mathcal{V} \subset \mathcal{S}$, let

$$\mathcal{V} := \{a \in \mathcal{S} | a(-t) = 0, \forall t \in \mathbb{Z}; ||a||_2 < \infty\}$$

Here $\mathcal{V}$ does not include a signal $v \in S$ such that $||v||_2 = \infty$. For such signals where $||\mathcal{T}_t v||_2 \to \infty$ for $t \to \infty$ a new signal space $\mathcal{V}_e$ is defined. However it is still required

that for any finite $t < \infty$, $||\mathcal{T}_t v||_2$ is bounded. The extended signal space is thus defined as,

$$\mathcal{V}_e := \{v \in \mathcal{S} | \, \forall t \in \mathbb{Z} : \mathcal{T}_t v \in \mathcal{V}\}$$

The extended signal space thus includes signals that can grow without bound, however they are bounded for a finite time horizon.

The notation developed so far is now applied to a feedback control system illustrated below.

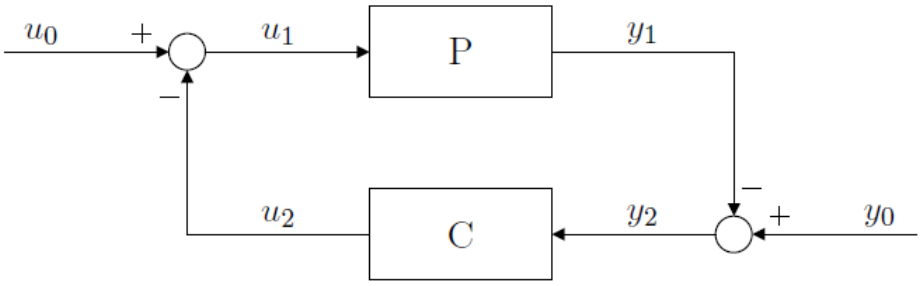

Figure 3.1: Control loop.

Let $m \in \mathbb{N}$ be the dimension of the input space, and $o \in \mathbb{N}$ be the dimension of the output space, the input and output signal spaces are defined as,

$$\mathcal{U} := \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_{m} = \mathcal{V}^m, \;\; \mathcal{Y} := \underbrace{\mathcal{V} \times \cdots \times \mathcal{V}}_{o} = \mathcal{V}^o$$

The corresponding extended signal spaces $\mathcal{U}_e$, $\mathcal{Y}_e$ are defined accordingly. Given a plant,

$$P : \mathcal{U}_e \to \mathcal{Y}_e \mid \; P(0) = 0 \tag{3.4}$$

and controller,

$$C : \mathcal{Y}_e \to \mathcal{U}_e \mid \; C(0) = 0 \tag{3.5}$$

The closed loop system is defined using the following equations

$$y_1 = Pu_1 \tag{3.6}$$
$$u_0 = u_1 + u_2 \tag{3.7}$$
$$y_0 = y_1 + y_2 \tag{3.8}$$
$$u_2 = Cy_2 \tag{3.9}$$

$w_0 = (u_0, y_0)^T \in \mathcal{W}$ represents the plant input and output disturbances, $w_1 = (u_1, y_1)^T \in \mathcal{W}_e$ represents the plants true input and output and $w_2 = (u_2, y_2)^T \in \mathcal{W}_e$ represents the plant's measured input and outputs.

For the purpose of the application of the EMMSAC algorithm, it is necessary that the closed loop system satisfy two conditions. These are that the closed loop system be *Well Posed* and *Gain Stable*. These terms are defined below.

Definition 3.1. *A closed loop system [P,C] given by equations 3.6-3.9 is said to be well-posed if for all $w_0 \in \mathcal{W}$ there exists a unique solution $(w_1, w_2) \in \mathcal{W}_e \times \mathcal{W}_e$.*

It has been demonstrated previously by Buchstaller (2010) that this property also hold true for linear switched systems.

Definition 3.2. *A closed loop system [P,C] given by equations 3.6-3.9 is said to be gain stable if there exists an $M > 0$ such that:*

$$\sup_{w_0 \in \mathcal{W}, w_0 \neq 0} \frac{\left|\left|\Pi_{P//C} w_0\right|\right|_2}{||w_0||_2} = ||\Pi_{P//C}||_2 < M < \infty \tag{3.10}$$

where $\Pi_{P//C}$ is the map from the disturbances $w_0 \in \mathcal{W}$ to the plant signals $w_1 \in \mathcal{W}_e$,

$$\Pi_{P//C} : \mathcal{W} \to \mathcal{W}_e \times \mathcal{W}_e : w_0 \mapsto (w_1, w_2) \tag{3.11}$$

3.1.2 Plant Set

Before the formal specification of the plant set is provided, it is useful to consider a simple example of a possible plant set. As an example assume that the process that needs to be controlled can be represented with a discrete time first order transfer function given below,

$$P_p = \frac{p}{z - p}$$

where p is an uncertain parameter in the range $0 < p \leq 1$, a possible set of candidate plants would then be,

$$\frac{0.1}{z-0.1}, \frac{0.2}{z-0.2}, \ldots, \frac{1}{z-1}$$

For the purpose of this thesis however, all plants in the plant set are always represented using the state space representation. Let $\mathcal{P}$ be the parameter set which represents the plants. Since state space plants are used, this set can be represented as,

$$\mathcal{P} := \left\{ (A,B,C,D) \in \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times m} \times \mathbb{R}^{o\times n} \times \mathbb{R}^{o\times m} \middle| \begin{array}{ll} (A,B) & \text{is controllable} \\ (A,C) & \text{is observable} \\ n \geq 1 & n,m,o \in \mathbb{N} \end{array} \right\}$$

Here n is the system order and m, o are number of system inputs and outputs as defined previously. A single plant P_p in the plant set can be represented as,

$$P_p : \mathcal{U}_e \to \mathcal{Y}_e : u_1^p \mapsto y_1^p, p = (A_p, B_p, C_p, D_p) \in \mathcal{P} \tag{3.12}$$

The mapping from the plant input to output is achieved using the following relations,

$$x_p(k+1) = A_p x_p(k) + B_p u_1^p \tag{3.13}$$
$$y_1^p = C_p x_p(k) + D_p u_1^p \tag{3.14}$$
$$x_p(-k) = 0,\ k \in \mathbb{N} \tag{3.15}$$

Here x_p is the state vector for plant P_p and k is the discrete time step.

3.1.3 Controller Set

The controller set consists of a controller corresponding to every plant in the plant model set and is given through the design method $K : \mathcal{P} \to \mathcal{C}$ where $\mathcal{C}$ is the parametrised set of all controllers. The minimal requirement for each plant controller pair is that it be gain stable. The controllers corresponding to each plant can be generated using any method feasible for the application for instance LQR, pole placement or an exhaustive search.

3.1.4 The *Closeness* Function

Before a controller can be switched into the control loop, it needs to be determined which plant in the plant model set is *closest* to the true plant P_{p*}. This is achieved

through a closeness function which returns a residual signal $r_p(k)$ for each plant P_p. This signal is a positive scalar that measures the closeness of plant P_{p*} to P_p using the true plant's measured input and output signals $w_2 = (u_2, y_2)^T$. By convention used in this research, the lower the residual signal the closer is a plant to the true plant. The normal functioning of the EMMSAC algorithm requires the evaluation of the residual signal $d_p(k)$ for the entire plant model set at each time step k. The controller corresponding to the plant with the lowest value for $r_p(k)$ is switched into the loop. Figure 3.2 illustrates the true plant, one candidate plant and the corresponding signals.

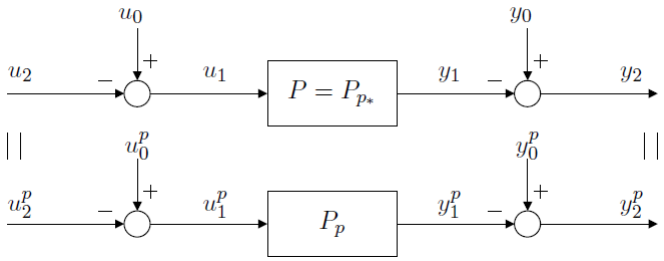

Figure 3.2: Disturbances and observations (Obtained from Buchstaller (2010)).

The residual signal is evaluated using a *disturbance estimator* which is described next.

The Disturbance Estimator: The purpose of the disturbance estimator is taken as input the parameters of the candidate plant P_p and the true plant's measured signals $w_2 = (u_2, y_2)^T$ and determine the smallest *size* disturbance signals $w_o^p = (u_0, y_0)^T$ which are *explain* the true plant's measured signals. The size of the disturbance signals is measured in terms of the 2-norm and is the required residual. By the term *explain* it is meant that disturbance signals be consistent with the measured signals and the candidate plant, i.e they should satisfy the following relation over the interval $[a, b]$.

$$(y_0^p - y_2) = P_p(u_0^p - u_2) \tag{3.16}$$

Formally, such disturbance signals will be referred to as *weakly consistent* disturbance signals and are defined below.

Definition 3.3. Let a $\leq$ b, a,b $\in \mathbb{N}$. The set of weakly consistent disturbance signals $N_p^{[a,b]}(w_2)$ to a plant $P_p, p \in P$ and the observation $w_2 = (u_2, y_2)^T$ is defined by,

$$\mathcal{N}_p^{[a,b]}(w_2) := \left\{ v \in \mathcal{W}|_{[a,b]} \middle| \begin{array}{c} \exists (u_o^p, y_o^p)^T \in \mathcal{W}_\urcorner \; s.t \\ \mathcal{R}_{b-a,b} P_p (u_o^p - u_2) = \mathcal{R}_{b-a,b}(y_o^p - y_2), \\ v = (\mathcal{R}_{b-a,b} u_o^p, \mathcal{R}_{b-a,b} y_o^p) \end{array} \right\}$$

where $\mathcal{W}_e = \mathcal{U}_e \times \mathcal{Y}_e$ and $\mathcal{R}$ is the *restriction operator*.

Estimator Implementation: Now given the plant $P_p, p \in \mathcal{P}$, the residual for the observed signal $w_2 \in \mathcal{W}_e$ can be given by the residual operator X as follows,

$$X(w_2)(k)(p) = r_p^A[k] = \inf\{r \geq 0 |\ r = ||v||_2, v \in \mathcal{N}_p^{[0,k]}(w_2)\} \tag{3.17}$$

Here $r_p[k]$ is the residual of the plant P_p at time instant k. The simplest method to implement this estimator is to formulate the residual determination as a constrained optimization problem given below.

$$w_o^p \in \inf\left\{r \geq |r = ||v||_2, v \in \mathcal{N}_p^{[0,k]}(w_2)\right\}, a \leq b \in \mathbb{N} \tag{3.18}$$

Clearly it is not feasible to implement X directly because the computational complexity involved in computing it increases with time.

However as has been demonstrated by Buchstaller (2010), it is not necessary to exhaustively search for the smallest 2-normed disturbance signals which are weakly consistent with the candidate plant and the true plant's measured signals. As the author has demonstrated, the 2-norm of the smallest disturbance signal is equal to the residual of a Kalman filter corresponding to the candidate plant and the true plant's measured signals. This is achieved using the following relations,

$$\hat{x}(k+\frac{1}{2}) = \hat{x}(k) + \Sigma(k)C_p^T[C_p\Sigma(k)C_p^T + I]^{-1}[y_2(k) - C_p\hat{x}(k)] \tag{3.19}$$

$$\Sigma(k+\frac{1}{2}) = \Sigma(k) - \Sigma(k)C_p^T[C_p\Sigma(k)C_p^T + I]^{-1}C_p\Sigma(k) \tag{3.20}$$

$$\hat{x}(k+1) = A_p\hat{x}(k+\frac{1}{2}) + B_p u_2(k) \tag{3.21}$$

$$\Sigma(k+1) = A_p\Sigma(k+\frac{1}{2})A_p^T + B_pB_p^T \tag{3.22}$$

$$\widetilde{y}(k) = C_p\hat{x}(k) \tag{3.23}$$

Here k indicates the time step, A_p, B_p, C_p are the state space matrices corresponding to the system P_p, $\hat{x}(k+\frac{1}{2})$ and $\hat{x}(k+1)$ are the a priori and posteriori estimated state vectors, $\Sigma(k+\frac{1}{2})$ and $\Sigma(k+1)$ are the a priori and posteriori error covariance matrices, $u_2(k)$ and $y_2(k)$ are the measured input and output signals for the true plant P_p^* and $\widetilde{y}(k)$ is the estimated system output.

The Kalman residual of the plant P_p at time step T is denoted with $r_p(T)$ and defined as,

$$r_p(T) = \left[\sum_{k=0}^{T} ||y_2(k) - \widetilde{y}_1(k)||^2_{[C\Sigma(k)C^T+I]^{-1}}\right]^{\frac{1}{2}} \tag{3.24}$$

The residual is a measure of the *closeness* of plant P_p to P_p^*. The lower the residual, the closer the two plants are. The residual is zero if the plants are identical.

Buchstaller (2010) states the result that the Kalman filter residual equals the size of the smallest disturbance in the form of the following theorem.

Theorem 3.4. *Let $p = (A_p, B_p, C_p) \in \mathcal{P}$ and suppose C_p is full row rank. The Kalman filter equations 3.19-3.23 with initial conditions $\hat{x} = 0$ and $\sigma(0) = \sigma(0)^T = 0$ describe a deterministic least-squares filter initialised to zero:*

$$\sum_{k=0}^{T} ||y_2(k) - \widetilde{y}_1(k)||^2_{[C\Sigma(k)C^T+I]^{-1}} = r_p^2(T) = \inf_{(u_0^p, y_0^p) \in \mathcal{N}_p^{[0,T]}(w_2)} (||u_0^p||_2^2 + ||y_0^p||_2^2) \quad (3.25)$$

3.1.5 The Switching Algorithm

The principle of the switching algorithm is simply to select a plant from the plant model set that has the least residual. The switching algorithm thus generates a signal q_f that points to the plant with minimal residual and can therefore be written as,

$$q_f(k) := \arg\min_{p \in \mathcal{P}} \; r_p[k], \; \forall k \in \mathbb{N} \quad (3.26)$$

Now, at each time instant k the switching signal q_f puts a controller corresponding to the least residual plant into the loop. The complete algorithm is illustrated in Figure 3.3.

The switching can be done at each time instant or after a chosen delay Δ. The purpose of the delay is to reduce the number of times switching is done, since excessive switching might reduce closed loop system performance or make it unstable. Interestingly rapid switching between two stable linear systems can cause instability. Hespanha et al. (2003) provides an example using two stable LTI systems with intersecting state trajectories. The author then illustrates that sufficiently fast periodic switching between the two systems can cause instability.

The current implementation of the EMMSAC uses a fixed delay and will be explained in the next section.

3.2 Application of EMMSAC for AFM Control

This section illustrates the applicability of the EMMSAC algorithm for AFM control. This is done through simulation results where the AFM control loop is simulated for two samples, one metallic and the second one a polymer. The motivation behind this choice is to demonstrate that separate controllers are needed for the samples and that the EMMSAC algorithm can detect which sample is being scanned and switch the corresponding controller into the loop.

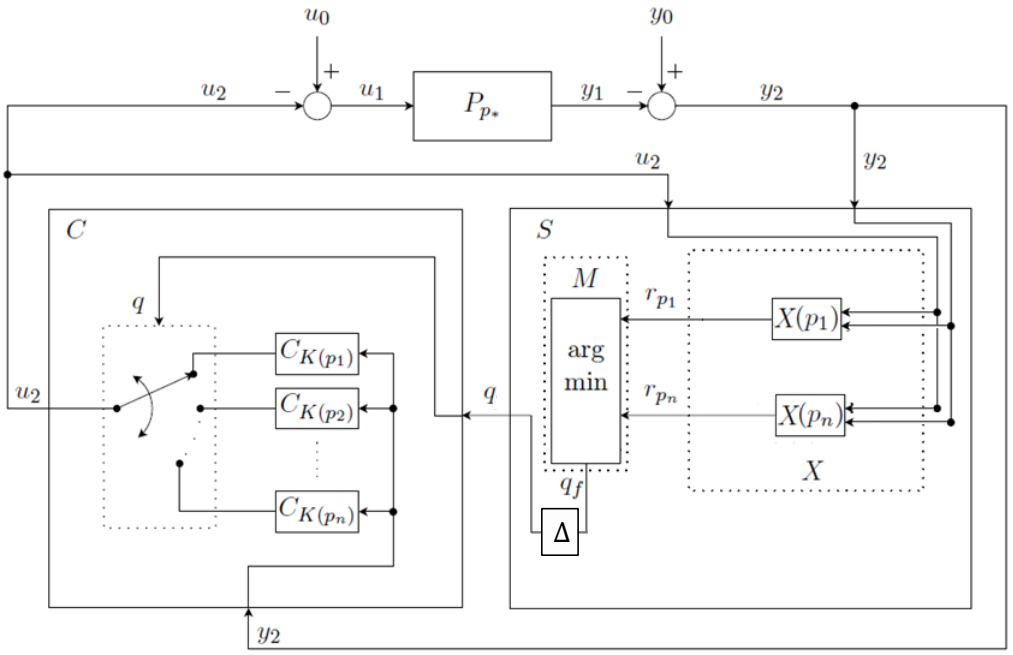

Figure 3.3: The complete EMMSAC algorithm (Obtained and modified from Buchstaller (2010)).

Before the simulation details can be described it must be made clear that the original EMMSAC algorithm is modified in two ways to enable usage for AFM control.

Firstly, it is observed that the cantilever plant dynamics vary with a variation in controller gains. This is because of the non-linear variation of the tip-sample interaction force with respect to the tip-sample separation. If weak gains are switched in, the tip will mostly vibrate far from the surface or in the attractive region of the tip-sample interaction force. If strong gains are used, the tip will mostly vibrate in the repulsive region, i.e it will penetrate the surface. As a consequence, a fixed pre-chosen grid of trial controllers is simulated for each sample and the cantilever plant input and output signals are saved. These are used to generate a least residual plant corresponding to each controller test. Thus a cluster of least residual plants is generated for each sample. The plant set then consists of all the plants in both clusters.

Next, one Kalman estimator is generated corresponding to each plant in the set. During AFM operation the residuals for all Kalman estimators are estimated, and the mean residual for each cluster is determined. The cluster with the lesser mean residual is considered to be closest to the true AFM plant.

Secondly rather than having one stabilizing controller for each plant, a suitable controller is chosen for each cluster. The suitability of a controller is determined using a performance metric described in this section. The selected controller is chosen from the

grid of trial controllers. Since there are two clusters in this case, the final controller set consists of two controllers one for each sample. During AFM operation, once the cluster with the lesser mean residual is determined, the controller corresponding to that cluster is switched into the AFM control loop. A detailed description of this entire procedure is provided next in this section.

The simulations are performed using the same simplified AFM control loop as illustrated in Figure 2.13. Two samples are used which have drastically different properties, i.e. the first one is Fused Silica SiO_2 and the second one is a generic soft biological sample. The material properties as obtained from Stark et al. (2004) and Hoelscher and Schwarz (2007) respectively are provided in Table 3.1. All remaining simulation parameters are exactly the same as those illustrated in Table 2.2.

Parameter	**Fused Silica**	**Biological Sample**
Hamaker constant (J)	6.4×10^{-20}	0.2×10^{-18}
Elastic modulus (GPa)	70	1
Poison ratio	0.17	0.3
Intermolecular Separation (nm)	0.16	0.3

Table 3.1: Sample material properties.

For the purpose of the current simulation the measured system input $u_2 = f_D$, and the measured system output $y_2 = d_{TM}$. Ideally the cantilever base excitation force F_b should be used as input. However since in the case of the experimental results explained in the next chapter, it is not possible to measure F_b. This is because the Z piezo actuator has no position measurement sensors and its transfer function is not known. This means that either of h or f_D can be used. In this case f_D is used. This is because high pass filtering of the cantilever output implies that control signal h has relatively lesser impact on the output d_{TM} since h has a much lower frequency as compared to f_D.

The generation of the plant and controller sets as described in the next two sections is done by simulating single line scans for a set of forty trial controllers C_T give below,

$$(K_p, K_i) \in C_T = [0.01, 0.02, \ldots, 0.1] \times [0.001, 0.003, 0.005, 0.007] \tag{3.27}$$

The range and separation for the controller gains in this set is chosen empirically. Each simulated scan line generates five signals of interest namely, A_{sp}, A, h, f_D and d_{TM}. These are placed in a data set D_j^i given below,

$$D_j^i = [A_{sp},\ A,\ h,\ f_D,\ d_{TM}] \in \mathbb{R}^{250001 \times 5} | 1 \leq i \leq 2, 1 \leq j \leq 40 \tag{3.28}$$

Here the superscript i corresponds to the sample material type ($i = 1$ for Fused Silica and $i = 2$ for Polymer) and the subscript j corresponds to the controller number. The dither piezo input f_D equals u_2, and cantilever deflection d_{TM} equals y_2. Each column vector in a data subset has a length of 250001. This originates from the fact that, (as mentioned in Table 2.2) each simulation assumes the scan time T to be 0.05s with a sample time T_s of $0.2\mu s$ (Sample rate of 5MSPS). The term MSPS is an abbreviation for mega samples per second. The complete data set $\mathbb{D}$ is given below,

$$\mathbb{D} = \begin{bmatrix} D_1^1 & D_2^1 & \dots & D_{40}^1 \\ D_1^2 & D_2^2 & \dots & D_{40}^2 \end{bmatrix} \tag{3.29}$$

The next two subsections illustrate how this data is used for the generation of the plant and controller sets followed the simulation results for the complete EMMSAC algorithm in the last subsection.

3.2.1 Plant Set

Once the data set is recorded, the data subsets $D_j^i \;\; \forall \; i \;\; \in [1, 2], \;\; j \in [1, 2, \dots, 40]$ are used to generate corresponding least residual plants G_j^i, using the mapping procedure Φ_p,

$$\Phi_p : D_j^i \rightarrow G_j^i \;\; \forall \; i \;\; \in [1, 2], \;\; j \in [1, 2, \dots, 40] \tag{3.30}$$

For the purpose of clarity, each plant in the plant set which was referred to as P_p shall now be referred to as G_j^i. This mapping is done by using the piezo inputs $u_2 = f_D$ and cantilever deflection $y_2 = d_{TM}$ from the selected data subsets D_j^i and then finding the least residual second order LTI plant G_j^i given below,

$$G_j^i = P_p = \frac{k_o \omega_n^2}{s^2 + \frac{\omega_n}{Q} s + \omega_n^2} \tag{3.31}$$

Here k_o is the system gain, w_n is the resonance frequency in rads^{-1} and Q is the quality factor. The choice of this structure is motivated from the fact that the transfer function from the piezo input voltage to the cantilever deflection can be well approximated by a second order LTI system.

For the purpose of implementing the EMMSAC the discrete time state space matrices corresponding to the transfer function are determined using zero order hold. The sample time T_s used for this purpose is selected to be $0.2\mu s$. This makes the sample rate 5 MSPS (mega samples per second). The sample rate is selected to be well above the cantilever

resonance which is 65.497 kHz as mentioned in Table 2.1. Lower sample rates might be feasible as well, however this sample rate is used for the current implementation.

Given a data set D_j^i the corresponding observations $w_2 = [u_2\ y_2]$ are extracted. The closest plant G_j^i corresponding to these observations is determined by using the closeness function described previously. This is done by solving the following optimization problem,

$$G_j^i = \arg\min_p r_p(250001) \tag{3.32}$$

Here the subscript p again refers to the parameters needed to specify a candidate plant. Since only three parameters need to be specified p is defined as,

$$p = (Q\ \omega_n\ k_o) \tag{3.33}$$

The optimization is subject to the constraints $0 < Q \leq 500$, $2\pi(6500) < \omega_n \leq 2\pi(650,000)$, and $0 < k_o \leq 100$. The constraints for the quality factor stem from the fact it is always positive and most commercially available cantilevers have a quality factor less than 500. Likewise the upper and lower constrains for the resonance frequency are selected to be a tenth and ten times the nominal value of 65 kHz. This is done since no a-priori information about the effect of tip sample interaction on the resonance frequency is available. Finally, the limits for gain k_0 are empirically selected to be between 0 and 100. The optimization is performed using the Simplex search algorithm. The final plant set $\mathbb{G}$ can therefore be represented in the following manner,

$$\mathbb{G} = \begin{bmatrix} G_1^1 & G_2^1 & \dots & G_{40}^1 \\ G_1^2 & G_2^2 & \dots & G_{40}^2 \end{bmatrix} \tag{3.34}$$

The plants in this set are illustrated in Figure 3.4.

As illustrated in the figure, the two samples have distinct least residual plant clusters. In addition the parameters of the least residual plants are effected by the choice of the PI controller gains used to generate the scan line data.

3.2.2 Controller Set

The controller set is determined by finding the *best* controller corresponding to the row data subsets corresponding to each sample, i.e data sets $D_1^1,\ \dots\ D_{40}^1$ and $D_1^2,\ \dots\ D_{40}^2$. Ideally the best controller is the one that generates the least squared error between the amplitude set-point A_{sp} and A. The performance metric is thus simply given as,

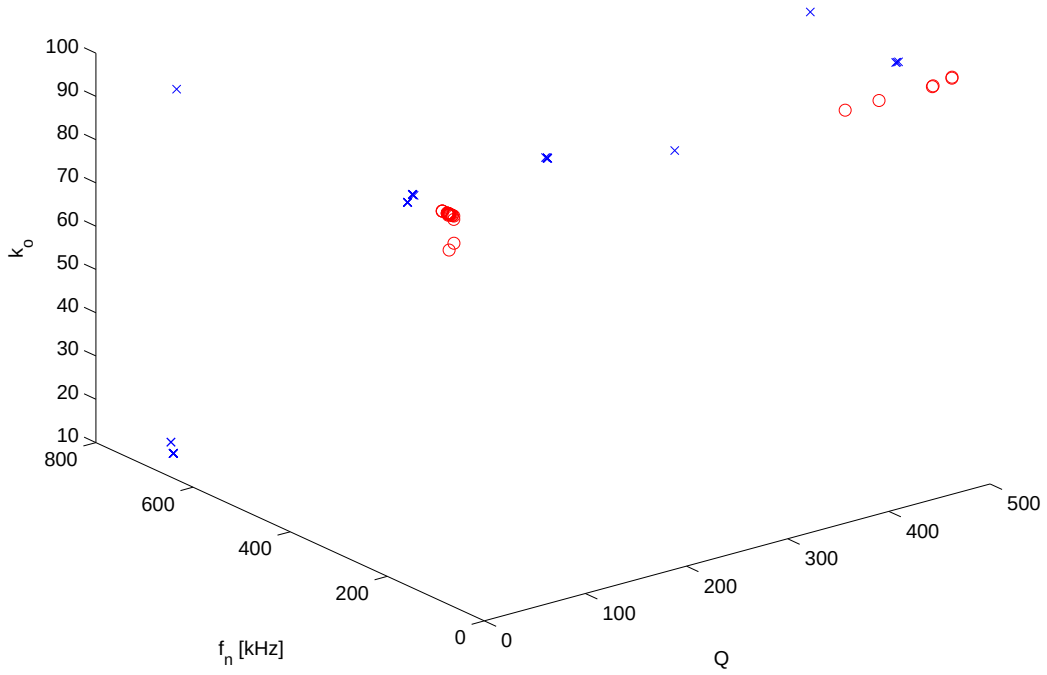

Figure 3.4: Least residual plants for samples one ('x') and two ('o').

$$\chi_j^i = \sum_{k=37501}^{250001} \left(e_j^i(k)\right)^2 \tag{3.35}$$

where,

$$e_j^i = A_{sp} - A \Big| \begin{array}{l} \forall\ i \in [1,2] \\ \forall\ j \in [1,2,3,\ldots,40] \end{array} \tag{3.36}$$

The upper limit in the summation is the number of elements in the error vector. The lower limit corresponds to 15% of the vector length. Ideally the summation of the squared error signal should begin from $k = 1$, however this will include the initial transient behaviour of the error signal. Since the objective of the controller is to estimate the sample topography, therefore the initial 15% of the error signal is excluded from the computation. This can be done only in the case of a simulated experiment, where it is known exactly when the sample features of interest begin.

The mapping procedure which maps each data subset to the corresponding best controller is given as,

$$\Phi_c : D_j^i \rightarrow C_k \Big| \begin{array}{l} \forall\ \ i \in [1,2] \\ \forall\ \ j \in [1,2,3,\ldots,40] \\ \forall\ \ k \in [1,2] \end{array} \tag{3.37}$$

The mapping Φ_c is done by first extracting the amplitude vector A and the set point amplitude A_{sp} from the relevant data set D_j^i and finding the error e_j^i. Next the best controller C_k corresponding to the row of data subsets for each as sample is determined as follows,

$$C_k = C_T(j^*) \;\; \forall \;\; k \in [1, 2] \tag{3.38}$$

where,

$$j^* = \arg\min_j \chi_j^i \Big| \begin{array}{l} \forall \;\; i \in [1, 2] \\ \forall \;\; j \in [1, 2, 3, \ldots, 40] \end{array} \tag{3.39}$$

and C_T is the set of 100 trial controllers as mentioned in Equation 3.27.

The final controller set then consists of two controllers one corresponding to each sample, and referred to as C_E and is illustrated in Table 3.2.

Table 3.2: Controller set.

Sample. No.	K_p	K_i
1	0.04	0.007
2	0.06	0.007

Figure 3.5 compares the performance of the two controllers for each of the two samples. The performance improvement in terms of the performance metric χ is illustrated in the same figures.

Here C_1 and C_2 are the optimal controllers for samples 1(Silica) and 2(Soft Biological Sample).

These results are obtained using the same control loop as illustrated in Figure 2.13 and using the parameters mentioned in Table 2.2. As illustrated, a separate controller is needed for each of the two samples. While an optimal controller for one sample can reconstruct the topography and can do the same for another sample, this generates clear distortions in the reconstruction.

Even though the effect of using the correct controller is visually and numerically clear in the case of the simulated results, it is observed during the experimental verification that the squared error metric alone is not sufficient. In particular two limitations were observed. Firstly, the metric does not sufficiently take into account the oscillatory nature of the estimation. While there is a substantial reduction in the squared error χ for the case of the soft sample when the correct controller is switched in, switching in the correct controller for the first sample reduces the squared error only marginally.

Secondly, it does not take into account the difference between the trace and retrace estimations. Ideally the trace and retrace topography estimations should be as similar

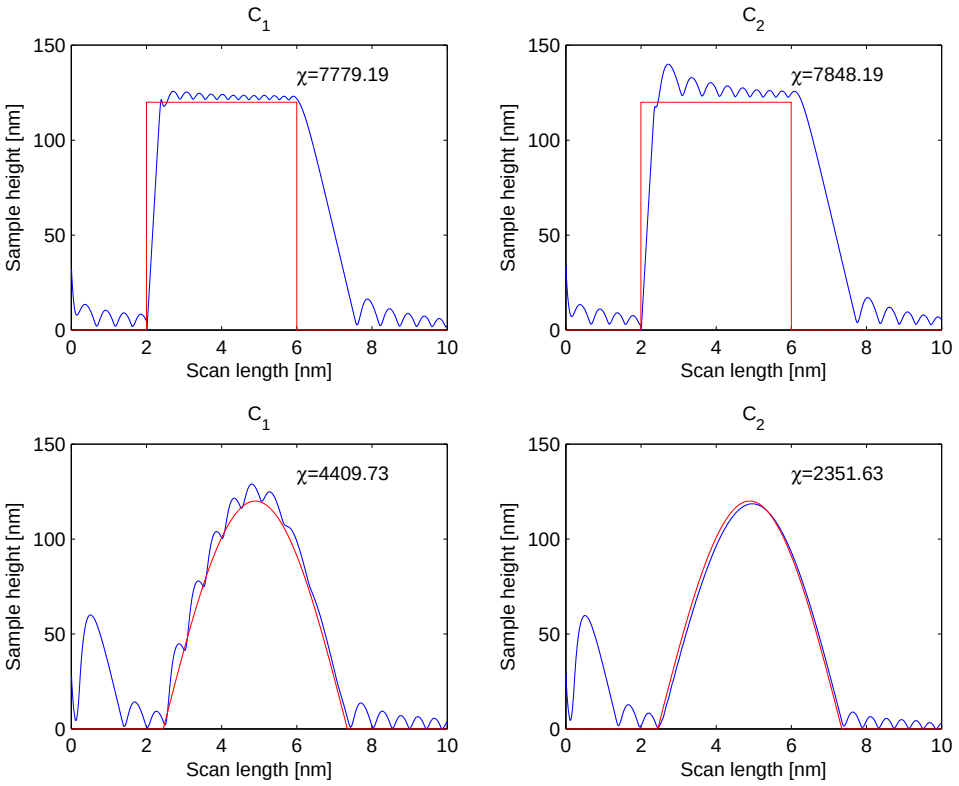

Figure 3.5: Controller performance comparison for Sample 1 (top row) and Sample 2 (bottom row)

as possible. Therefore, two additional constraints are introduced and are explained in detail in the next chapter which illustrates the experimental results.

3.2.3 Simulation Results

The complete MMAC AFM control simulation is carried out by using the AFM and Topography parameters as provided in Tables 2.2 and 3.1. The control loop is augmented with an additional block corresponding to the EMMSAC algorithm as shown in Figure 3.6.

After each scan line is simulated the AFM input-output data $w_2 = [u_2 \; y_2]$ is used to determine the switching signal q. Here the plant input is the cantilever excitation force $u_2 = f_D$ whereas the plant output is the cantilever deflection signal $y_2 = d_{TM}$. The switching signal is then generated using the mapping procedure ϕ_q given below,

$$\Phi_q : w_2 \rightarrow q \;\; \forall\, i \in [1, 2] \;\; \forall\, j \in [1, 2, \ldots, 40] \tag{3.40}$$

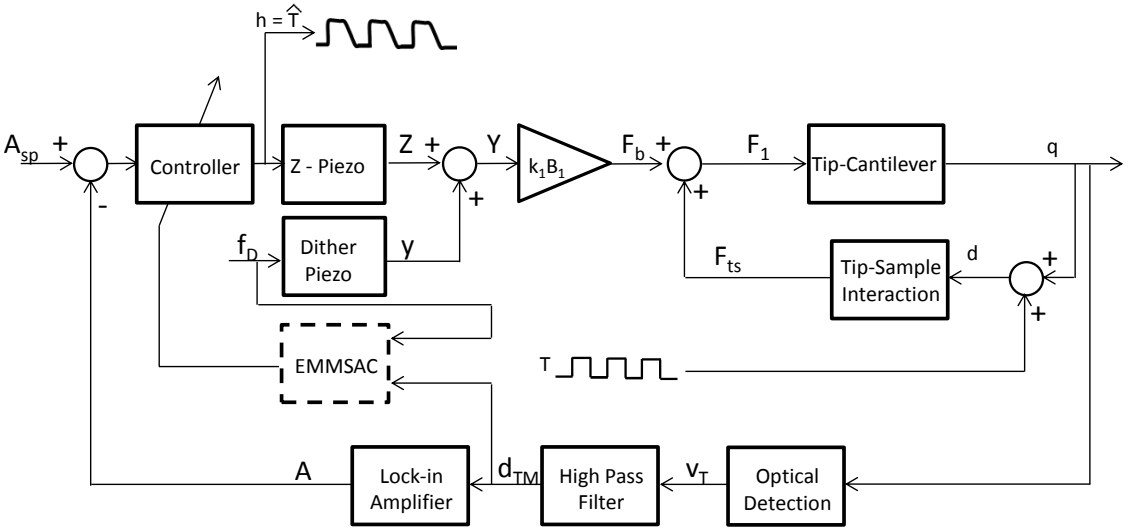

Figure 3.6: The AFM control loop with automated controller tuning using EMMSAC.

The mapping is done by computing the residuals corresponding to the plant input-output w_2 for each plant in the plant set $\mathbb{G}$ and recording them in a residual matrix $\mathbb{R}$ given below,

$$\mathbb{R} = \begin{bmatrix} r_1^1 & r_2^1 & \dots & r_{40}^1 \\ r_1^2 & r_2^2 & \dots & r_{40}^2 \end{bmatrix} \quad \mathbb{M} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} \tag{3.41}$$

The mean residual for each row m_i is evaluated. Next the residuals which are in excess of one standard deviation away from the mean are excluded. The row mean is now evaluated again and written in a matrix $\mathbb{M}$. This exclusion of outlying residuals is motivated from the observation that a few residuals are unusually high and do not represent the actual mean of the residuals in a row. The subscript i in m_i now corresponds to the i^{th} row of the residual matrix $\mathbb{R}$. The switching signal q then equals the row index with the least mean residual.

$$q = \arg\min_i m_i \;\; \forall\, i \in [1, 2] \tag{3.42}$$

Once the switching signal is determined, the corresponding controller is switched into the loop. The steps involved in the procedure are summarised below,

Step 1: The first scan line is simulated, and the signals u_2 and y_2 are saved. The first scan line is always simulated using the false controller. This means that controller C_2 is used for sample 1 and vice versa.

Step 2: The data obtained is used to determine the switching signal q using the mapping ϕ_q as mentioned previously.

Step 3: The remaining scan lines are simulated using the recommended controller.

Before the simulation results can be presented, the simulations initial conditions and the EMMSAC delay need to be described.

Initial Conditions: It must be noted that all the scan lines are simulated with the same initial conditions. The initial cantilever base height Z is always set to be slightly greater than the amplitude of free vibration A_o. This choice is motivated by the fact that in a physical AFM experiment the initial cantilever base height is generally greater than the amplitude of free vibration. Specifically in this case the initial cantilever base height is set to be 10% greater than A_o. The cantilever tip position and velocity are also initially zero. The reason for keeping the initial conditions same for each simulated scan line is that the controller performance varies substantially even for small variations in initial conditions.

In the case of the hardware implementation described in the next chapter, it is not possible to exactly set the initial conditions at the start of each line. It is observed that the controller performance varies from scan line to scan line even when the same controller is used and the same sample region is scanned. As a consequence the performance results for the hardware experiments are averaged for a larger number of trials. These results shall be discussed in detail in Section 4.2.

EMMSAC Delay Δ: For the purpose of the current research, it is observed that switching controllers is suitable only after an entire scan line has been completed. This is so because if the controllers are switched in earlier, it will be difficult to estimate controller performance. This is true because the data of a complete scan line is needed to determine how well the sample topography has been estimated. Therefore a delay was not explicitly implemented.

The remaining parts of this section illustrate the performance improvement obtained using the EMMSAC algorithm.

Figure 3.7 illustrates the performance before and after the first switch for sample one (Fused Silica). The loop is started with the false controller. As can be seen that there is an overshoot in the estimated sample topography. After the switch the reconstruction is clearly better since there is no over shoot and the topography is distorted lesser by oscillations. This is quantified by a reduction in the summed squared error χ.

The next Figure 3.8 illustrates the estimated sample surface. It is clear that only the first line is distorted because the simulation is initiated with a suboptimal controller. Figure 3.9 illustrates the sample estimation with the false controller.

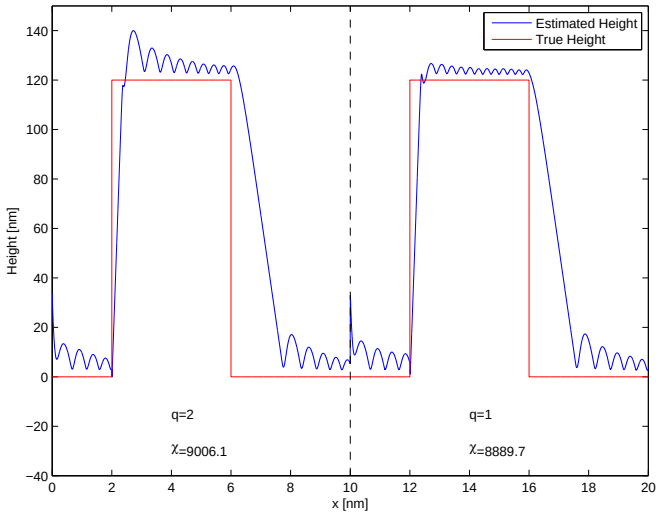

Figure 3.7: EMMSAC scan lines for sample one (Fused Silica).

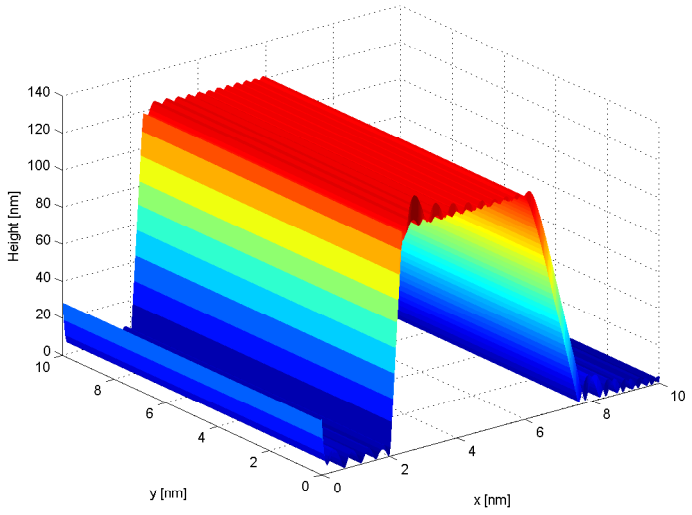

Figure 3.8: Estimated sample topography for sample one using EMMSAC.

Figures 3.10 and 3.11 illustrate the simulation results for sample two (Biological specimen). Once again the results demonstrate that the EMMSAC determines the correct controller and improves performance. The last figure 3.12 illustrates the sample reconstruction with the false controller.

A consistent distortion is observed at the start of the scan line. However once the control signal settles down, the topography is estimated correctly.

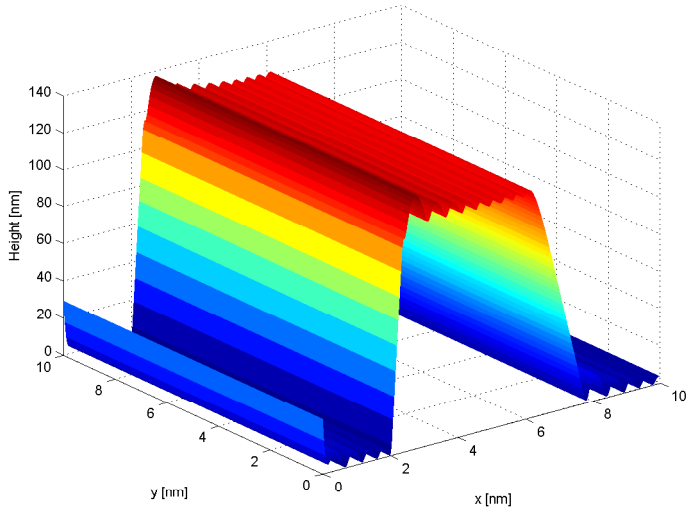

Figure 3.9: Estimated sample topography for sample one using controller C_2.

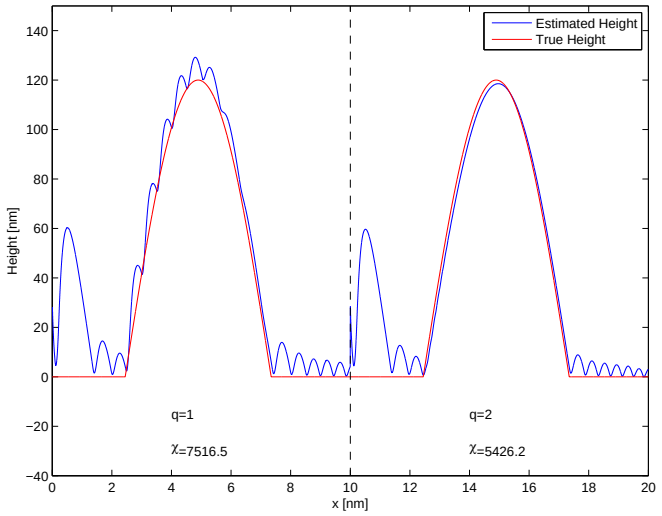

Figure 3.10: EMMSAC scan lines for sample two (Biological specimen).

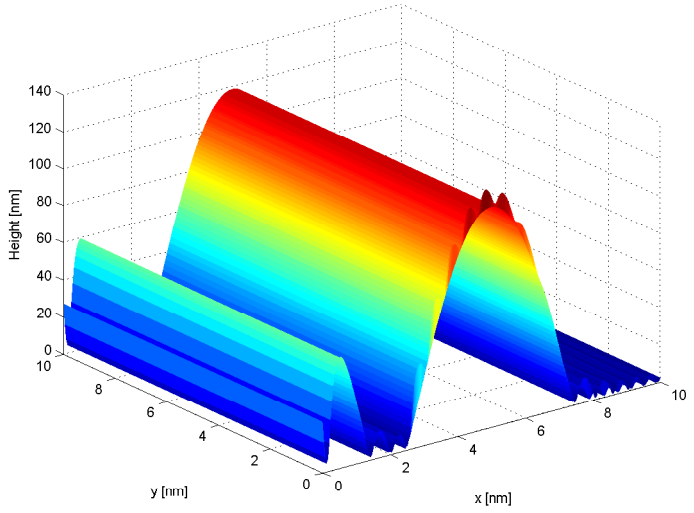

Figure 3.11: Estimated sample topography for sample two using EMMSAC

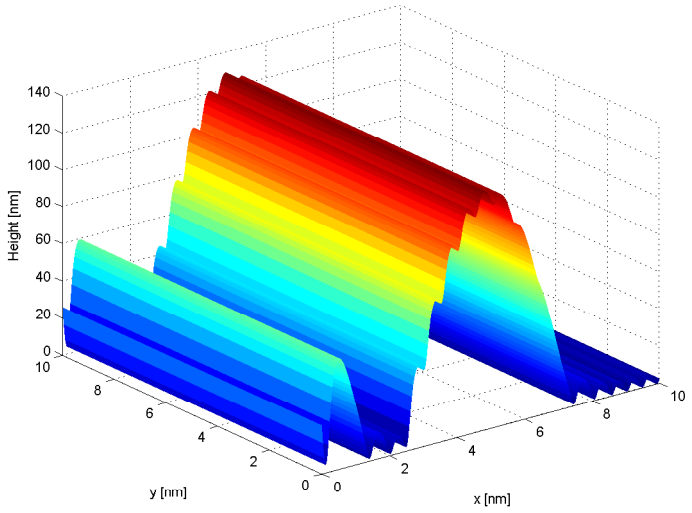

Figure 3.12: Estimated sample topography for sample two using controller C_1

The next chapter explains the implementation of the EMMSAC algorithm on a commercial AFM, and illustrates the experimental results obtained.

Chapter 4

Application of EMMSAC on AFM Hardware

This chapter provides the experimental results which verify that the EMMSAC algorithm can be used to generate images of a sample without manual tuning. The existing commercial AFM used for this purpose has an analog PI control system, with no possibility of interfacing to a desktop PC. Therefore a customised digital system is implemented. This is done using a Digital Signal Processor (DSP) from Texas Instruments. The first section (4.1) provides the implementation details for this hardware.

The verification of the EMMSAC algorithm is first done in an *offline* manner, where a large set of experimental data is used for verification of the EMMSAC performance in the context of a benchmark. The meaning of the term offline and the benchmark are explained in section 4.2. Lastly, Sections 4.3 and 4.4 present images obtained using EMMSAC and manual tuning respectively.

In the next chapter towards the end of Section 5.1 it will be explained how the automated tuning is also useful for reduction in AFM image generation times.

4.1 Controller Implementation Using Digital Signal Processor

The sole motivation of developing a custom control system was to enable access to the control signals. It was determined that there are two possibilities to implement such a controller namely Digital Signal Processors (DSP) and Field Programmable Gate Arrays (FPGA). Of these two the first one was selected because of the greater ease in programming DSPs. DSPs can be programmed in C/C++ whereas FPGAs can be programmed using two low level programming languages Verilog® and VHDL®. Both require greater development time as compared to C/C++.

Amongst the vast range of DSPs available the DSP chosen for the current application was the TMS320C6713 DSP from Texas Instruments. This decision was made based on a number of factors namely, relatively low cost, availability in the form of a DSP kit which eliminates the requirement for making a substantial amount of interface circuitry and reasonably fast clock frequency of 225 MHz. In addition this DSP allows floating point operations which greatly simplifies programming over the previously used fixed point operations. Finally and most importantly, this DSP has a key enabling feature called Real Time Data Exchange (RTDX). This allows data to be exchanged between the host PC and the target DSP *on the fly*, i.e without interrupting real time code execution. This implies that if a PID controller is running on the processor, the controller gains can be changed without halting the processor. This is especially important for the implementation of the EMMSAC algorithm, since it relies on changing the controller parameters while the controller is functioning.

The remaining parts of this section explain in detail the data acquisition components used with the DSP kit, details of the digital controller data flow and methods used for verification of the controller.

4.1.1 Data Acquisition

This section explains in detail all the circuitry assembled for acquiring data from and sending to the Nanonics AFM. All Nanonics I/O signals are in the range of ± 10 V. Furthermore, it is required that the measurement circuit has an input impedance greater than $10k\Omega$. In addition the circuit that sends signals to the AFM must have an output impedance no greater than 1 $k\Omega$. The current measurement circuit has an input impedance of $10^{13}\Omega||6pF$ Texas Instruments (b).

Data is read into the DSP kit using a 4 channel, 16 bit ADC running at 250 kSPS (kilo samples per second) the ADS 8361 from Texas Instruments. On the output side the data is written out using the 16 bit DAC 8831 also from the same manufacturer. This DAC has a maximum update rate of 1 MSPS. Both ADC and DAC are provided by Texas Instruments in the form of evaluation boards with substantial amounts of interface circuitry fabricated on them. These boards fit on the DSP kit through an interface board referred to as the 5-6K Interface Board. The ADC is used without an anti-aliasing filter because the frequency spectrum of the amplitude signal $A(t)$ which is read by the ADC is well below the sampling frequency. This is experimentally verified in the section 4.1.3.

Although the evaluation boards have all the necessary components which provide serial data communication between the ADC/DAC chips and the DSP kit, they lack the necessary signal conditioning circuits that will enable connecting the ADC/DAC evaluation boards to the Nanonics AFM. The problem is that all Nanonics AFM analog I/O vary in a ± 10 V range, however the ADC is unipolar, i.e it accepts voltages in the 0-5 V

range. Similarly the DAC output voltage range is limited to ± 2.5V range. In addition the DAC output needs to be filtered to eliminate the noise generated by the DAC itself.

To compensate for this two circuits were implemented. The first one illustrated in Figure 4.1(left) (obtained from Texas Instruments (a)) converts a ± 10V signal to a 0-5V one. The operational amplifier selected for this case is the OPA132 because of its very high $10^{13}\Omega||6pF$ input impedance. The input circuit has two components. Operational amplifier labelled A is configured as a differential amplifier. Operational amplifier B a unity follower. The unity follower takes the 2.5 reference voltage from the ADC and passes it to the differential amplifier to offset its output by 2.5V. This reference voltage is also passed to the ADC for use as a reference. The ADC records all voltages greater than 2.5V as positive and the others as negative. The output of the differential amplifier then equals

$$V_o = \left(1 + \frac{R_1}{R_3}\right)\left[\frac{R_2 V_{in} + R_4 REF_{OUT}}{R_2 + R_4}\right] \tag{4.1}$$

This expression is obtained through a simple superposition procedure. This procedure uses the fact that the operational amplifier A's output is a superposition of inverting and non-inverting feedback configurations.

The second one Figure 4.1 (right) provides low pass filtering for the DAC output. Both figures are obtained from the component specification sheets. The low pass filter is a simple first order system with the DC gain of unity and a cut off frequency of 15 Hz. While this low pass filtering removes the noise from the DAC output it also severely constrains the maximum scan rate of the AFM. Since the controller output cannot be greater than 15 Hz, this means that the sample must be rastered slow enough. Slow enough in this case means that the frequency content of the topography signal with respect to time must be substantially below the filter cut off frequency. As an example if the sample consists of perfect rectangular pits, the scan rate should be slow enough so that no more than one pit passes beneath the cantilever in a second. This would imply that the topography signal will be a 1 Hz square wave with respect to time. Given the fact that a square wave has odd harmonics, the seventh harmonic, which equals 15Hz will be reconstructible. Since the seventh harmonic can be obtained from the controller signal, it will be possible to reconstruct the original topography.

The reason for such extreme low pass filtering is that the DAC output is corrupted with noise at all frequencies. While the magnitude of the noise in any particular frequency range is relatively low, it is significant for the current implementation because the variation in the controller signal during a Nanonics AFM scan is no greater than $\frac{1}{2}$ volts. Although the filtering severely constrains the scan speed, this controller does enable application of the EMMSAC. Furthermore, it must also be taken into account that this slow scan speed is still slightly faster than the recommended scan speed for this AFM.

Therefore, while it will be clearly better to investigate scanning at higher frequencies, the current implementation is sufficient for the current instrument.

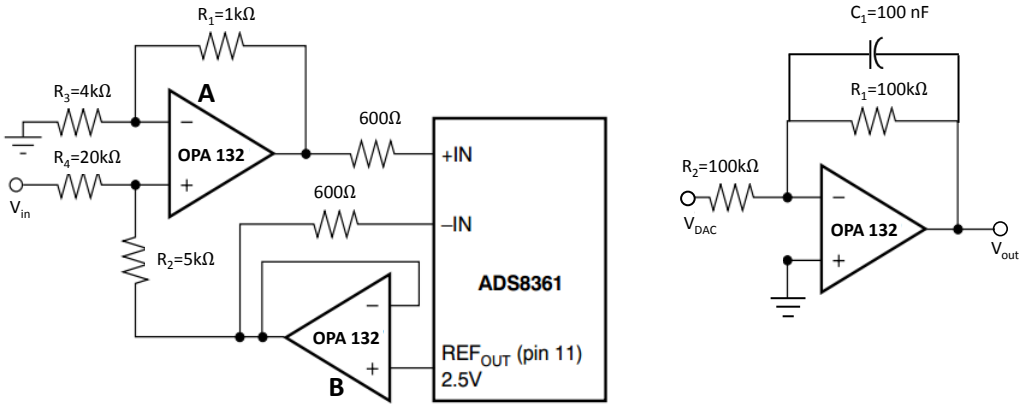

Figure 4.1: ADC (left) and DAC (right) interface circuits.

4.1.2 DSP Controller Data Flow

The entire DSP controller algorithm is *Interrupt Driven*, i.e the processor waits until the ADC has read a data sample and it is ready to be read. The ADC reads samples at the rate of 250 kSPS using timing signals generated by the ADC evaluation module. Once the ADC has read a sample it generates a signal referred to as an *Interrupt*. The processor upon receiving this interrupt runs the control law on the sample and writes the controller output to the DAC. All these steps are explained in detail below and illustrated in the data flow graph in Figure 4.2.

The first step is to initialize the processor. This includes the usual procedures including defining variables, enabling interrupts and initializing the serial interface.

Once the processor is initialized it stays in idle state until the ADC generates an interrupt. This initiates execution of the corresponding *Interrupt Service Routine*(ISR). Once control is transferred to the ISR the first step is to read the data from the ADC and store it in a temporary buffer of the type INT32. The next step is to identify the ADC channel at which the sample has been read. The ADC has four input channels (A0,A1,B0,B1) and reads either two or all four channels simultaneously. Currently it is operated in a mode that allows reading two data channels simultaneously namely A1 and B1, where the AFM amplitude signal is connected to the B1 channel. Channel A1 is not used in this application. This channel identification is done by making use of the

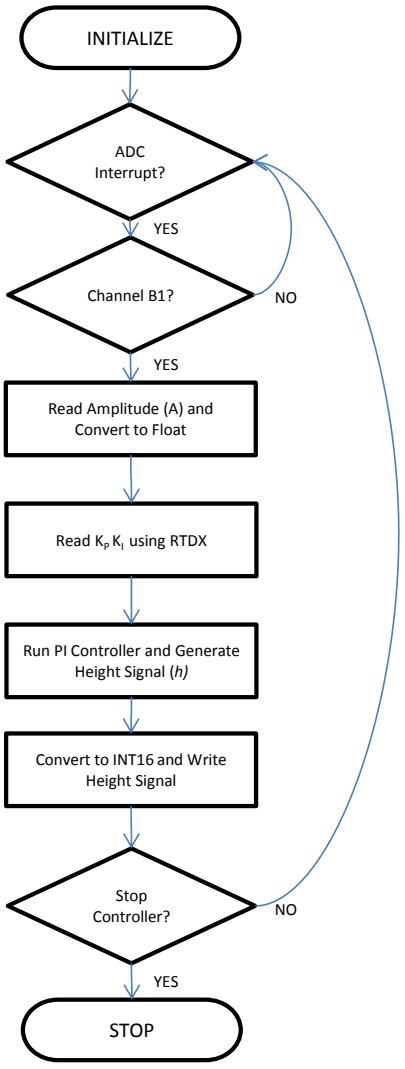

Figure 4.2: DSP controller data flow.

format in which the data is sent serially from the ADC to the DSP. This is illustrated in Figure 4.3.

CLOCK CYCLE	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
SERIAL DATA	CH 0/1	CH A/B	DB15	DB14	DB13	DB12	DB11	DB10	DB9	DB8	DB7	DB6	DB5	DB4	DB3	DB2	DB1	DB0	0	0

Figure 4.3: ADC serial data format (Texas Instruments (a)).

The first row indicates the clock cycles and the second lists the purpose of the data bit transferred at each clock cycle. Here DB is abbreviation for Data Byte. As can be seen all data transfers between ADC and DSP occur in frames of 20 bits. The first two bits

serve the purpose of channel identification. Bits 3-18 contain the discretized input signal value in twos complement form. The last two bits are always zero.

Since the last two bits are always zero they are discarded by executing a bit wise right shift on the data stored in the buffer. The status of bit 2 is then read by performing a bit wise AND operation with 0×00010000 and comparing with 0×00010000. In case of a match the data is from channel B1. The status of bit 1 is irrelevant since in the current ADC mode only channels A1 and B1 are read.

Once it is verified that the current sample is from the correct channel, data bits 3-18 are stored in a data buffer y of type INT16. Next the data is converted to a floating variable using a standard C/C++ casting operation. At this stage the processor reads the PID controller gain parameters K_p, K_i, using the *Real Time Data Exchange* (RTDX) interface. This is implemented simply by using the non-blocking routine provided by Texas Instruments. The advantage of using a non-blocking routine is that it checks for new data on the RTDX connection and returns control immediately to the ISR, thus allowing continued real time execution. On the transmission side the RTDX channel can be written to directly by Matlab using a very simple interface.

Once both controller parameters and the latest sample have been stored in y, the controller output is determined using the following control law,

$$\begin{aligned} e(i) &= y_{sp} - y(i) \\ u(i) &= K_p e(i) + K_i T_s \sum_{i=0}^{i} e(i) \end{aligned} \tag{4.2}$$

where y_{sp} is the set point, T_s is the sample time, e is the error between the set point and the current value, and i is the discrete time variable. It must be noted that no anti-wind up is implemented with the integrator. This is done because the output range of the DAC is $\pm 2.5V$, which is sufficient to control the piezo-actuator without saturation.

Finally, once the controller output is generated it is converted to the INT16 data type again and written to the DAC through the serial interface. The DAC is a single channel device so no channel identification bits need to be added. The only modification that needs to be made is that since the ADC sends the data in two's complement form, this must be compensated for by executing a XOR operation. Before the data is written to the DAC it is XORed with 0×8000 and then sent to the DAC.

Figure 4.4 illustrates the final functional block diagram of the entire DSP based controller setup. While the DSP kit will run the controller, the DAQ will be used to observe the AFM Height (h), Amplitude (A), dither piezo input (f_D) and tapping mode deflection (d_{TM}) signals. These will be sent to the PC where the EMMSAC algorithm will use

the f_D and d_{TM} signals to generate the optimal set of controller gains. The Height and Amplitude signals are used for constructing the sample image and evaluating controller performance respectively. The set point amplitude A_{sp} is saved into the DSP and PC at the start of the imaging process. Once the optimal controller gains are determined by the EMMSAC algorithm running on the PC, they are sent to the DSP using the RTDX interface. Here McBSP 0 and McBSP 1 refer to the Multi Channel Buffered Serial Port interface between the ADC/DAC and the C6713 DSP.

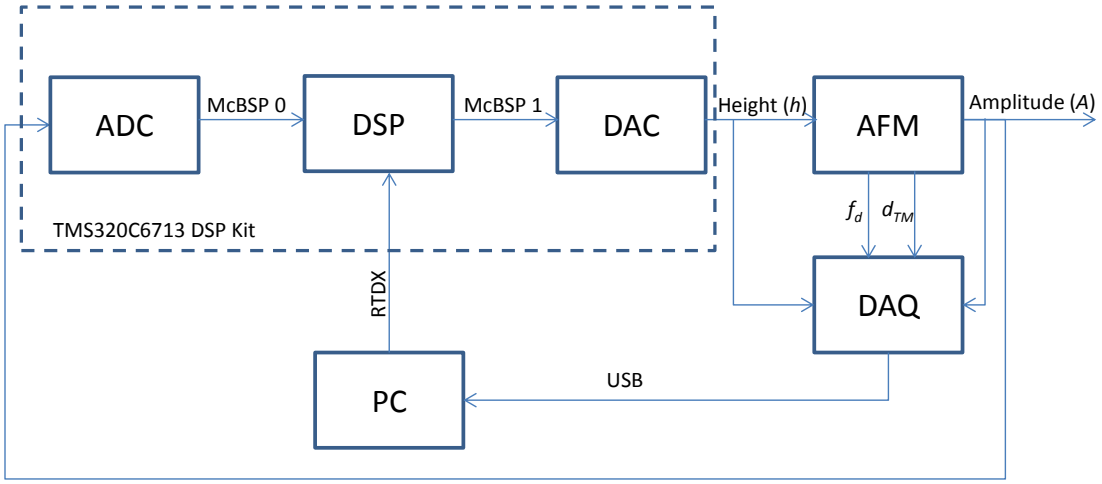

Figure 4.4: Functional block diagram.

4.1.3 DSP Controller Verification

The purpose of this section is to describe the procedure used to test the controller implemented on the DSP and determine how closely it matches the behaviour of an ideal PI controller. The first step in this this direction is to determine the sample time T_s of the DSP controller. The sample time in this case is defined as the time between the acquisition of two data samples. This interval includes the time needed to get the controller parameters using RTDX, generate the controller output and writing it to the DAC. All these steps are collectively referred to as *Data Processing* and assumed to take time T_p. This is illustrated in Figure 4.5.

The Code Composer Studio used to program the DSP provides a feature to measure the time interval between two events. This was used to determine the sample time, and in this case T_s is 7.5μs. A sample time of 7.5μs means a sample rate of 133kSPS. This sample rate is sufficient for the current application. The reason is that the Amplitude signal from the AFM, which the DSP controller is meant to measure has its highest frequency approximately less than 500 Hz. This can be verified using the frequency spectrum of a second long segment of the Amplitude signal as illustrated in Figure 4.6. As can be seen in the bottom plot the frequency content is much less than 500 Hz. This

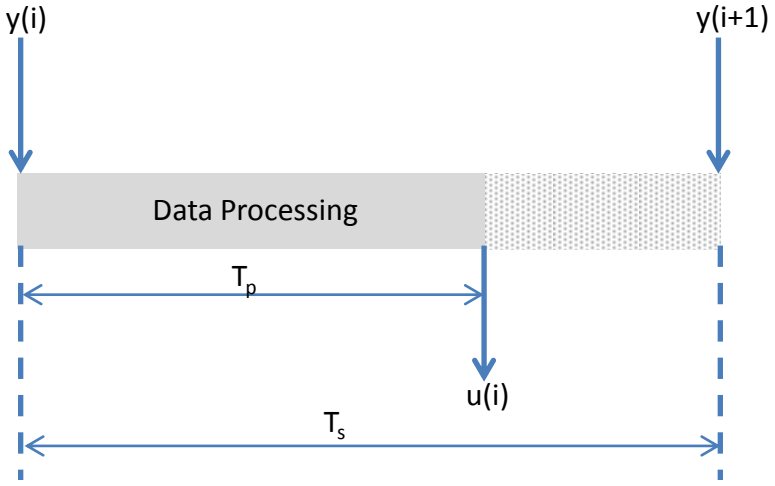

Figure 4.5: Sample time T_s.

signal was recorded using a National Instruments DAQ sampling at a rate of 1MSPS with a resolution of 16 bits.

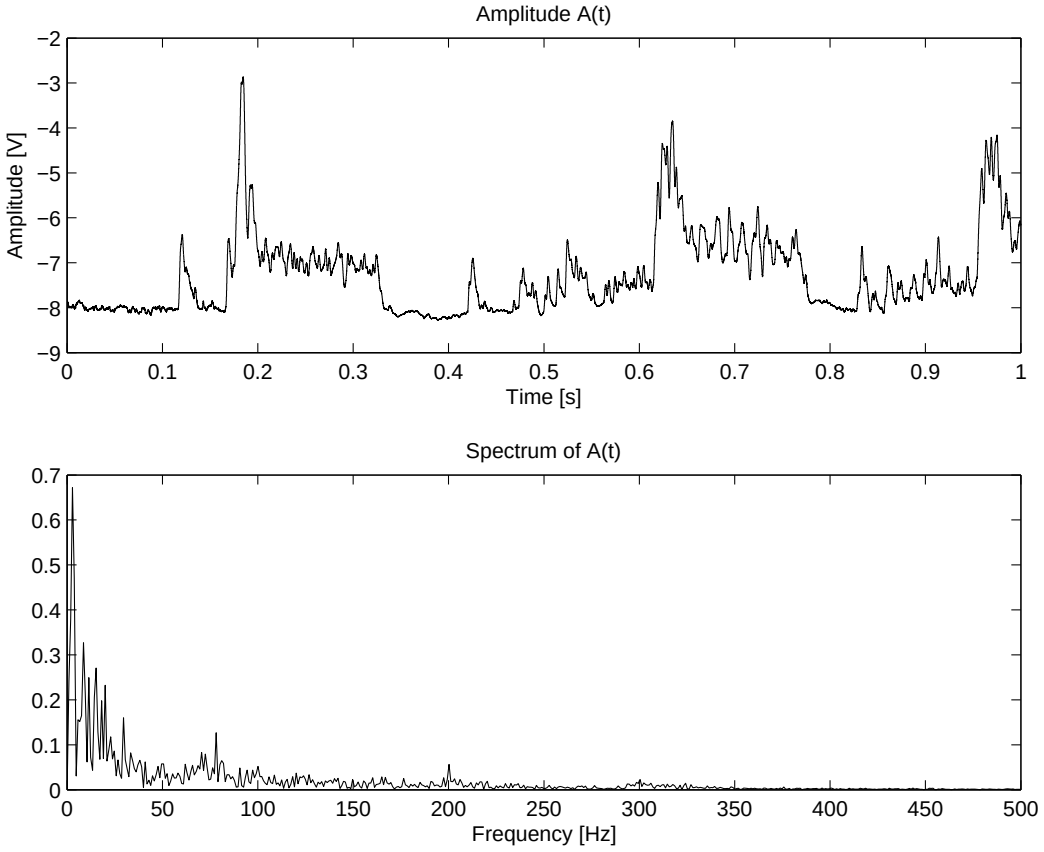

Figure 4.6: Frequency spectrum of A(t).

In the next step the two components of the PI controller were verified separately by generating Bode plots and comparing with plots generated in Matlab using the same

controller parameters. In all cases a sinusoidal input with unit amplitude was provided to the DSP controller. Since the DAC can generate no more than 2.5 V, the maximum gain during the verification procedure is $20log_{10}2.5 = 7.95dB$. The lowest possible gain is determined by the resolution of the DAQ card used to measure the response of the DSP controller. The DAQ can measure voltages in the range of ± 10 V with a resolution of 16 bits. Therefore the gain that can be measured is $20log_{10}\frac{20}{2^{16}} = -70.30dB$. The frequency range for all plots is from 10 to 1000 Hz. The Bode plots were generated with out using the low pass filter at the DAC output. The magnitude and phase of the noise corrupted DAC output was estimated using a regression procedure explained in this section. These plots are illustrated in Figures 4.7, 4.8 respectively. Each plot shows the ideal magnitude and phase curves generated by Matlab (solid line) and those from the DSP controller (dashed line). The maximum gain errors for the P and I controllers are 0.06 dB and 1.18 dB respectively. In the case of phase the DSP controller's phase deviates from the ideal values as frequency increases. This is due to the latency of the digital processor. However this error is not significant for the current application since the controller works at a frequency less than 10 Hz. The maximum phase errors for the P and I controllers is 9.42° and 10.1° respectively.

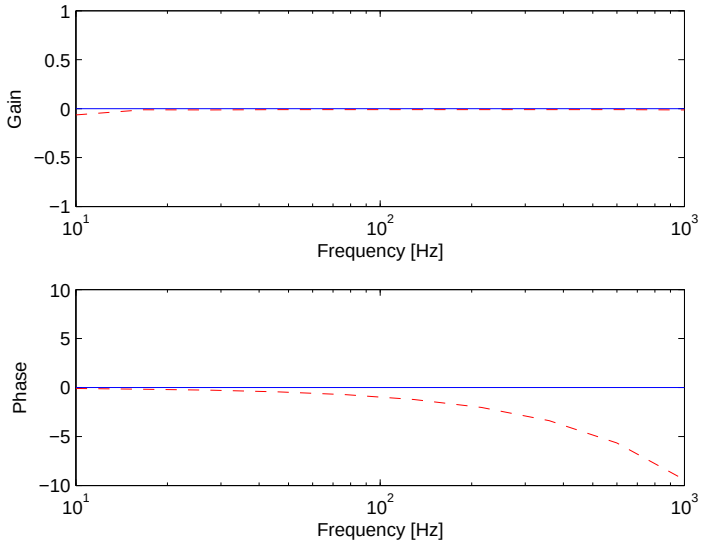

Figure 4.7: Proportional controller Bode plot.

The Bode plots are obtained using a simple procedure. The DSP is sequentially provided ten sinusoidal signals $x_i (i = 1, 2, \ldots, 10)$ with unity amplitude and logarithmically spaced frequencies f_i between 10 and 1000 Hz. The corresponding output signals y_i are recorded. Each signal segment lasts one second and all data is generated and recorded at the rate of 1M samples per second. The amplitude and phase of the input/output signals are determined using a regression method. The phase of the two signals is then used to determine the phase shift of the output with respect to the input. This method assumes that the input and output signals are sinusoids of the form,

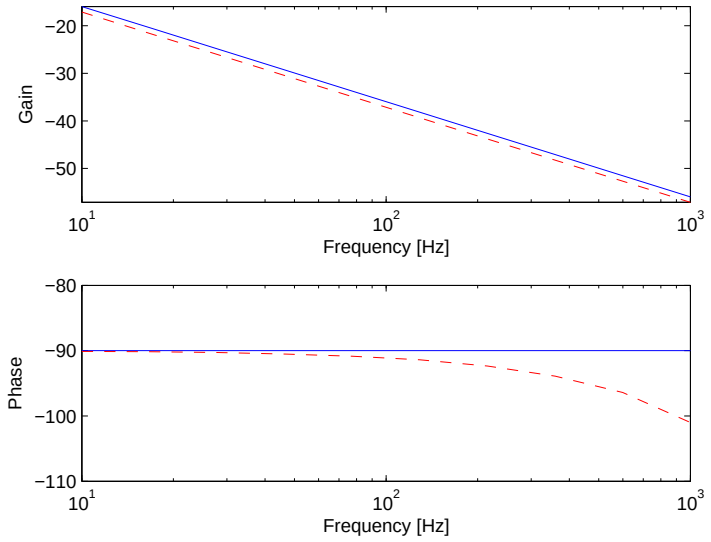

Figure 4.8: Integral controller Bode plot.

$$\begin{aligned} x_i(t) &= M_x \cos(2\pi f_i t + \phi_x) + k_x \\ y_i(t) &= M_y \cos(2\pi f_i t + \phi_y) + k_y \end{aligned} \tag{4.3}$$

where $t \in [0\ 1]$ is the time variable, M is the amplitude,ϕ is the signal phase and k is the signal offset. The method then makes use of the fact that the sinusoidal signals can also be written as,

$$\begin{aligned} x_i &= A_x \cos(2\pi f_i t) + B_x \sin(2\pi f_i t) + k_x \\ y_i &= A_y \cos(2\pi f_i t) + B_y \sin(2\pi f_i t) + k_y \end{aligned} \tag{4.4}$$

This immediately leads to the determination of amplitude and phase of both signals using the relations

$$\begin{aligned} M_x &= \sqrt{A_x^2 + B_x^2} \qquad \phi_x = \tan^{-1} \frac{-B_x}{A_x} \\ M_y &= \sqrt{A_y^2 + B_y^2} \qquad \phi_y = \tan^{-1} \frac{-B_y}{A_y} \end{aligned} \tag{4.5}$$

The system gain G_i for the i^{th} signal is then simply M_y and the corresponding phase shift $\Delta\phi_i$ is $\phi_y - \phi_x$. The regression procedure is applied by writing equations 4.4 as two matrix equations and solving for k_x, A_x, B_x and k_y, A_y, B_y respectively.

$$\begin{bmatrix} x_i \end{bmatrix} = \begin{bmatrix} \vec{1} & \vec{c_i} & \vec{s_i} \end{bmatrix} \times \begin{bmatrix} k_x \\ A_x \\ B_x \end{bmatrix} \tag{4.6}$$

$$\begin{bmatrix} y_i \end{bmatrix} = \begin{bmatrix} \vec{1} & \vec{c_i} & \vec{s_i} \end{bmatrix} \times \begin{bmatrix} k_y \\ A_y \\ B_y \end{bmatrix}$$

Here, $\vec{1}$ is a vector with all entries equal to unity and has a length equal to the signal segments x_i and y_i, $c_i = \cos 2\pi f_i t$ and $s_i = \sin 2\pi f_i t$

The final step in the verification procedure is to input a pre-recorded AFM amplitude signal to the DSP controller and compare its output with the output generated using the Nanonics controller. As can be seen in Figure 4.9 the DSP and the Nanonics controller signals are similar. The difference however is not relevant since the correctness of the PI controller has already been verified using Bode plots. This plot is generated to illustrate the fact that the DSP controller can be used to generate a sample image. The sample being viewed in this case is a rectangular metallic grid with a 3μ m pitch. As it can been the figure both controllers can extract the sample topography features. The only addition to get this result was a 15 Hz low pass filter after the DAC output to remove noise generated by the DAC. The PI gains in the DSP controller were set to $K_p = 0.4$ $K_i = 7$. An additional offset of 0.8[V] had to be added to the DSP output to bring it close to the output of the Nanonics controller. This offset occurs most likely due to the Nanonics controllers internal implementation and can be easily compensated for in the DSP code.

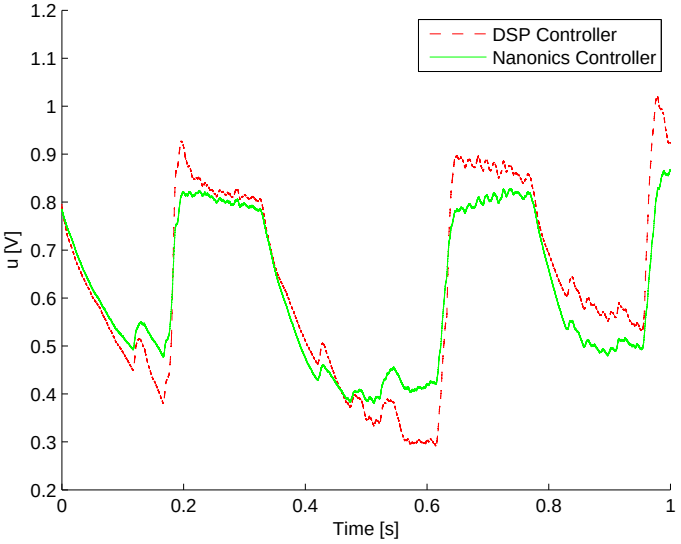

Figure 4.9: DSP and Nanonics controller outputs.

Figure 4.10 illustrate the final customized hardware developed. The DSP kit itself is located right beneath the visible circuitry. The figure illustrates the ADC, DAC modules along with the op-amp based interface circuit.

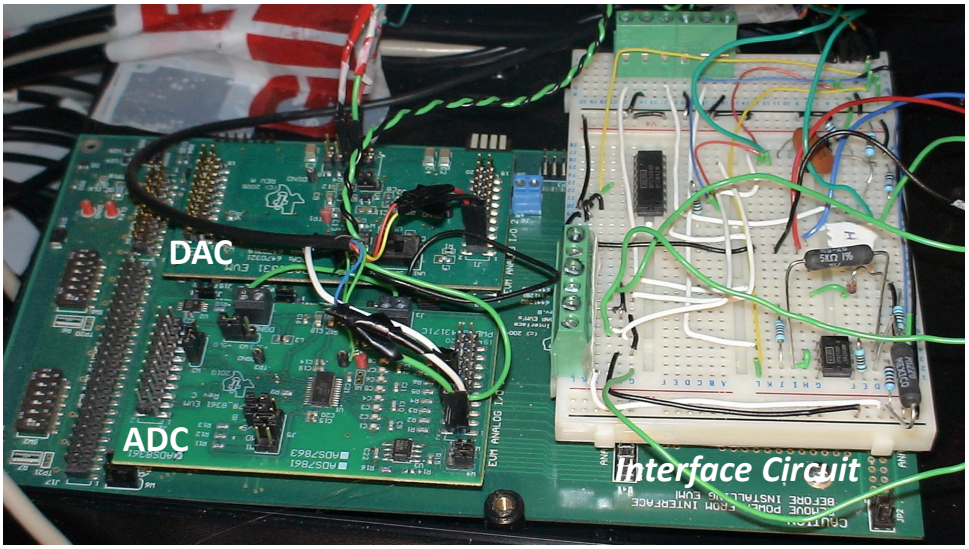

Figure 4.10: The customized controller hardware.

4.2 Offline EMMSAC Verification

The purpose of the offline EMMSAC verification is two-fold. Firstly, to demonstrate that the EMMSAC switching signal points to the *right* group of plants. Secondly, it shows that the EMMSAC based controller provides a better performance as compared to a fixed controller in the context of a benchmark. The term *right plant group* refers to the fact that, the least residual plants generated using more recent AFM data generally generate lesser Kalman Filter residuals. This is because the AFM plant parameters vary temporally even when the piezo-actuators, cantilever and sample are not changed. Therefore, the right group of least residual plants are those which are generated using AFM data which is temporally closer. This means that if one group of least residual plants is generated using a year old AFM data, and the other is generated using an hour old AFM data using the same equipment, the later group will be referred to as the right group of plants. This is so because the later group of least residual plants is likely to be a more accurate description of the AFM dynamics.

The offline EMMSAC verification was performed by collecting AFM data on two separate days using a fixed set of trial controllers C_T. Next, a fraction of the data from each day was used to generate least residual plants. The plants were then tested against the

remaining data from each day to determine if the least residual plants generated using data from the same day actually have a lesser residual.

The performance benchmark for this verification is the squared 2-norm of the error between the set-point amplitude A_{sp} and the actual amplitude A, i.e $||e||_2^2 = ||A_{sp} - A||_2^2$.

The remaining part of this section provides a detailed explanation of the offline EMMSAC verification procedure. Subsection 4.2.1 describes the experimental set-up used along with a description of the data recorded. This is followed by an explanation of the process used to generate the least residual plant set and the corresponding controller set in Subsections 4.2.2 and 4.2.3 respectively. Subsection 4.2.4 illustrates that the EMMSAC algorithm's switching signal points to the right group of plants.

4.2.1 Experimental Setup

All experiments have been performed on the Nanonics Imaging CryoView 2000 AFM. The cantilever obtained from Nanosensors is made of *n-silicon* with a length of $240\mu m$. The resonance frequency and quality factor as measured through a separate frequency sweep experiment are 65.5 kHz and 180.4 respectively. The imaged sample is the TGZ 02 calibration grid from Micro Masch. The calibration grid contains rectangular SiO_2 steps on Si wafer with a step height of $120 \pm 0.5nm$ and a pitch of 3.0 μm. The complete AFM control loop is illustrated in Figure 4.11.

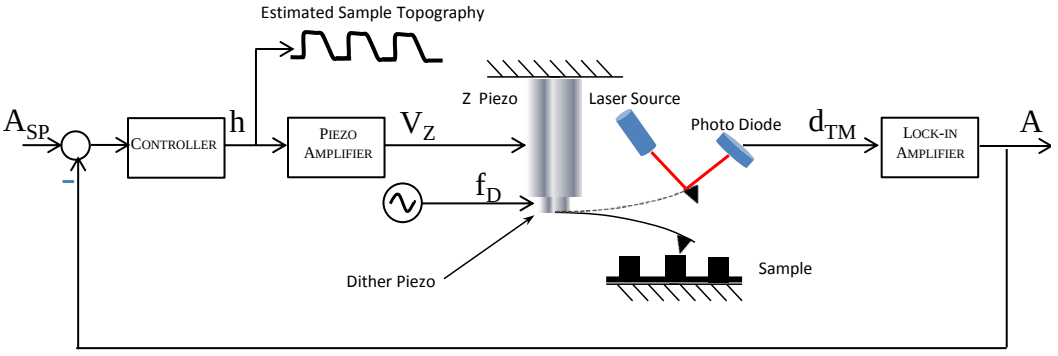

Figure 4.11: The AFM experimental setup.

In this figure, h is the height signal generated by the controller. This is amplified by a piezo-amplifier to generate V_z which is passed to the piezo-actuator. The dither piezo-actuator is actuated by an excitation signal f_D, which is a sinusoid with a frequency equal to the cantilever resonance frequency and a *sufficient* amplitude. The meaning of the term sufficient is explained next.

The excitation from the dither-piezo causes the cantilever tip to vibrate. This vibration is detected by the optical detection system and is referred to as the *tapping mode deflection*

signal d_{TM}. Finally, a lock-in amplifier is used to determine the amplitude A of d_{TM}. This is compared against the set point amplitude A_{sp} to generate the error signal, which is provided as input to the controller. Ideally the set point should be specified in nanometers. However, this is not currently possible with the existing hardware. The output from the lock-in amplifier has a range of 0 to 10 volts. Here 0 implies a static cantilever tip and 10 indicates the maximum measurable amplitude. According to the procedure specified by Nanonics first the tip is driven far from the surface so that there is no tip sample interaction. Next the amplitude of the excitation signal f_D is selected so that the amplitude A equals 9 volts. Following the manufacturer's recommendations the set point amplitude A_{sp} is then always selected to be be 1.6 volts below A. This means that the set point amplitude is 7.4 volts or approximately 80% of the amplitude of free vibration A. This selection originates from the fact that if the set point is selected to be too low for instance 50%, then the tip will vibrate too close to the sample surface. This will cause the tip to wear down or break. Likewise, if the set point is too high for instance 95%, the error will be difficult to measure and controller implementation will not be feasible.

In this case $u_2 = f_D$, $y_2 = d_{TM}$. The output disturbance y_0 principally originates from the sample topography. Ideally the effect of the z-piezo input voltage should be taken into account to calculate the force acting on the cantilever base F_b. Due to the open loop nature of the current piezo-actuators, their dynamics cannot be determined. As a consequence only the dither piezo input f_D and the photodiode output d_{TM} is used. The relation between f_D and d_{TM} is linear and can be approximated by a second order transfer function. Due to the relatively simple plant structure, the controller structure is empirically selected to be a proportional integral controller.

Before the EMMSAC algorithm can be applied for image generation a sufficient amount of data must be recorded for generation of the plant and controller sets. This is done in the following manner. A single line of the sample is repeatedly scanned with a one hundred separate PI controllers belonging to the following trial controller set C_T,

$$(K_p, K_i) \in C_T = [0.1, 0.2, \ldots, 1.0] \times [10, 20, \ldots, 100] \tag{4.7}$$

The line length is selected equal to the pitch of sample i.e, $3\mu m$. This ensures that sufficient topography variation occurs during each scan. Each line scan requires one second to complete and all five signals generated in the loop namely, A_{sp},A,h,f_D and d_{TM} are recorded. This is done with a National Instruments DAQ (NI USB 6356) with a sample rate of 10^6 samples per second. The PI controllers are implemented on the TMS320C6713 DSK from Texas Instruments using 16 bit A/D and D/A converters. The testing of each controller thus generates a data subset D_j^i given below,

$$D_j^i = [A_{sp},\ A,\ h,\ f_D,\ d_{TM}] \in \mathbb{R}^{10^6 \times 5} | 1 \leq j \leq 100, 1 \leq i \leq 10 \tag{4.8}$$

Here the subscript j indicates that the data subset is generated though **trial** of the j^{th} controller from the set C_T in the loop. The superscript i refers to the **iteration**. Each iteration consists of trials of all the 100 controllers in the set C_T.

In all two **experiments** were performed where each experiment consists of five iterations. The final data set thus generated is $\mathbb{D}$ is illustrated in Figure 4.12.

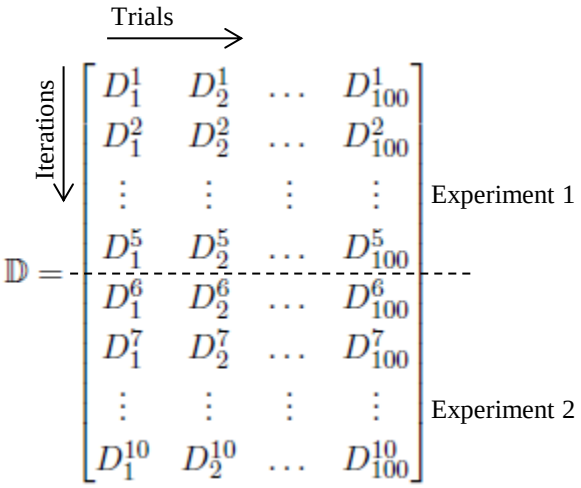

Figure 4.12: Experimental data set.

4.2.2 Plant Set

Once the data set is recorded the data subsets $D_j^i \ \forall\ i\ \in [1,6],\ \ j \in [1,2,\ldots,100]$ are used to generate corresponding least residual plants G_j^i. Only the data sets collected in the first iterations in each experiment are used for generation of the least residual plants. The remaining data sets in each experiment are used for verification of the EMMSAC procedure as explained in Section 4.2.4.

The least residual plant G_j^i are obtained from the data sets $D_i^j \ \forall\ i\ \in [1,6],\ \ j\ \in [1,2,\ldots,100]$ using the mapping procedure Φ_p,

$$\Phi_p : D_j^i \rightarrow G_j^i \ \ \forall\ i\ \in [1,6],\ \ j \in [1,2,\ldots,100] \tag{4.9}$$

This is the same mapping procedure described in section 3.2.1. The optimization is performed subject to the constraints $0 < Q \leq 500$, $2\pi(1000) < \omega_n \leq 2\pi(10^5)$, and $0 < k_o \leq 100$. The final plant set obtained is provided below.

$$\mathbb{G} = \begin{bmatrix} G_1^1 & G_2^1 & \dots & G_{100}^1 \\ G_1^6 & G_2^6 & \dots & G_{100}^6 \end{bmatrix} \tag{4.10}$$

The plants in this set are illustrated in Figure 4.16.

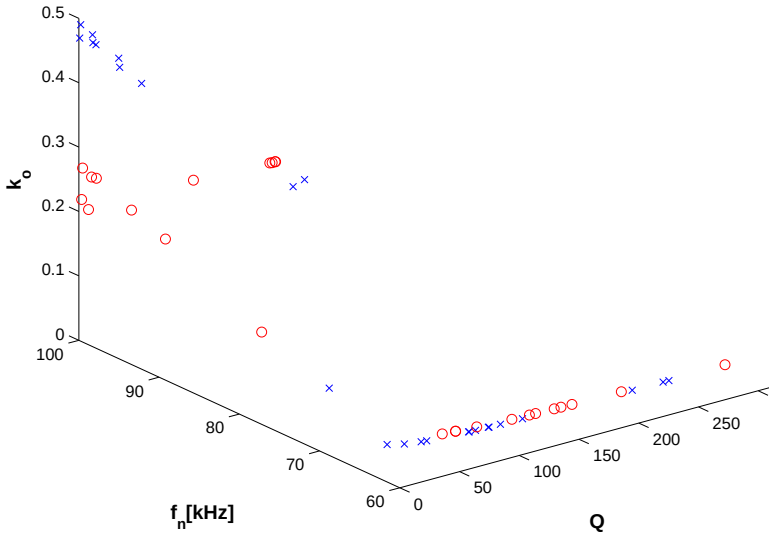

Figure 4.13: Plant set obtained for experiment one (Blue X) and two (Red O).

4.2.3 Controller Set

The controller set is determined by finding the *best* controller corresponding to the data subsets in the first iteration in each experiment, i.e data sets D_1^1, ... D_{100}^1 and D_1^6, ... D_{100}^6. Ideally the best controller is the one that generates the least squared error between the amplitude set-point A_{sp} and A. The performance metric χ_j^i is thus defined by as follows,

$$\chi_j^i = \sum_{k=1}^{10^6} \left(e_j^i(k)\right)^2 \tag{4.11}$$

where,

$$e_j^i = A_{sp} - A \Big| \begin{matrix} \forall \;\; i \in [1,6] \\ \forall \;\; j \in [1,2,3,\dots,100] \end{matrix} \tag{4.12}$$

The error signals in each case were sampled at a rate of 10^6 samples per second (1 MSPS). As illustrated in Figure 4.6, the frequency content of the amplitude signal $A(t)$ is well below 500 Hz. Clearly the frequency content of the error signal is exactly the

same. The sample rate is therefore greater than necessary. Given the large number of error vectors corresponding to each data set, the error vectors were first down sampled by a factor of 20. This is done to reduce the time needed to determine the controller set. In addition since the error signal is noisy it is low pass filtered using a $4^{th}-$ order low pass Butterworth filter with a cut-off frequency of 100 Hz. The filtering is performed on the down sampled signals in Matlab. Since the scan rate is low, i.e 0.5 Hz or 1 scan line per second it is unlikely that any meaningful information can be extracted from the signal at higher frequencies.

The motivation behind the low pass filtering is that it is difficult to evaluate EMMSAC performance using a noisy error signal. The final controller set is obtained using the same mapping procedure Φ_c described in section 3.2.2. It consists of two controllers one corresponding to each experiment and is illustrated in Table 4.1.

To conclude two operations are performed on the error signal before performance evaluation, i) Down sampling ii) Low pass filtering. Since the frequency content of the error signal is well below 500 Hz, the down sampled and filtered metric is identical to the one originally suggested in Section 3.2.2. These operations are performed only for the case of the offline EMMSAC verification.

Table 4.1: Controller set.

Exp. No.	K_p	K_i
1	0.2	90
2	0.9	70

4.2.4 EMMSAC Verification Results

The purpose of this section is to verify that switching signal points to the correct group of plants in the plant set using the cantilever input (drive signal f_D) and output (tapping mode deflection d_{TM}). This is done by extracting the signals f_D and d_{TM} from all the data subsets and evaluating the switching signal q using the mapping Φ_q, given below,

$$\Phi_q : D_j^i \rightarrow q \;\; \forall\, i \in [1, 2, \ldots, 10] \;\; \forall\, j \in [1, 2, \ldots, 100] \tag{4.13}$$

The mapping is done by extracting the signals f_D and d_{TM} from all the data subsets and evaluating the residuals for each plant in the plant set and recording them in a residual matrix $\mathbb{R}$ given below,

$$\mathbb{R} = \begin{bmatrix} r_1^1 & r_2^1 & \ldots & r_{100}^1 \\ r_1^6 & r_2^6 & \ldots & r_{100}^6 \end{bmatrix} \quad \mathbb{M} = \begin{bmatrix} m_1 \\ m_2 \end{bmatrix} \tag{4.14}$$

The mean residual for each row m_i is evaluated next and written in a matrix $\mathbb{M}$. The subscript i in m_i now corresponds to the i^{th} row of the residual matrix $\mathbb{R}$. The switching signal q then equals the row index with the least mean residual.

$$q = \arg\min_i m_i \quad \forall\, i \in [1, 2] \tag{4.15}$$

Ideally the switching signal generated from a subset in an experiment should point to the same experiment. For instance the switching signal q generated from data sets $D_j^i \quad \forall i \quad \in [1, 2, 3, 4, 5] \quad \forall\ j \in [1, 2, 3, \ldots, 100]$ should equal 1. Likewise for the data sets $D_j^i \quad \forall i \quad \in [6, 7, 8, 9, 10] \quad \forall\ j \in [1, 2, 3, \ldots, 100]$ the switching signal should equal 2. Figure 4.14 illustrates the results for the switching signal. As can be seen that the switching signal is able correctly identify which experiments the data sets belong too. The implication of this is that the switching signal points to those plants which correspond to the data most recently collected.

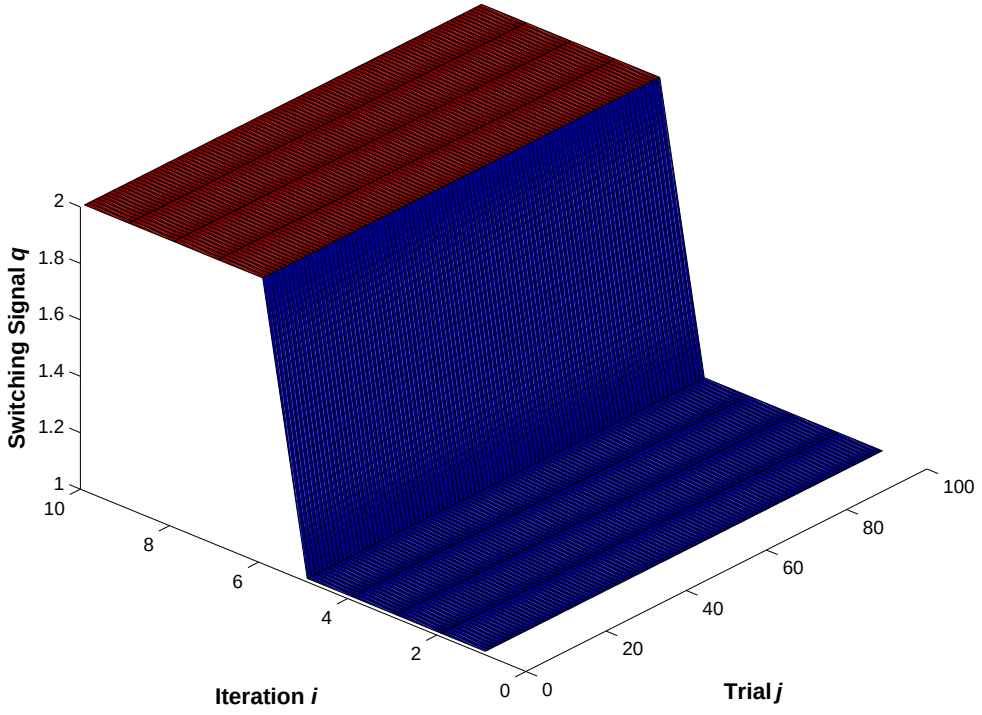

Figure 4.14: Least residual based identification results.

The fact that the EMMSAC is able to identify correctly every time is due to the relatively simple nature of the offline experiment. Only one sample is used, and the data for experiment 1 and 2 is collected on days which were in excess of two months apart. Such identification results may not occur if the data is collected at a smaller interval or if more than one samples are used. However the fundamental objective of this experiment is to

show that Kalman residuals have the ability to detect changes in the dynamics of an atomic force microscopy experiment. In addition it also reveals that this identification can be done using a small number of low order least residual plants.

Next, the performance of each controller in the controller set C_E is compared against the EMMSAC recommended controller in Figure 4.15. As can be seen that the EMMSAC controller performs *slightly* better than the rest of the controllers. The term slight is used because while there is a relatively bigger difference between the performance of C_1 and C_{EMMSAC}, there is a much lesser difference between C_2 and C_{EMMSAC}. This is quantified using the mean squared errors m_1, m_2 and m_{EMMSAC} as illustrated in the Figure 4.15. The mean squared error for the EMMSAC controller m_E is still the lowest. The mean is computed by excluding the performance for iterations 1 and 6. This is because the data sets collected in these iterations were used to determine controllers C_1 and C_2.

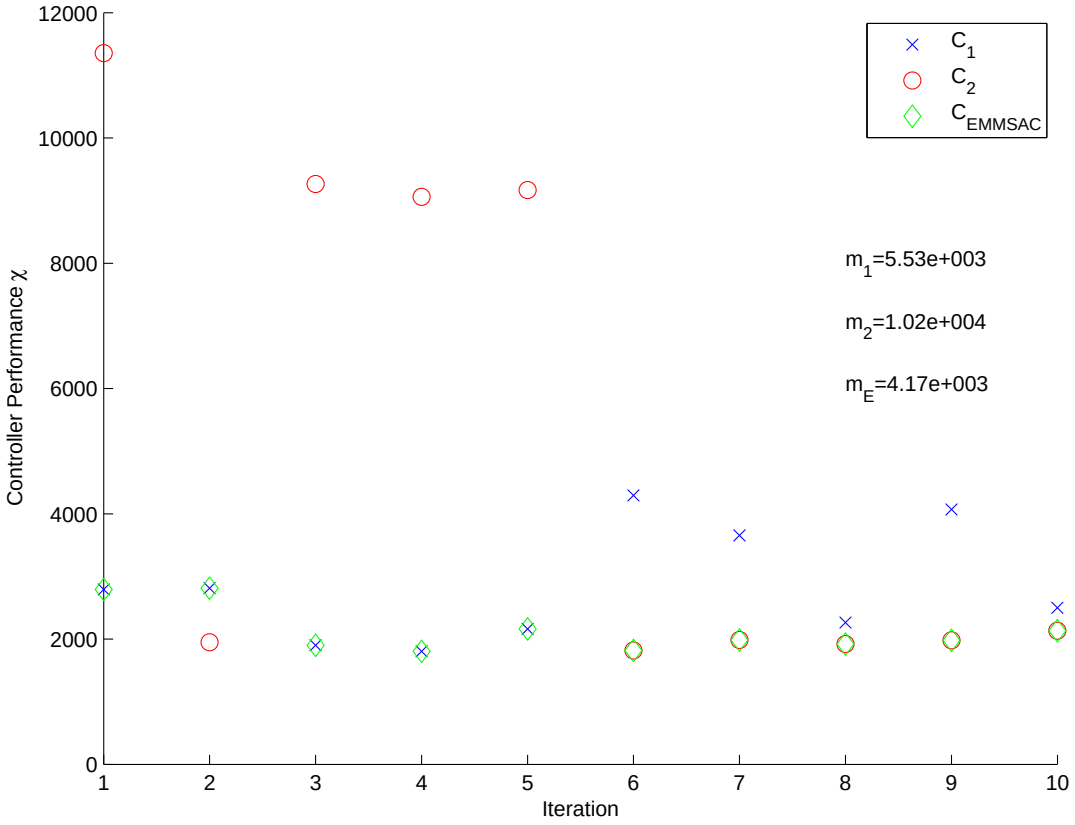

Figure 4.15: Comparison of controller performance.

4.3 Image Generation Using EMMSAC Algorithm

The purpose of this section is to illustrate the imaging results obtained using the EMMSAC algorithm for two samples. The first sample is a calibration grating TGZ01 from MikroMasch. The grating consists of 1-D arrays of rectangular SiO_2 steps on a Silicon wafer. The step height is 120 nm with a pitch of 3 μm. The second sample is a DVD which is made of a relatively softer polymer. This sample consists of 1-D *tracks* which have a pitch of 0.75 μm. Each scan covers and area of $3 \times 1.17\mu$m and the scans are carried out at a rate of 0.5 Hz i.e. each scan line takes one second for completion. The relatively small scan area is a consequence of limitations imposed by the AFM and signal processing hardware. However since the structures being viewed vary periodically and have a pitch less than or equal to 3 μm, this scan size is sufficient for these samples.

The experimental set up is exactly the same as mentioned in Section 4.2.1. The first subsection describes the generation of the plant set. Due to the reasons mentioned in Section 3.2.2, the controller set is generated in a slightly different manner and is described in the second subsection. The last subsection describes the images obtained using the EMMSAC algorithm.

4.3.1 Plant Set

The plant set is generated by testing a set of trial PI controllers in the set C_T given below,

$$(K_p, K_i) \in C_T = [0.1, 0.2, \ldots, 1.0] \times [10, 20, \ldots, 100] \tag{4.16}$$

Here K_p and K_i represent the proportional and integral gains of the PI controller used in the AFM control loop. The range and separation of the gains is chosen empirically. This combination of P and I gains results in 100 controllers in the trial controller set C_T.

The plant set is determined through a series of data collection experiments, one for each sample. In each experiment all the controllers given in the trial set C_T are tested twice, once for a *trace* line and once for a *retrace* one. The terms trace and retrace refer to consecutive scan lines in a raster scan pattern. Each scan line takes one second to execute. During this scan five signals are recorded namely $A_{sp}, A, h, f_D, d_{TM}$. The testing of each controller generates two data sets D^i_{jT} and D^i_{jR} given below,

$$\begin{aligned} D^i_{jT} &= [A_{sp},\ A,\ h,\ f_D,\ d_{TM}] \in \mathbb{R}^{10^6 \times 5} | 1 \leq j \leq 100, 1 \leq i \leq 2 \\ D^i_{jR} &= [A_{sp}, A,\ h,\ f_D,\ d_{TM}] \in \mathbb{R}^{10^6 \times 5} | 1 \leq j \leq 100, 1 \leq i \leq 2 \end{aligned} \tag{4.17}$$

Here j refers to the controller being tested, i equals the sample. In this case $i = 1$ corresponds to the TGZ01 calibration grating and $i = 2$ corresponds to the DVD. The subscripts T and R correspond to trace and retrace lines respectively. The complete data $\mathbb{D}$ set is thus given as,

$$\mathbb{D} = \begin{bmatrix} D^1_{1T} & \dots & D^1_{100T} & D^1_{1R} & \dots & D^1_{100R} \\ D^2_{1T} & \dots & D^2_{100T} & D^2_{1R} & \dots & D^2_{100R} \end{bmatrix} \tag{4.18}$$

Next, the dither-piezo input f_D and cantilever deflection d_{TM} in each data set D^i_j are used to determine the closest second order LTI plant G^i_j using the same mapping Φ_p described in the previous section. Each data set leads to one plant. The final plant set $\mathbb{G}$ can thus be given as,

$$\mathbb{G} = \begin{bmatrix} G^1_{1T} & \dots & G^1_{100T} & G^1_{1R} & \dots & G^1_{100R} \\ G^2_{1T} & \dots & G^2_{100T} & G^2_{1R} & \dots & G^2_{100R} \end{bmatrix} \tag{4.19}$$

Here each column corresponds to the testing of a single controller and each row corresponds to one system identification experiment. The set thus contains 400 plants. Although it would be ideal to use the entire plant set for evaluation of the switching signal using the mapping Φ_q, this is computationally infeasible for image generation on a physical system.

Therefore a reduced plant set $\mathbb{G}_R$ is used. The set contains only the trace plants, i.e the plants with the subscript T in the original plant set $\mathbb{G}$. Amongst the trace plants, it uses only the plants which were generated using the odd controller gain entries. This means that plants generated using $K_p = [0.1, 0.3, 0.5, 0.7, 0.9]$ and $K_i = [10, 30, 50, 70, 90]$ were used. The reduced plant set thus has 25 plants in each row, and 50 plants in all.

Figure 4.16 illustrates the reduced plant set. As can be seen, the plants corresponding to the two sample cluster at slightly different locations.

4.3.2 Controller Set

The system identification data is used once again to select the best controller for each experiment. This is done by evaluating the error signal for the trace $e_T = A_{sp} - A$ and retrace $e_R = A_{sp} - A$ for each controller test in an experiment. The performance metric χ for each controller is then given below,

$$\chi = \sum_{k=1}^{10^6} e_T^2(k) + \sum_{k=1}^{10^6} e_R^2(k) \tag{4.20}$$

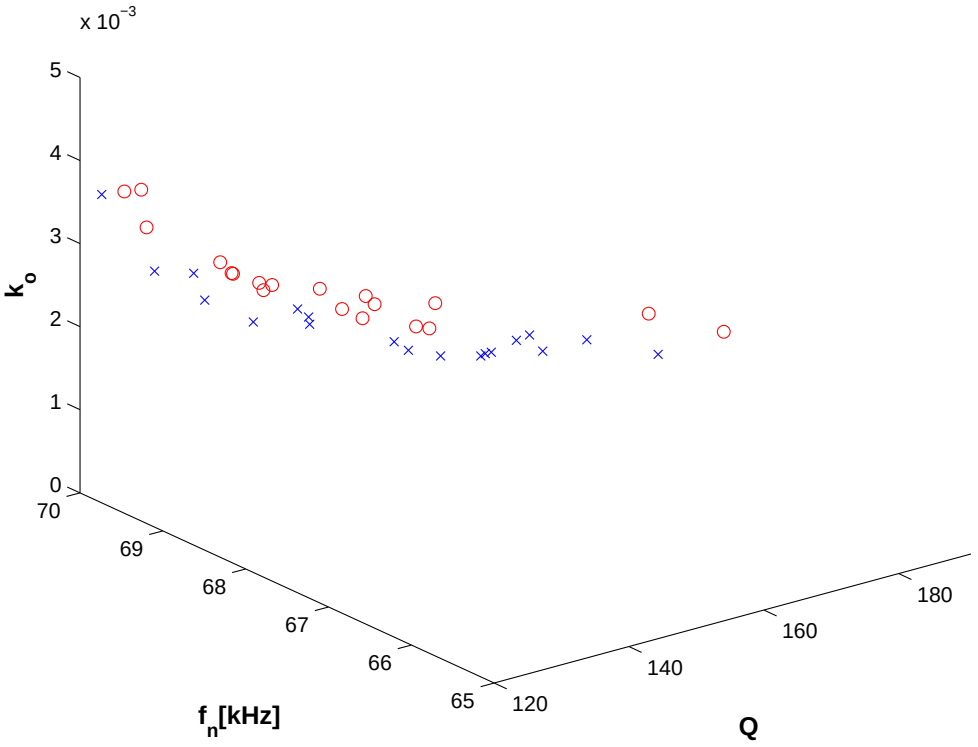

Figure 4.16: Plant set. (Blue X:TGZ Red O:DVD)

Here e_T is the error for the trace line and e_R is the error for the retrace line. For each experiment the controller with the lowest χ is selected. However as mentioned previously in Section 2.7.1 the squared 2-norm of the error signal does not take into account if the control signal used to construct the sample image is oscillatory and if the trace and retrace signals are close. As a consequence two constraints need to be used. Firstly, controllers that result in a *large* difference between trace and retrace height signals are discarded. Secondly, controllers that cause *significant* oscillations in the height signal are also discarded. Finally, controllers that result in height signals that are nearly *flat* i.e., indicating no interaction with the sample are also discarded. These constraints are imposed by two metrics given below,

$$TR = \sum_{k=1}^{10^6} (h_T(k) - h_R(k))^2 \tag{4.21}$$

where h_T is the height signal for a trace and h_R is the height signal for a retrace line. A higher value of TR indicates greater mismatch between trace and retrace lines. In addition a very low value implies that the height signals are nearly flat and there is no significant interaction between the cantilever and the sample.

The second metric PSD is the power spectral density of the height signals for frequencies greater than 10 Hz. This frequency was chosen empirically based on the a priori

knowledge of the sample topographies used in this contribution and the scan speed. If a controller results in a height signal with substantial energy in frequencies above 10 Hz, this implies that the gains are too high and are producing erroneous height signals. Likewise, a very low PSD value again indicates flat height signals and therefore no significant interaction between cantilever and sample. The measure is given below,

$$PSD = \sum_{f=10}^{500000} |H_T(f)|^2 + \sum_{f=10}^{500000} |H_R(f)|^2 \tag{4.22}$$

Here H_T and H_R are the Fast Fourier Transforms of the trace and retrace height signals respectively. The upper frequency limit in the PSD calculation i.e. 500,000 Hz is chosen to be half of the signal sampling frequency of the DAQ card.

Finally a metric matrix Φ is constructed for each experiment as follows,

$$\Phi^i = \begin{bmatrix} \chi_1 & \chi_2 & \cdots & \chi_{100} \\ TR_1 & TR_2 & \dots & TR_{100} \\ PSD_1 & PSD_2 & \dots & PSD_{100} \end{bmatrix} \tag{4.23}$$

Here the superscript i corresponds to the sample. Now the best controller for the sample is chosen using the following relation,

$$C^i = \arg\min_j \chi_j \Big| \begin{matrix} TR_L < TR_j < TR_U \\ PSD_L < PSD_j < PSD_U \end{matrix} \quad \forall \;\; 1 \leq j \;\; \leq 100 \tag{4.24}$$

here TR_L, TR_U are the 5^{th} and 80^{th} percentile for all TR values for the i^{th} sample. Likewise PSD_L, PSD_U are the 5^{th} and 30^{th} percentiles for all PSD values in the i^{th} sample. These limits are chosen empirically and are meant to discard controllers with unacceptable performance.

The final controller set then consists of two controllers, one corresponding to each sample and is illustrated in Table 4.2.

Table 4.2: Controller set.

q	K_p	K_i
0	0	50
1	0.3	90
2	0.4	60

Here $q = 0$ corresponds to the default controller used to initiate the imaging process for both sample. The controller is deliberately chosen to be suboptimal. This results in

poor performance in the first. This is done to determine if the EMMSAC algorithm can identify the correct sample despite a poorly performing controller.

4.3.3 Image Generation

The complete AFM control loop along with the EMMSAC implementation is illustrated in Figure 4.17.

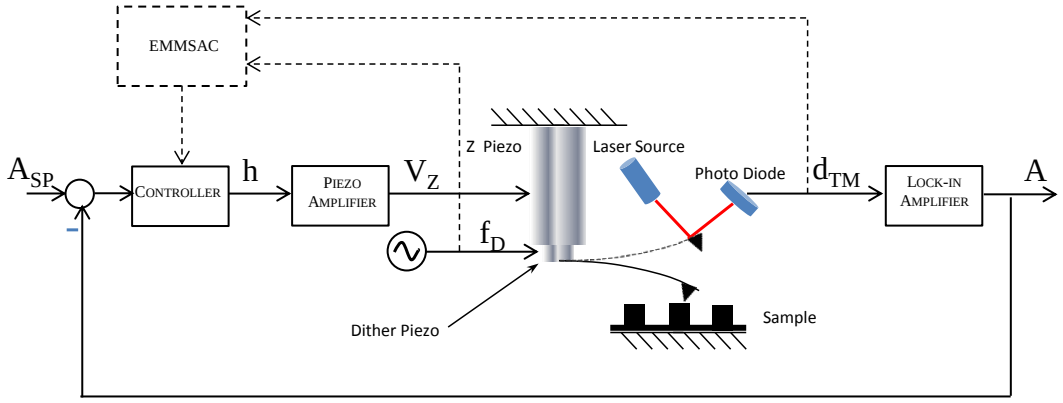

Figure 4.17: Implementation of the EMMSAC algorithm with the AFM in the loop.

As illustrated, the entire AFM control loop remains unchanged. The only addition is the EMMSAC block which uses the cantilever drive signal f_D and the tapping mode deflection d_{TM} as its inputs and generates the switching signal as its output. Due the computationally intensive nature of the EMMSAC, the algorithm is run at three uniformly spaced intervals during the scan starting immediately after the first line. During each EMMSAC run, the dither-piezo input f_D and the cantilever deflection signal d_{TM} for a single trace line are saved. Next the switching signal q is evaluated using the mapping Φ_q as described in the previous section.

Finally the controller corresponding to the switching signal is put into the AFM control loop and is allowed to stay in the loop until the start of the next consecutive EMMSAC run. During each EMMSAC run a total of 2×25 plant residuals are evaluated. This requires 50 second on an Intel 2.67 GHz desktop. During this interval while the residuals are being computed the AFM scan is halted, however the DSP continues to Amplitude signal (A) by controlling the Height (h). The generation of the complete image requires approximately 10 minutes.

Figure 4.18 illustrate the images obtained and the switching signal after each EMMSAC run. The TGZ 02 calibration sample consists of SiO_2 steps on Si wafer with a height of 120 nm and pitch of 3μm. In Figure 4.18 a single step can be viewed. Ideally a

greater scan area is preferable to enable viewing of a greater number of steps, however the current AFM instrument and the control hardware limit the maximum permissible scan area to approximately $3 \times 1.17\mu$m. Each scan consists of 100 scan lines.

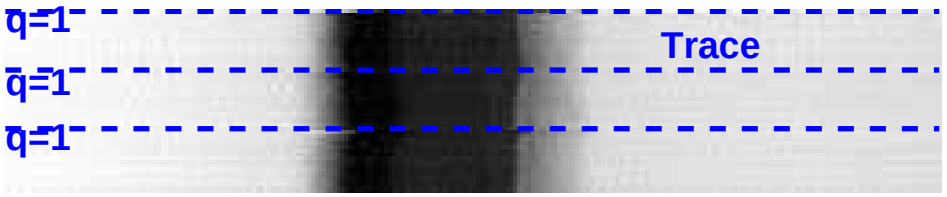

Figure 4.18: TGZ calibration sample image.

Figures 4.19 illustrate the sample heights for a single trace and retrace line. The height signals are chosen for illustrating the results since the sample feature heights are a priori known and hence serve as a bench mark. As can be seen, that although there is still a small amount on noise present in the system, the topography for the TGZ 02 sample has an error of approximately 20 nm,the trace and retrace signals are fairly similar and the sample topography is clearly revealed.

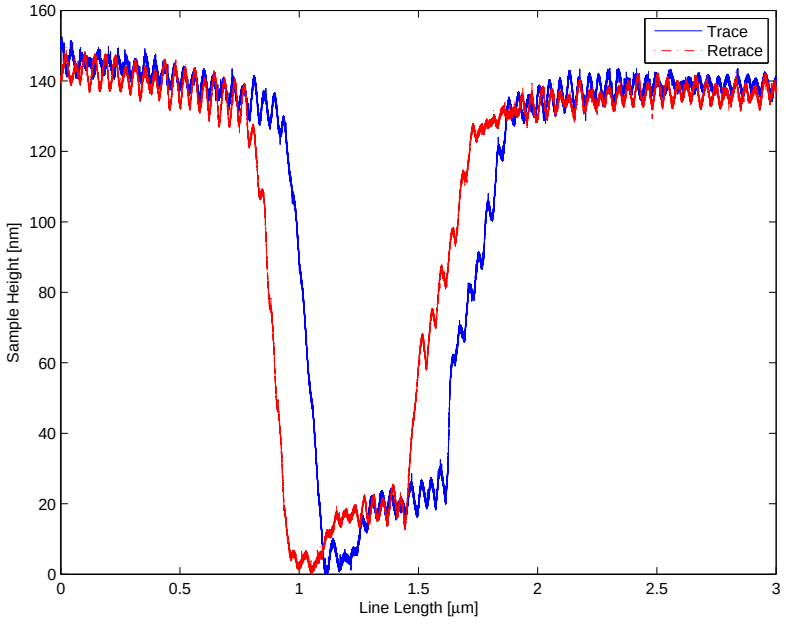

Figure 4.19: TGZ calibration sample single line image.

Finally, Figure 4.20 reveals the controller performance for every line scanned. The controller used for line 1 is $K_p = 0$ and $K_i = 5$. This is a known suboptimal controller. It is chosen to determine if the EMMSAC can generate an image even when starting

from a poorly tuned controller. As can be seen the sum squared error χ drops after the first line and remains below the starting value.

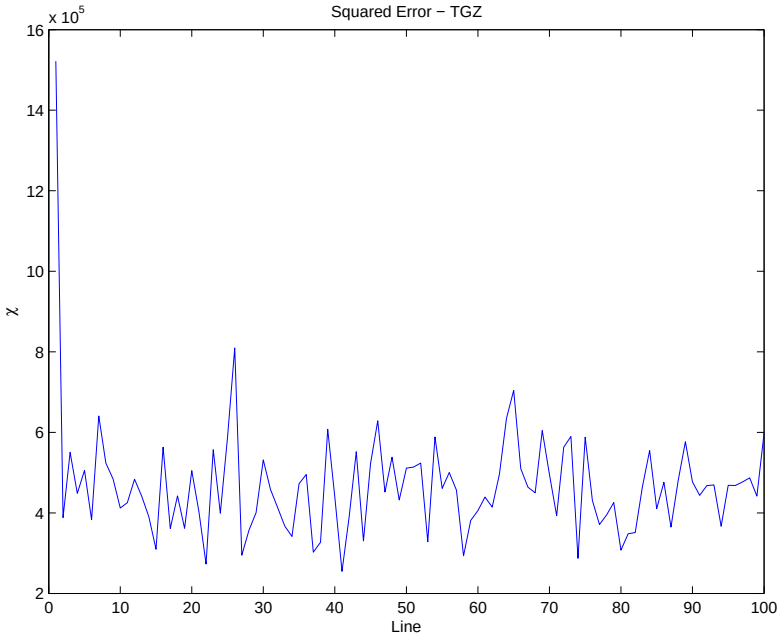

Figure 4.20: EMMSAC controller performance on TGZ sample.

Figures 4.21, 4.22 and 4.23 illustrate the same results for the case of the DVD sample.

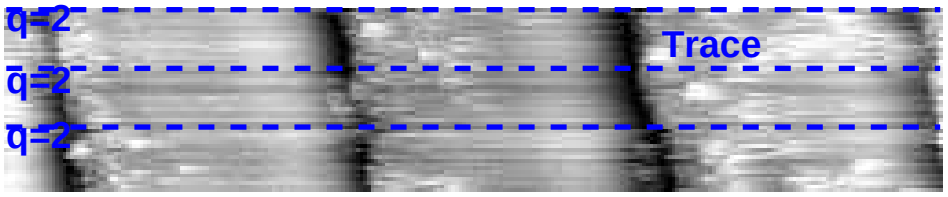

Figure 4.21: DVD sample image.

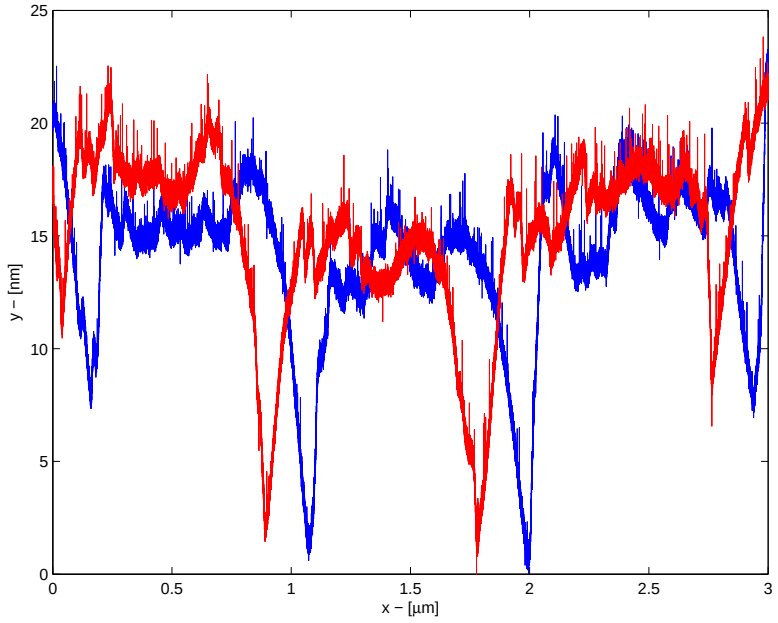

Figure 4.22: DVD single line image.

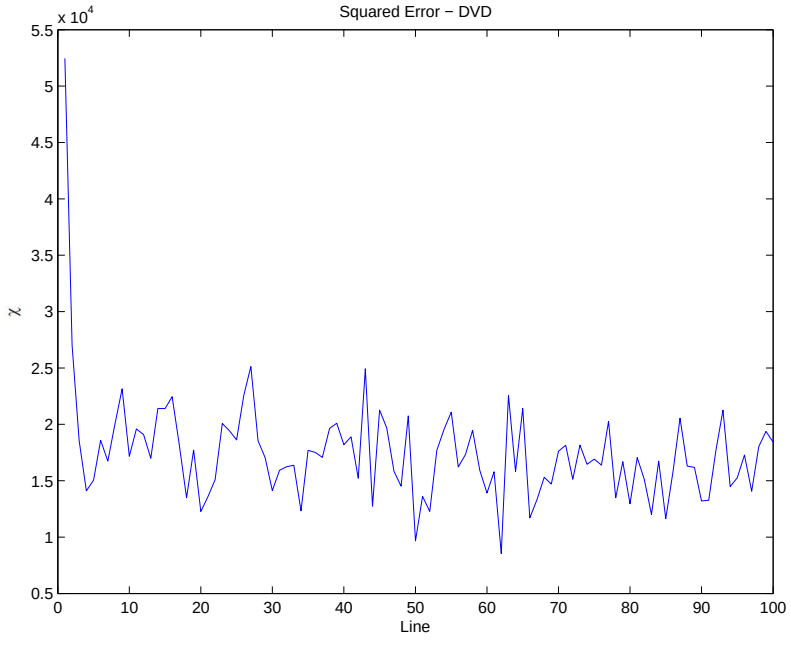

Figure 4.23: EMMSAC controller performance on DVD sample.

4.4 Image Generation Using Manual Tuning

This section presents the images generated for the same sample using manual tuning of the PI controllers. Figures 4.24 and 4.25 illustrate the sample image and a single scan line respectively.

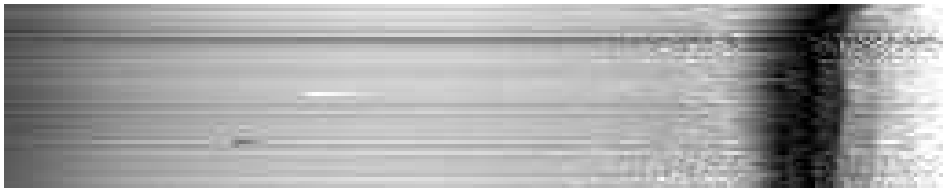

Figure 4.24: TGZ calibration sample image.

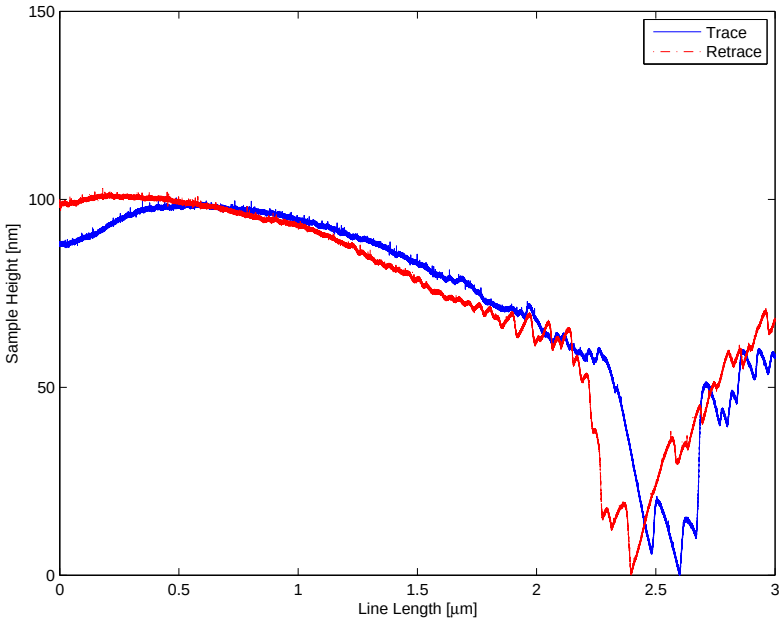

Figure 4.25: TGZ calibration sample single line image.

The controllers were tuned as the image was being generated by observing the sum squared error for each line. Figures 4.26 and 4.27 illustrate the squared error and PI gains which were used during the tuning procedure. The integrator gain in Figure 4.27 has been scaled down by a factor of ten for the purpose of illustration.

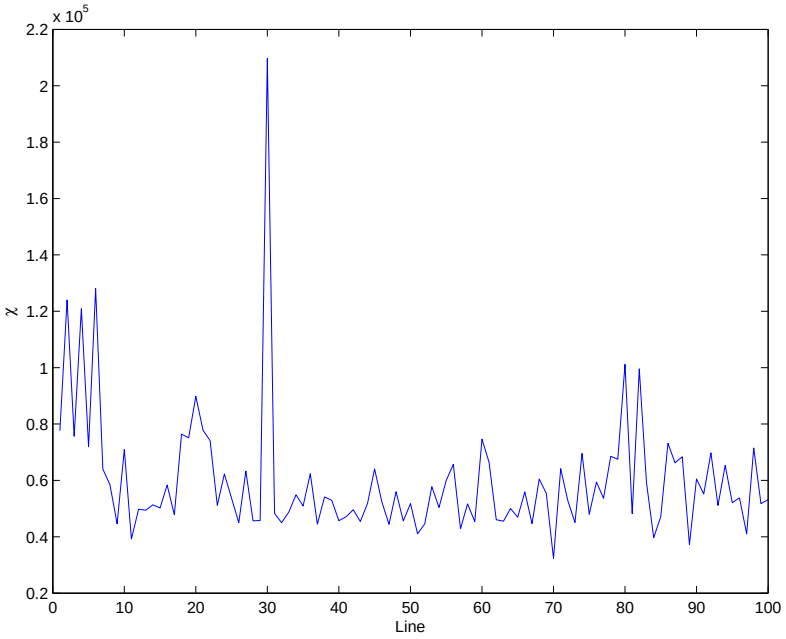

Figure 4.26: Hand-tuned controller performance on TGZ sample.

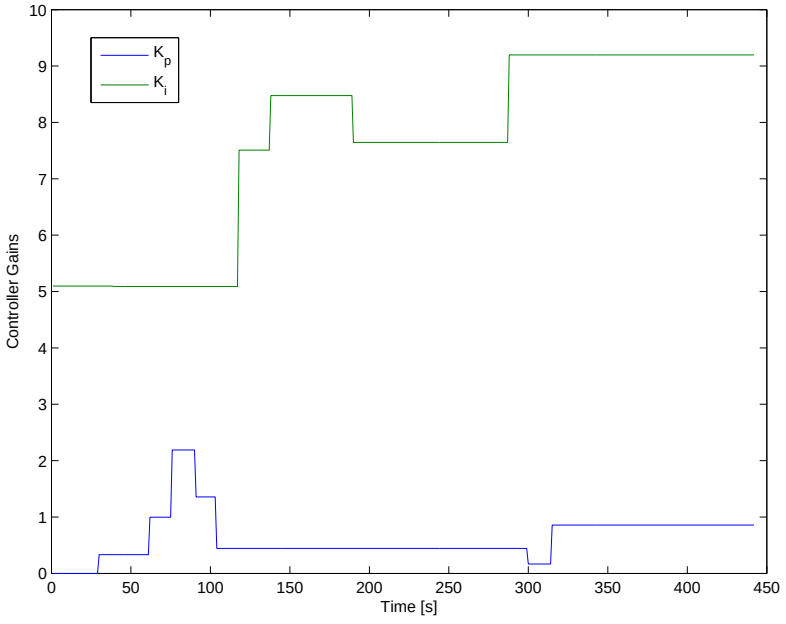

Figure 4.27: Manual K_p, K_i gain selection.

While both approaches (automated and manual) can estimate the sample topography, the advantage of using the automated approach is that it eliminates the manual gain tuning effort illustrated in Figure 4.27. Although the hand-tuned controller gets a better performance in terms of the performance metric χ, the relevant aspect in this case is that the EMMSAC controller generates images without any human effort. Clearly the

performance of the EMMSAC can be improved, however the objective of generating a sample image without hand-tuning is attained by the EMMSAC system.

4.5 Conclusions

The experimental results presented in this chapter clearly demonstrate that the EMMSAC algorithm can automate the controller tuning procedure which otherwise needs to be performed manually. The algorithm can identify temporal variations in the AFM system as well as a change of sample. In each case the best controller is switched in, which is exactly what a human user would do. It takes away the need for manual tuning of the controller gains and clearly exhibits the potential of making AFM easier to use.

It is clear that the image quality needs to be improved. This can be achieved by making the generation of the Kalman filter residuals faster and having a higher number of candidate plants and their controllers. It may also be beneficial to investigate adaptive plant sets as compared to fixed ones as in the current case.

Lastly, a better performance metric needs to be determined. The current metric of choice has two limitations which are detailed below.

• The first limitation is that the error signal is affected by the system noise. As evident from the Figures 4.20 and 4.23 in the last two sections, the sum squared error varies considerably from one scan line to the next even in the same experiment with the same sample and controller. This makes it difficult to determine the effectiveness of switching in the best controller. The ideal situation would be one where switching in the false controller clearly deteriorates the image quality, as illustrated in the simulation results in Section 3.2.3.

• The second limitation is that the metric is not strongly correlated with image quality. This is due to at least two reasons. Firstly, the tip vibration amplitude A does not vary linearly with respect to the cantilever base height Z. If Z is reduced, i.e the cantilever base moves towards the sample surface the amplitude reduces linearly. However if the base is moved away from the surface, the amplitude increases at first but then saturates. This is simply because there is a physical limit to the vibration amplitude for any cantilever. The variation of A with respect to Z is illustrated below,

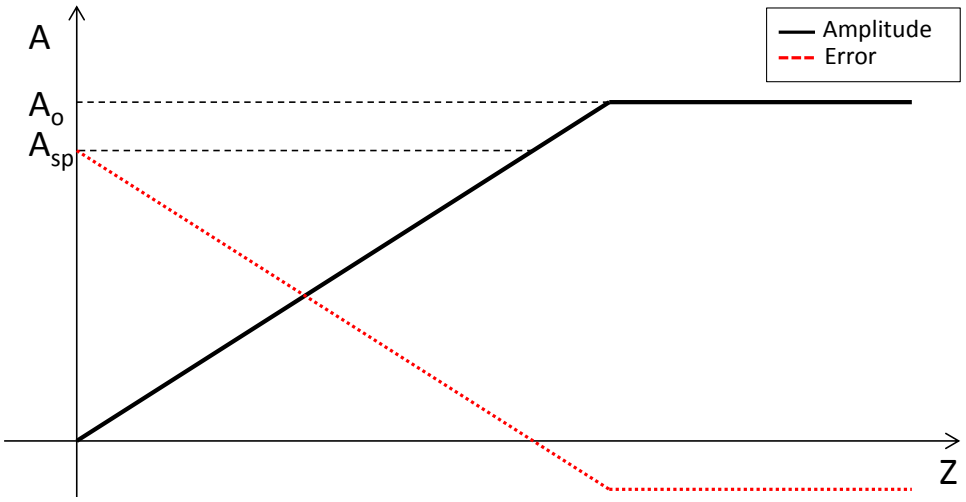

Figure 4.28: Variation of amplitude A and error e with respect to cantilever base distance (Z) from sample surface.

As illustrated, if the cantilever base moves away from the sample surface, the error $e = A_{sp} - A$ increases a little and then saturates. Thus if a weak controller is switched in, the cantilever base will stay relatively far from the sample surface but error will saturate to a small negative value. It will not reflect the ineffectiveness of the controller and the estimated topography will be relatively flat.

The second reason why the metric is not correlated well with image quality is that a strong controller can result in an oscillatory control signal. Since that the image is generated from the control signal, the oscillations show as distortions which immediately reduces the image quality. However the current metric does not reflect this in a substantial manner.

To conclude, the experimental results reported in this chapter demonstrate that EMMSAC algorithm can use AFM data to switch in the correct a priori known optimal controller. While the Kalman residual based identification mechanism works, the performance improvement obtained is marginal. This is however not a limitation of the algorithm itself but of the performance metric used. In addition the abstract and subjective nature of image quality must also be taken into account. The issue of human perception of image quality has been addressed previously in literature, for instance by Wang et al. (2004). The authors note that image quality has traditionally been measured using the difference between a reference image and a distorted image. This is often done through metric like the Mean Squared Error(MSE) or the Peak Signal to Noise Ratio (PSNR). They observe that human vision is adapted for *extracting structural information from a scene*. In this context they suggest a Structural Similarity Index and demonstrate that it can be used as a performance measure for image quality. Such a performance measure is relevant to the case of AFMs since a reference measure

is not available and the end users generally have a priori information about the possible structure that they wish to view.

The next chapter investigates two novel strategies to reduce imaging times in AFMs. The first one generates sample images using coarse scanning trajectories and compressive sensing. The second one speeds up imaging by scanning the sample in a *parallel* manner. This is done by using the entire length of the beam, instead of just the tip.

Chapter 5

Faster AFM Imaging

While the EMMSAC based automated controller tuning procedure described previously makes AFMs more accessible to the general scientific community, the issue of long imaging times still needs to be investigated. The purpose of this chapter is to illustrate two methods that have the potential to reduce AFM imaging times, namely *Compressive Sensing* and *Full Beam Imaging*. The remaining two sections explain these methods in detail. After the description of the Compressive Sensing method it will also be made clear that the automated tuning procedure is beneficial to this method as well.

5.1 Compressive Sensing

One key disadvantage associated with AFMs is that the cantilever needs to be moved sequentially over the entire sample in a raster pattern. For the purpose of generating a complete image of the sample the raster pattern needs to be fine enough. Coarser trajectories can be used, but while this will improve imaging speed it will be at the expense of missed regions in the sample image. However if the sample is periodic in nature, i.e. it has repeating patterns then it is possible to build a sample image using relatively coarser trajectories. In this context an emerging signal processing method called compressive sensing offers a possible solution. This method functions by using a limiting number of measurements from a sub-sampled signal and searching for the sparsest equivalent representation in a suitably chosen transformed domain. The sparse equivalent is then used to generate an estimate of the original sub-sampled signal using the same transformation. This method is functional even when the sampling rate is below the minimal limit determined by the Nyquist principle (Baranuik (2007)). The remaining parts of this section describe the concept of compressive sensing in detail and how this is used to generate images of AFM samples using coarse trajectories.

According to the Nyquist principle a signal must be sampled at twice the signal bandwidth to avoid losing any information. This requires many applications to sample signals

at very high rates, which in turn necessitates greater storage space and the need for compression. The conventional method therefore is to sample signals at very high rates, compress them and then continue with storage or transmission. This signal compression relies on the principle that the signal being measured can be represented in terms of a set of basis vectors. In addition it is also required that out of this generic set of basis vectors only a small number of vectors is sufficient to represent the signal. Let the signal be an N element vector $x \in \mathbb{R}^N$. This can then be represented using a set of N basis vectors as follows,

$$x = \sum_{i=1}^{N} p_i \psi_i \tag{5.1}$$

or in matrix notation,

$$x = \psi p \tag{5.2}$$

where $\psi_i \in \mathbb{R}^N$ is the i^{th} basis vector, $\psi = [\psi_1 \psi_2 \dots \psi_N]$ and $p \in \mathbb{R}^N$ is a vector of weighting coefficients. Signal compression then relies on the fact that the signal can be represented using only $K << N$ of these coefficients, i.e only K coefficients are non-zero and the remaining are zero. Once these coefficients are determined, only the K non-zero ones are either stored or transmitted. Such a signal is referred to as *K-Sparse* or *Compressible*.

Figure 5.1 illustrates the K-sparse concept using a simple illustration. It shows a unity amplitude one hertz square wave that is reconstructed using three and seven harmonics. As can be seen that the reconstructions although not, accurate do represent the shape of the original signal. Even though there are infinitely many harmonics in a square wave, the signal shape can be approximated using a small number of these and ignoring the rest.

There is one major disadvantage associated with this procedure. Acquiring a signal at high sampling rates can be expensive and time-consuming. This is especially true for the current application. Scanning a topography using an AFM is a slow process, and it will be very beneficial if the topography can be reconstructed without having to scan the entire surface. The remaining parts of this section explain what is compressive sensing, how does it work despite sampling at sub-Nyquist rates and how is this applicable to atomic force microscopy.

Compressive sensing assumes that the signal of interest is K-Sparse, and defines a new form of measurement. Instead of measuring a signal sample by sample in the time domain, each measurement is a dot product of the complete signal x and the *Measurement Vector* ϕ_i. Let the i^{th} measurement be y_i, and let there be $M < N$ measurements. The

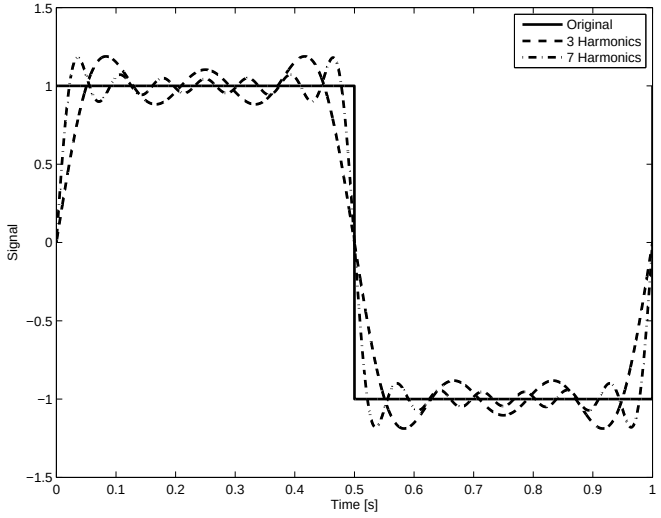

Figure 5.1: Square wave reconstruction using three and seven harmonics.

i^{th} measurement is then $y_i =< x, \phi_i >$ and the complete vector of M such measurements can then be written as,

$$y = \phi x = \phi\psi s = As \tag{5.3}$$

where ϕ is a matrix whose i^{th} row is given by ${\phi_i}^T$. This matrix is referred to as the *Measurement Matrix*. A is an $M \times N$ matrix which results from the multiplication of ϕ, ψ and s is the sparse representation of x in the transformed domain so that $x = \psi s$. The next problem is the selection of the appropriate measurement matrix ϕ, basis matrix ψ and a method to recover the original signal x from the measurements y. These details are given below.

Measurement Matrix ϕ: It is now well established (Baranuik (2007)) that any measurement matrix is suitable if the resulting A matrix satisfies the following condition,

$$1 - \epsilon \leq \frac{||Av||_2}{||v||_2} \leq 1 + \epsilon \tag{5.4}$$

where $v \in \mathbb{R}^N$ is any arbitrary $3K$ sparse vector, and ϵ is any constant greater than zero. This condition is known as the *Restricted Isometry Property*. Furthermore it is also possible to prove that a measurement matrix can exhibit the restricted isometry property with high probability (Candes et al. (2006)) if two conditions are satisfied. Firstly, all the entries are independent and identicaly distributed random numbers. These should have a Gaussian probability distribution function with a mean of zero and variance $\frac{1}{N}$, i.e. $\phi_{ij} \sim \mathcal{N}(0, \frac{1}{N})$. Secondly, the number of measurements is sufficient, i.e $M \geq cKlog\left(\frac{N}{K}\right) << N$ where c is a small constant. An interesting property of a Gaussian measurement matrix is that $A = \phi\psi$ will also be independent identically distributed Gaussian, thus enabling the usage of any orthonormal basis ψ.

Basis Matrix ψ: The selection of the basis depends entirely on the application and are selected empirically. For most applications involving imaging of natural scenes the *Discrete Cosine Transform* DCT basis is suitable (Taubman and Marcellin (2002)). Alternatively the use of wavelets and curvelets etc. has also been reported in literature (Ma (2010)).

Signal Recovery Algorithm: The signal recovery problem now is to simply determine a sparse $N \times 1$ vector s' such that $y = As'$. The classical approach is to determine a minimal L_2 norm solution which yields the solution,

$$\begin{aligned} \hat{s} &= \text{argmin}||s'||_2 \;\; s.t \;\; y = As' \\ \hat{s} &= A^T \left(AA^T\right)^{-1} y \end{aligned} \tag{5.5}$$

Since attempting to find a minimal L_2 norm solution attempts to find a minimal energy solution, the solution found is usually not sparse and has lot of non-zero elements. Therefore the minimal L_2 norm solution is not feasible.

A natural solution to this problem is to find a solution vector s' that minimizes the L_0 solution. Since the L_0 norm counts the number of non-zero elements in a vector, it appears to be a suitable choice for finding sparse solution. The solution can thus be written as,

$$\hat{s} = \text{argmin}||s'||_0 \;\; s.t \;\; y = As' \tag{5.6}$$

While this method can determine a sparse signal, the problem is that the above mentioned optimization problem is NP complete (Baranuik (2007)) and therefore very time-consuming.

The next alternative is find a solution that minimizes the L_1 norm, i.e

$$\hat{s} = \text{argmin}||s'||_1 \;\; s.t \;\; y = As' \tag{5.7}$$

The L_1 norm minimization algorithm attempts to find the sparse solution in the $n - m$ dimensional shifted null space of A which is $\bar{\mathcal{N}} = \mathcal{N}(A) + s$. Here $\mathcal{N}$ is the null space of A. Fortunately, this is a convex optimization problem which is much more feasible computational.

The reason why the L_1 norm minimization is well suited for finding sparse solutions is better understood using the simplified illustration in Figure 5.2 originally reported in Baranuik (2007). Although the original optimization problem has more then three

dimensions, the figure attempts to explain the process using a simplified three dimensional illustration. The left part of the figure represents a L_2 hypersphere and the right one illustrates a L_1 hypersphere. The point s represents a sparse solution, which lies in the shifted null space of A represented by the hyperplane. This fact that the solution is sparse is illustrated by the fact that it lies on a coordinate axis. In general the hyperplane is likely to be inclined at an angle due to the random nature of A.

The optimization then proceeds by expanding the L_2 and L_1 hyperspheres until they intersect with the shifted null space at $\hat{s}$. Due to the *pointed* nature of the L_1 hypersphere it is more likely to intersect with the shifted null space hyperplane $\bar{N}$ on a coordinate axis. On the other hand the expanding L_2 ball is likely to intersect the hyperplane at a point that does not lie on any single axis and is instead a linear combination of them. This makes the solution determined by the L_2 norm minimizer less likely to be sparse.

Once s' is known the original signal x can be recovered using $\hat{x} = \psi\hat{s}$

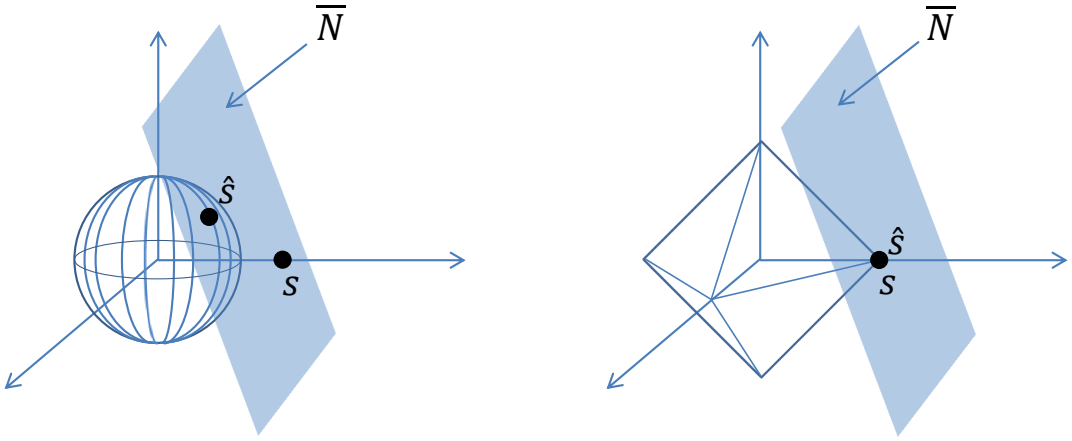

Figure 5.2: Graphical visualization for L_1, L_2 norm minimization (Baranuik (2007)).

5.1.1 Simulation Results

Although the concept of using Compressive Sensing for reducing AFM imaging times was developed independently during this research, it was reported first by Song et al. (2011). While the author provides experimental results using rectangular scan trajectories, this research provides experimental results using spiral scan trajectories. The advantage is that there are no drastic changes in direction, which in turn reduces the possibility of exciting the piezo-actuator harmonics. The precise meaning of the term scan trajectories and an illustration for the spiral scan trajectory shall be provided next in this section.

Before the concept can be applied it was tested on pre recorded images from AFMs. The unknown signal x in this case is simply the vectorized image shown in Figure 5.3. For these simulations a 128×128 pixel image of a VGRP-15M calibration sample is used. In this case x is a $1 \times n$ vector where $n = 128^2 = 16384$. The next step is to extract m measurements from x and generate the measurement vector y. This is done by multiplying x with the measurement matrix ϕ.

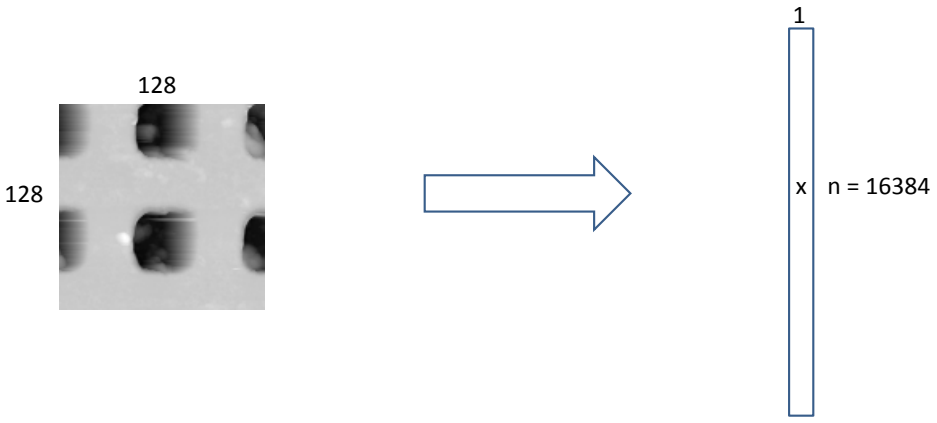

Figure 5.3: Image vectorization.

The measurement matrix is generated simply by *overlaying* the selected trajectory over the image. This is illustrated using a trajectory example in Figure 5.4 . If the trajectory passes through the region covered by a pixel, that pixel is considered to be known otherwise it is unknown. The + operator here indicates the overlay operation. The number of measurements m is then equal to the number of known pixels. The rows of the measurement matrix are then generated by setting each entry corresponding to a known pixel equal to one and the others to zero. This will generate a measurement matrix with all rows identical. It was observed that such a measurement matrix does not work. A possible cause might be that this measurement matrix does not satisfy the Restricted Isometry Property. This problem was solved by randomly setting all ones in the matrix to zero with a probability of 0.5. The motivation behind this solution is to duplicate the randomness of measurement matrix as is done in the original compressive sensing algorithm. Clearly this needs to be investigated further. The problem is that it is not possible to take inner products with the entire unknown signal x therefore the original Gaussian random measurement matrices cannot be used.

The overlay procedure is described in detail next.

Overlay Procedure : The complete scan data is recorded in the data matrix Ξ given below,

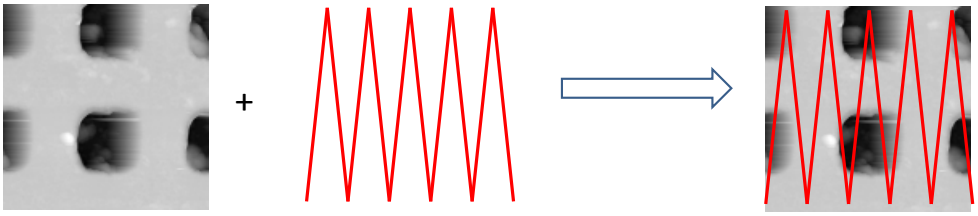

Figure 5.4: Trajectory overlay illustration.

$$\Xi = \begin{bmatrix} x_1 & y_1 & h_1 \\ x_2 & y_2 & h_2 \\ \vdots & \vdots & \vdots \\ x_N & y_N & h_N \end{bmatrix} \tag{5.8}$$

The i^{th} pair of points $x_i, y_i \in \begin{bmatrix} -\frac{L}{2} & \frac{L}{2} \end{bmatrix}$ in the first two columns represents the i^{th} point on the trajectory. The corresponding topography at that point is then represented by h_i. Here N is the total number of points on the trajectory and L is the length of the square scan area. Next, the trajectory vectors are discretized. The discretization process maps each trajectory point to a square grid of $n_p \times n_p$ pixels. For the current example the number of pixels $n_p = 128$. The discretization map $\mathcal{Q}$ is given below,

$$\mathcal{Q} : \Xi \rightarrow \Xi^D \tag{5.9}$$

where,

$$\Xi^D = \begin{bmatrix} x_1^D & y_1^D & h_1 \\ x_2^D & y_2^D & h_2 \\ \vdots & \vdots & \vdots \\ x_N^D & y_N^D & h_N \end{bmatrix} \tag{5.10}$$

The i^{th} pair of points $x_i^D, y_i^D \in [1 \quad n_p]$ in the first two columns of Ξ^D represents the i^{th} discretized point on the trajectory. Both $x_i^D, y_i^D \in \mathbb{N} \quad \forall\, i \in [1 \quad N]$. The discretization map $\mathcal{Q}$ can be easily understood through the following figures,

The fact that the discretization map for the vertical coordinate is different from the horizontal coordinate is due to the fact that during mapping an area onto an image the origin is shifted to the top left as illustrated in Figure 5.6.

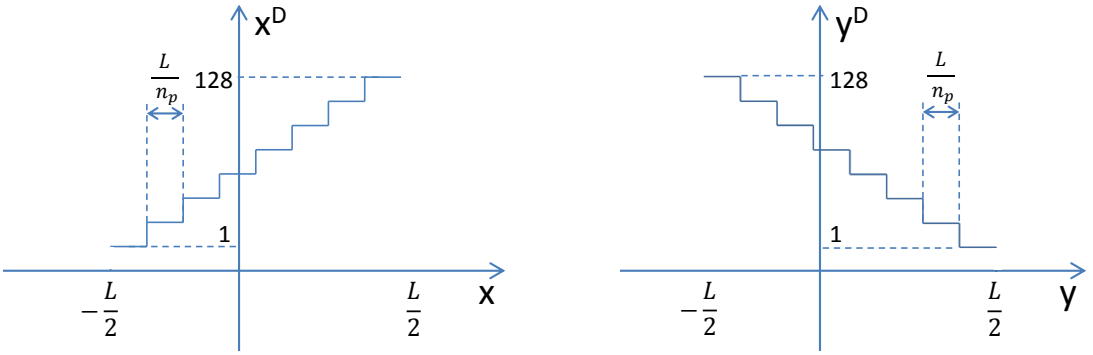

Figure 5.5: Discretization map $\mathcal{Q}$.

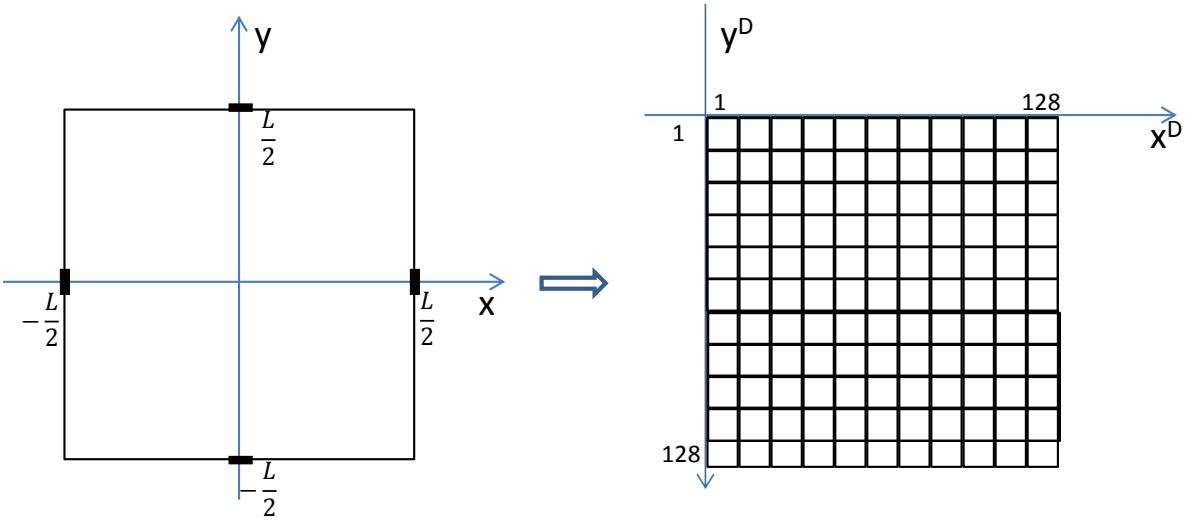

Figure 5.6: Mapping an area to an image.

In the next step two matrices $X, I \in \mathbb{R}^{n_p \times n_p}$ are created with all entries initially equal to zero. These matrices are used to generate the vector x and the measurement matrix ϕ. This is done by setting all entries in X corresponding to $x_i^D, y_i^D \quad \forall \; i \in [1 \quad N]$ equal to h_i and, all entries in I corresponding to $x_i^D, y_i^D \quad \forall \; i \in [1 \quad N]$ equal to unity. Finally both matrices are column wise vectorized to yield x and ι. The measurement matrix is then generated using,

$$\phi = \begin{bmatrix} \iota_1^T \\ \iota_2^T \\ \vdots \\ \iota_m^T \end{bmatrix} \times \Gamma \tag{5.11}$$

where, $\iota_1 = \iota_2 \ldots \iota_m = \iota$, m is the number of entries equalling unity in I and $\Gamma \in \mathbb{R}^{m \times n_p^2}$ is a random binary matrix. Each entry in Γ is an independent, identically distributed binary random variable $\sim Binary(\frac{1}{2}, \frac{1}{2})$. Thus each entry is either zero or unity with a probability of $\frac{1}{2}$.

Once the measurement matrix is obtained the measurement vector y is easily obtained. The next step is to select the appropriate basis matrix ψ. For the current implementation the basis is selected to be the *Discrete Time Transform* basis. This selection was made empirically by observing the fact that most compressive sensing application have used it as well, although it must still be investigated if a systematic basis selection procedure is possible. The basis matrix is constructed using the definition of the DCT transform. The i^{th} transformed element is given as,

$$x_i = w(i) \sum_{j=1}^{n} s_j cos\left(\frac{\pi(2j-1)(i-1)}{2n}\right) \quad i, j = 1, 2, \ldots n \tag{5.12}$$

where, $w(i) = \frac{1}{\sqrt{n}}$ if $i = 1$ and $w(i) = \sqrt{\frac{2}{n}}$ $2 \leq i \leq n$. This can be written in matrix form, so that each entry of the ψ matrix can be given as $\psi_{ij} = w(i)cos\left(\frac{\pi(2j-1)(i-1)}{2n}\right)$.

These steps are illustrated in Figure 5.7. The bottom part of the figure illustrates the A,s and y matrices that will be provided to the L_1 norm minimization algorithm.

The next step is to solve equation 5.7 directly using an L_1 norm minimizer L_1 Magic (Candes and Romberg). Figure 5.8 explains the simulation results carried out on a 128×128 pixel pre recorded image. The original image is shown in the top left corner. The first row illustrates five trajectories that were experimented with. The next row presents the images reconstructed using compressive sensing and the last one shows the images reconstructed using simple linear interpolation.

The interpolation is performed by approximating the unknown values between two given data points using a straight line. While it may be possible to get better interpolation results using other interpolation or possibly function approximation methods for instance splining, regression, weighted regression, kernel based interpolation etc., these are not the subject of this dissertation.

Table 5.1 provides the number of pixels sampled for each scan pattern, along with the mean squared errors (MSE) between the original image and each of the reconstructed images.

It is clear from the results that the compressive sensing has lesser MSE as compared with interpolation.

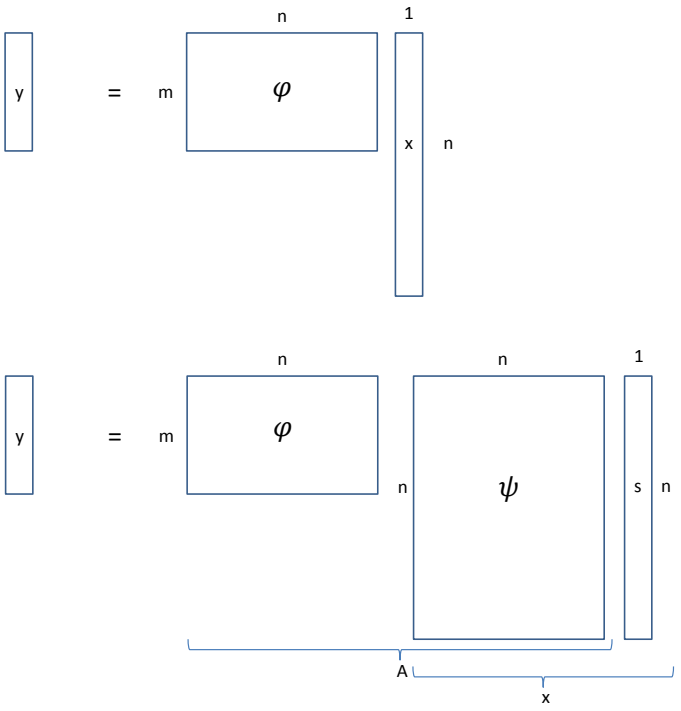

Figure 5.7: Compressive sensing matrix operations.

	Raster	**Random**	**Radial**	**Spiral**	**Lissajos**
Number of Pixels	1277	1639	1764	881	2096
MSE (Compressive Sensing)	52.98	115.63	308.31	156.82	91.28
MSE (Interpolation)	95.44	149.61	485.90	299.67	104.98

Table 5.1: Mean Squared Error between the original and reconstructed images.

5.1.2 Experimental Results

The simulation experiments reported in the previous section were extended to the MultiView 4000 AFM from Nanonics Imaging. The sample used for this purpose is the VGRP-15M calibration grid. This grid have 180nm square pits with a period of $10\mu m$. First a simple 256 line raster scan was performed. The probe starts scanning the sample from the bottom left. Each line upwards is referred to as a *trace* line and each returning line is a *retrace* line. Thus a 256 line scan consists of 128 trace lines and 128 retrace ones. The complete set of measurements during the scan are overlayed on a grid of 256×256 pixels. The complete procedure therefore results in a 256×256 pixel image. Figure 5.9 illustrates the result of this scan. As can be seen, each pit appears with a shadow. The reason for this is that there is a slight shift between the trace and retrace measurements. This is verified by separating the trace and retrace images as illustrated

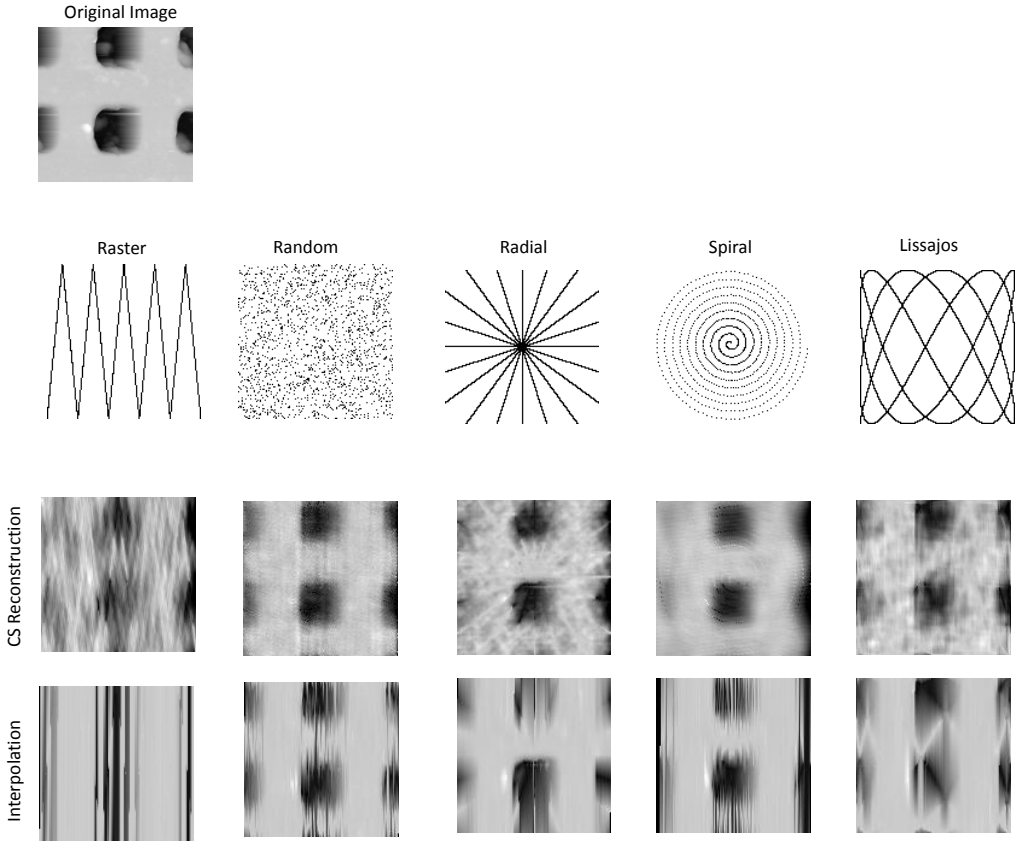

Figure 5.8: Compressive sensing results.

in Figure 5.10. Another phenomenon that can be observed is the fact that the image is brighter in the top left part and darker in the bottom right. This is due to either sample *tilt* or piezo creep, or possibly both acting simultaneously. These scan was completed in 38 minutes.

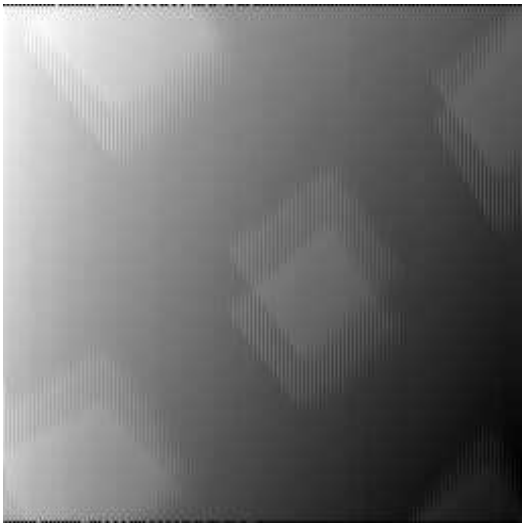

Figure 5.9: VGRP-15M calibration grid 256 line raster scan.

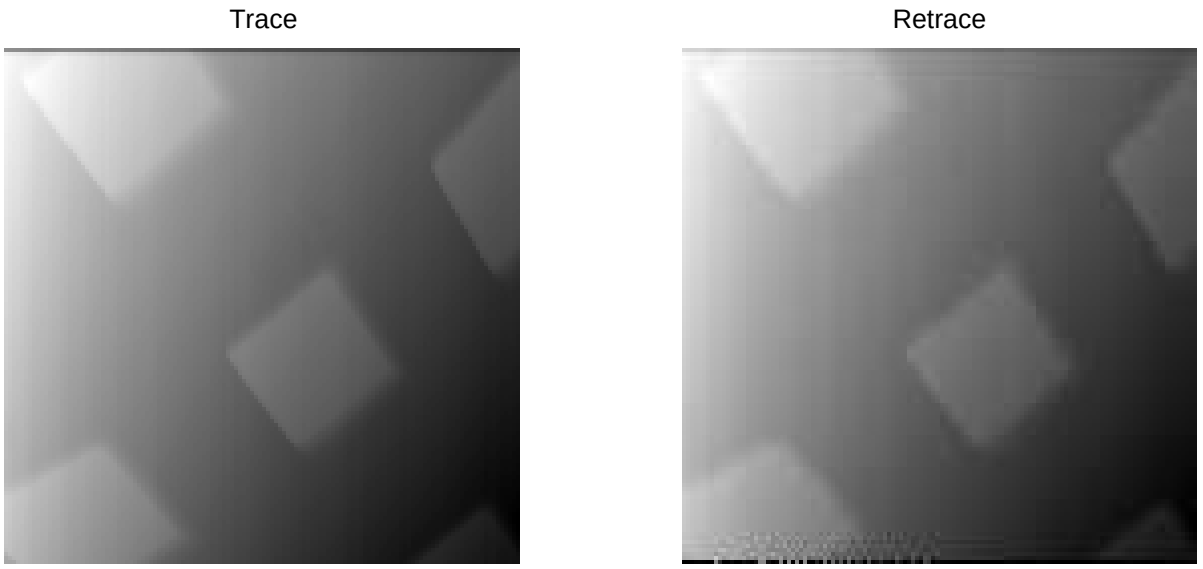

Figure 5.10: VGRP-15M calibration grid trace and retrace images.

Next a spiral scan was performed which took 3 minutes 16 seconds. The results for this scan are provided in Figure 5.11. The top left figure shows the trace image used as the reference. The top right figure shows the spiral scan trajectory. The image reconstructed using compressive sensing is illustrated in the bottom left figure. The last figure demonstrates the result obtained simply by interpolating the data obtained from the spiral scan. As can be seen the compressive sensing method is slightly better than the interpolation result. Using the raster scan as a reference image, the image reconstructed using the compressive sensing has a MSE of 1307.7 whereas the interpolated one has an MSE of 1517.3. Both compressive sensing and interpolation cannot detect all features. A possible reason is that the spiral trajectory does not pass over these features a sufficient number of times.

In addition it must also be noted that auto tuning developed previously is necessary for compressive sensing. This is so because if the trajectory is not in a simple rastering pattern, it becomes difficult for the user to manually tune the controller. In a conventional scenario, the user constantly compares trace and retrace signals and tunes the controller so that the two remain as similar as possible. For the case non-raster trajectories the concept of trace and retrace lines is difficult to implement. Furthermore, in the case of raster scanning the user can see the image as it is being generated line by line. If the controller is inadequate, the user has time to adjust the gains. In case of compressive sensing using non raster trajectories, the image is generated only after the scan is complete.

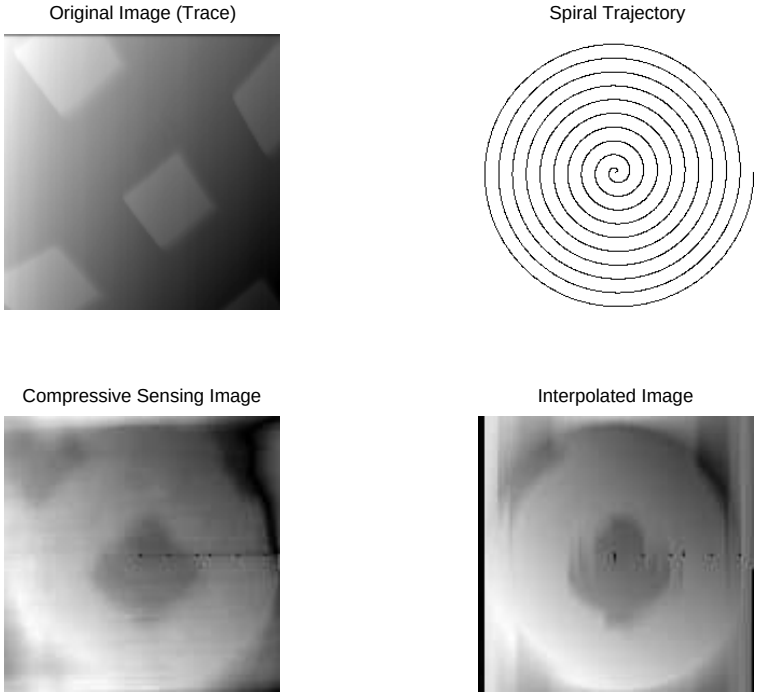

Figure 5.11: Compressive sensing results using MV-4000 for VGRP 15M calibration sample.

5.2 Imaging Using the Entire Cantilever Beam

As mentioned previously the reason for the slow imaging speed in AFMs is that the cantilever probe needs to be moved sequentially over the sample surface. In the last section, it was described how the scan trajectories can be made coarse and that the missing information can be estimated by using compressive sensing. This section provides an alternate method to reduce the imaging time. The principle concept suggested here is that rather than have one tip at the cantilever end and generate sample height measurements point by point, the cantilever can have a large number of tips along its entire length. This will enable simultaneous measurements across the entire beam length. The deflection profile along the entire beam length is then used to estimate the sample profile underneath the beam. The beam is then moved to the next line above the sample, the same process is repeated until the complete sample image is generated. The fact that a large number of parallel measurements can be generated along an entire line results in a substantial reduction in imaging time. This method is referred to as *Full Beam* Atomic Force Microscopy and is explained next.

The fundamental principle of the Full Beam atomic force microscopy as illustrated in Figure 5.12 is that the beam consists of N_T tips distributed uniformly across its length. Contrary to conventional atomic force microscopy where the beam is fixed at one end and free at the other, in this case the beam is pinned at both ends to a fixed support.

The sample is placed on a piezo-actuator at a distance Z_c below the pinned ends of the beam. The piezo-actuator enables both lateral and vertical movement of the sample.

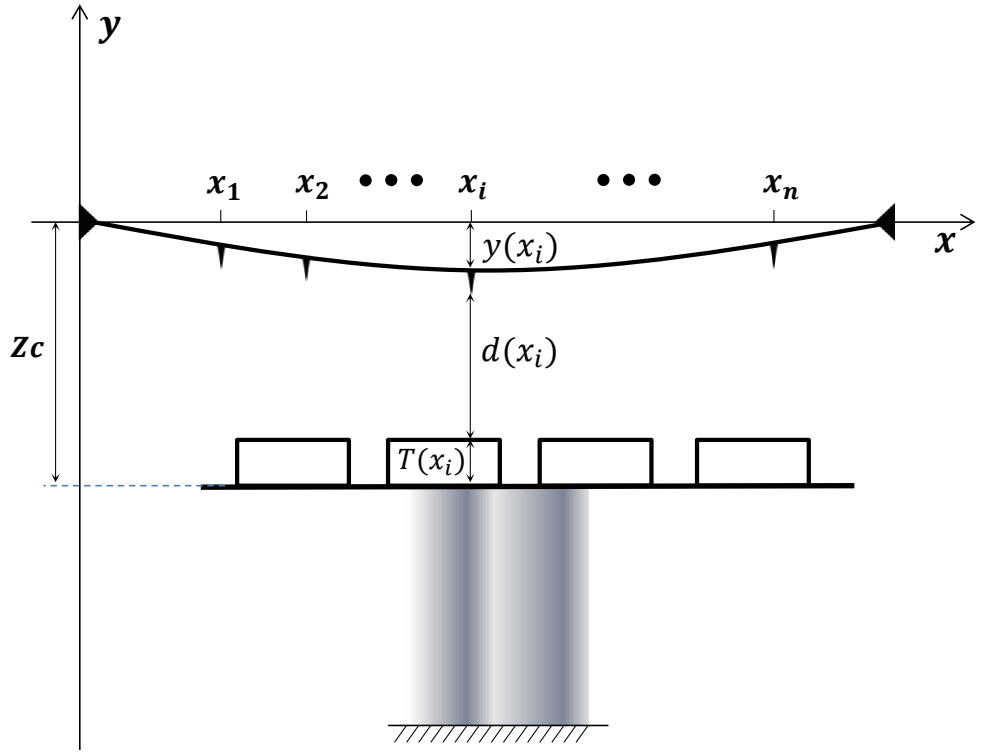

Figure 5.12: Full Beam Atomic Force Microscope setup.

The top end of the beam above each tip is illuminated with N_S laser spots that enables measurement of the deflection angle of the beam $\theta_i = \frac{d\theta}{dx}\big|_{x=x_i}$ for the i^{th} tip. For the purpose of the current contribution $N_T = N_S = n$. The vector of deflection angles $\vec{\theta} = [\theta_1,\ \theta_2,\ \ldots\ \theta_n]^T$ is used to evaluate the vector of forces and moments acting on the beam at the location of each tip $\vec{F} = [f_1,\ \tau_1,\ \ f_2,\ \tau_2,\ \ \ldots\ \ f_n,\ \tau_n]^T$. Here f_i is the tip sample interaction force acting on the i^{th} tip and τ_i is the moment acting on the beam at the location of each tip.

Once the force vector is known the sample topography can be estimated by inverting the tip sample interaction relation. This places two constraints on the full beam AFM set up given below,

$$a_o \leq |d(x_i)| \leq d_{max} \quad \forall\ 1 \leq i \leq n \tag{5.13}$$

$$|T(x_i)| \leq d_{max} \quad \forall\ 1 \leq i \leq n \tag{5.14}$$

Here a_o is the inter atomic separation for the sample material. It is the closest distance the tip can be from the sample surface without penetrating it. d_{max} is the maximum separation between the tip and sample after which the effect of the tip sample interaction force is too small to be measured. The motivation behind placing this constraint is to ensure that the tip sample interaction force is explained solely by the Van der Waals interaction force between tip and sample. The second constraint is a simple extension of the first one. If the sample topography has a variation greater than d_{max} then either some regions of the sample surface will be too far to exert any measurable force on the beam, or the tips will penetrate the sample. In both cases topography reconstruction will become impossible. The complete tip sample interaction force is explained by the following relation,

$$f_i = \begin{cases} -\frac{HR}{6d_i^2} & \text{if } d_i > a_o \\ -\frac{HR}{6a_o^2} + \frac{4}{3}E^*\sqrt{R}(a_o - d_i)^{\frac{3}{2}} & \text{if } d_i \leq a_o. \end{cases} \tag{5.15}$$

where $d_i = d(x_i)$ H is the Hammaker constant, R is the tip radius, E^* is the effective Young's Modulus. Figure 5.13 illustrates one possible tip sample interaction force curve for $E^* = 1.3$ GPa, $R = 10$ nm, $a_o = 0.16$ nm and $H = 7.1 \times 10^{-20}$ J.

As is clear from Equation 5.15 if the tip sample distance is greater than a_o the interaction force is explained by a single term which is the Van der Waals force. If the distance is less than a_o i.e, the tip starts penetrating the surface and the interaction force is explained by an additional term originating from the Derjaguin Muller Toporov (DMT) mode. In this case d_{max} is determined by using the fact that current AFMs can measure a minimum force F_{min} of 1 pico Newtons and then inverting the Van der Waals relation i.e, $d_{max} = \sqrt{\frac{-HR}{6 \times F_{min}}}$. Here the negative sign in the root is not a problem since the Van der Waals force always cause the tip to be attracted towards the surface thus giving F_{min} a negative sign as well. For the current values of H and R, $d_{max} = 15.38$ nm.

Finally the complete topography estimation process can be given as a mapping Φ,

$$\Phi : map\ \theta_i \rightarrow (f_i, y_i) \quad \forall\ 1 \leq i \leq n \tag{5.16}$$

where $y_i = y(x_i)$ is the beam deflection at $x = x_i$. Once the tip forces and beam deflections are known, the topography can be determined by the using the following relation obtained from Figure 5.12,

$$T_i = |y_i - Zc| - |d_i| \quad \forall\ 1 \leq i \leq n \tag{5.17}$$

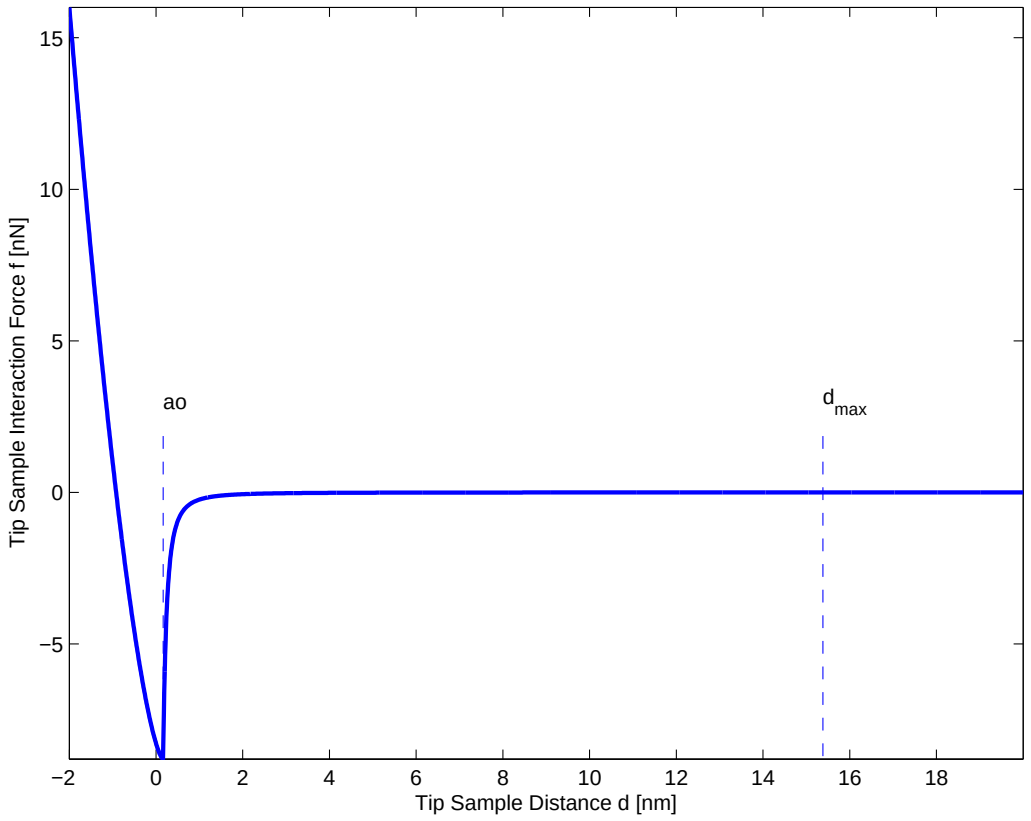

Figure 5.13: Tip sample interaction force.

where $T_i = T(x_i)$ is the topography at $x = x_i$. Using the constraints which ensure that the tip sample interaction force is explained only by the Van der Waals force, this relation becomes,

$$T_i = |y_i - Zc| - \sqrt{\frac{-HR}{6 \times f_i}} \quad \forall\, 1 \leq i \leq n \tag{5.18}$$

The mapping Φ can be obtained by using the Finite Element Model (FEM) for the beam. This is done by using the FEM matrices for the beam namely *Mass*, *Damping* and *Stiffness* matrices M,C and K respectively. A detailed description of how the FEM modelling of a cantilever beam is performed is provided in Appendix B. The complete matrix equation for the systems can then be written as,

$$M\ddot{\vec{X}} + C\dot{\vec{X}} + K\vec{X} = \vec{F} \tag{5.19}$$

where $\vec{X} = [y_1,\ \theta_1,\ \ y_2,\ \theta_2,\ \ \ldots\ \ y_{n+1},\ \theta_{n+1}]^T$, $\vec{F} = [f_1,\ \tau_1,\ \ f_2,\ \tau_2,\ \ \ldots\ \ f_{n+1},\ \tau_{n+1}]^T$ and n is the number of elements in the finite element model. The mass, damping and stiffness matrices can be constructed by using the corresponding mass, stiffness

and damping matrices for each element namely M_i, C_i and K_i and matrix assembly procedure. The element matrices used in this contribution are obtained from Meirovitch (2001) and are given below,

$$M_i = \frac{mL}{420n}\begin{bmatrix} 156 & 22 & 54 & -13 \\ 22 & 4 & 13 & -3 \\ 54 & 13 & 156 & -22 \\ -13 & -3 & -22 & 4 \end{bmatrix} \quad i = 1, 2, \ldots, n \tag{5.20}$$

$$K_i = \frac{EIn^3}{L^3}\begin{bmatrix} 12 & 6 & -12 & 6 \\ 6 & 4 & -6 & 2 \\ -12 & -6 & 12 & -6 \\ 6 & 2 & -6 & 4 \end{bmatrix} \quad i = 1, 2, \ldots, n \tag{5.21}$$

Here, m is the mass per unit length for the beam, E is the Elasticity Modulus, I is the moment of inertia through the transverse axis and L is the beam length. The element matrices are assembled using the assembly procedure mention in Meirovitch (2001) to obtain M and K. Once these are obtained, the stiffness matrix C is determined as a linear combination of M and K using Rayleigh Damping.

Assuming that $\ddot{\vec{X}}$ and $\dot{\vec{X}}$ are negligible we get,

$$K\begin{bmatrix} y_1 \\ \theta_1 \\ y_2 \\ \theta_2 \\ \vdots \\ y_{n+1} \\ \theta_{n+1} \end{bmatrix} = \begin{bmatrix} f_1 \\ \tau_1 \\ f_2 \\ \tau_2 \\ \vdots \\ f_{n+1} \\ \tau_{n+1} \end{bmatrix} \tag{5.22}$$

This assumption is reasonable if the beam is allowed to settle to a steady state before the topography estimation can begin.

The purpose of the mapping Φ is to use the known values of the measured deflection angles $\theta_1\ \theta_2\ \ldots\ \theta_{n+1}$ and determine $f_1\ f_2\ \ldots\ f_{n+1}$ and $x_1\ x_2\ \ldots\ x_{n+1}$. It is already known that while the tip sample interaction exerts forces on the tip but no moments, i.e $\tau_1 = \tau_2 \cdots = \tau_{n+1} = 0$.

The system of equations 5.22 is rearranged using a permutation matrix P as follows,

$$P\begin{bmatrix} y_1 \\ \theta_1 \\ y_2 \\ \theta_2 \\ \vdots \\ \vdots \\ y_{n+1} \\ \theta_{n+1} \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{n+1} \\ \hline \theta_1 \\ \theta_2 \\ \vdots \\ \theta_{n+1} \end{bmatrix} = \vec{\bar{X}} \tag{5.23}$$

and using the same permutation matrix,

$$P\begin{bmatrix} f_1 \\ \tau_1 \\ f_2 \\ \tau_2 \\ \vdots \\ \vdots \\ f_{n+1} \\ \tau_{n+1} \end{bmatrix} = \begin{bmatrix} f_1 \\ f_2 \\ \vdots \\ f_{n+1} \\ \hline \tau_1 \\ \tau_2 \\ \vdots \\ \tau_{n+1} \end{bmatrix} = \vec{\bar{F}} \tag{5.24}$$

This permutation matrix is then used to determine the rearranged system of equations as follows,

$$KP^{-1}P\vec{X} = P^{-1}P\vec{F} \tag{5.25}$$

$$PKP^{-1}\vec{\bar{X}} = \vec{\bar{F}} \tag{5.26}$$

$$\bar{K}\vec{\bar{X}} = \vec{\bar{F}} \tag{5.27}$$

The matrices $\bar{K}$, $\vec{\bar{X}}$ and $\vec{\bar{F}}$ are partitioned as follows,

$$\begin{bmatrix} K_{yy} & K_{y\theta} \\ K_{\theta y} & K_{\theta\theta} \end{bmatrix} \begin{bmatrix} \vec{y} \\ \vec{\theta} \end{bmatrix} = \begin{bmatrix} \vec{f} \\ \vec{0} \end{bmatrix} \tag{5.28}$$

Where $\vec{y} = [y_1,\ y_2,\ \ldots\ y_{n+1}]^T$ and $\vec{\theta} = [\theta_1,\ \theta_2,\ \ldots\ \theta_{n+1}]^T$ and $\vec{f} = [f_1,\ f_2,\ \ldots\ f_{n+1}]^T$. Finally, $\vec{y}$ can be determined as follows,

$$\vec{y} = -K_{\theta y}^{-1}K_{\theta\theta}\vec{\theta} \tag{5.29}$$

Once $\vec{y}$ is known, $\vec{f}$ is determined using Equation 5.28. To conclude the complete image generation using the full beam approach is accomplished in three steps given below,

Step 1: In the first scan line the vector of deflection angles $\vec{\theta} = [\theta_1 \ \theta_2 \ \dots \ \theta_n]^T$ is recorded.

Step 2: Using the mapping Φ mentioned in equation 5.16, the deflection vector $\vec{y} = [y_1 \ y_2 \ \dots \ y_{n+1}]^T$ and the tip force vector $\vec{f} = [f_1, \ f_2, \ \dots \ f_{n+1}]^T$ are generated.

Step 3: The two vectors $\vec{y}$, $\vec{f}$ and the current value of the control signal Z_c are used to estimate the topography T_i beneath each tip using the relation 5.18.

Step 4: The beam is moved over the next scan line and steps 1-3 are repeated until the topography of the entire sample has been estimated.

5.2.1 Full Beam AFM Control Loop

Contrary to the conventional AFM control loop where the AFM plant where both the control and error signals are time dependent scalars, the full beam AFM requires regulating the sample height Z_c as illustrated in Figure 5.12 to keep the vector of deflection angles $\vec{\theta}$ at a set point vector $\vec{\theta}_{sp}$. The error signal is therefore a vector given as,

$$\vec{e} = \vec{\theta}_{sp} - \vec{\theta} \tag{5.30}$$

The set point vector θ_{sp} is selected so that all points along the beam length stay in the desired region between a_o and d_{max}, i.e,

$$a_o \leq y_i \leq d_{max} \quad \forall \ 1 \leq i \leq n \tag{5.31}$$

This done by setting Zc to an arbitrarily chosen value between a_o and d_{max} (in this contribution $Z_c = -10$ nm), letting the surface be flat simulating the beam motion using Equation 5.19. The beam is assumed to be at rest initially and moves under the influence of the tip sample interaction force. Eventually, the beam attains a steady state so that $\ddot{\vec{y}} = \dot{\vec{y}} = \ddot{\vec{\theta}} = \dot{\vec{\theta}} = 0$. The value of $\vec{\theta}$ is used as the set point $\vec{\theta}_{sp}$.

The fact that the error signal is a vector gives rise to two problems. Firstly, it needs to be determined how *far* the vector $\vec{\theta}$ is from the set point vector $\vec{\theta}_{sp}$. Secondly, given a measure of how far $\vec{\theta}$ is from $\vec{\theta}_{sp}$, it must also be determined if the sample needs to be moved towards the beam or away from it.

While the problem of the distance metric can be solved by using the L_2 norm of the error signal vector as done in this contribution, the second problem requires utilization

of the beam geometry. This is illustrated in Figure 5.14. The figure shows the beam deflections y_i, deflection angles θ_i and the corresponding errors e_i when the beam is too close to a flat surface, when it is at a distance where $\vec{\theta} = \vec{\theta_{sp}}$ and when it is too far from the surface. For the purpose of this illustration, the figure was generated by setting $Z_c =$ 8,10 and 12 nm respectively.

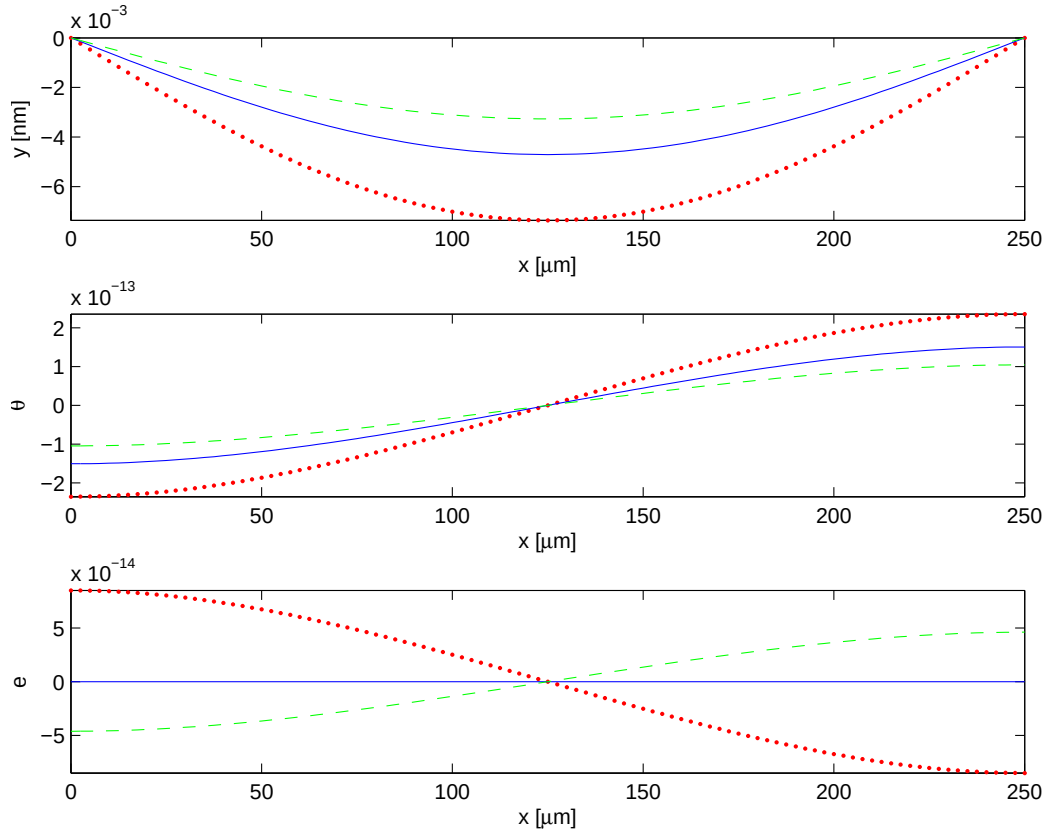

Figure 5.14: Determination of the error metric for three beam positions, near(..), at setpoint (–) and far(- -).

The top plot in Figure 5.14 illustrates the beam deflections for the three cases. As is clear from the plot, the beam shows greater deflection when closer to the surface because the magnitude of the tip sample interaction force is greater. Likewise the beam exhibits lesser deflection when farther away from the surface. The middle plot illustrates the deflection angles. For the first half of the beam before the middle point the deflection angles are all negative since the beam has a negative slope in this region. Likewise the slopes are positive for the second half of the beam. The third plot illustrates the deflection angle error. It is can be seen clearly that if the beam is too close to the surface the error values are positive for the first half of the beam and then negative and vice versa. This metric is given as follows,

$$E = sgn\left(\sum_{i=1}^{0.5(n+1)} e_i - \sum_{i=0.5(n+1)+1}^{n+1} e_i\right)|\vec{e}|_2^2 \tag{5.32}$$

The first term involving the signum function indicates if the surface needs to be moved towards the beam or away from it. The second term involving the L_2 norm is simply used to measure the size of the error. Finally, the controller attempts to minimize $E(t)$ by regulating the sample height Z_c. This contribution reports the results of using a simple Proportional controller along with a first order low pass filter to minimize oscillations in Z_c. These results are reported in the next section.

5.2.2 Simulation Results

This section presents simulation results for the case when the sample is moved laterally at a uniform velocity so that the topography of the entire surface may be estimated. The system is simulated by using the finite element model given in Equation 5.19. This is done by converting the M,C and K matrices into the corresponding state space matrices for the system.

The system is then simulated with a sample time of $T_s = 0.2\mu s$. The finite element model consists of $N_E = N_T = N_S = n = 100$ finite elements. Table I illustrates all the remaining simulation parameters for the sample and the beam.

Since there are a 100 tips on a beam with a length of 250μ m, the distance between tip base centres is 2.5μ m. While it would be a challenge to fabricate 100 tips in the length mentioned, the introduction of high aspect ratio tips for imaging narrow trenches in a sample surface provide a realistic possibility for this end. An example is provided in the form of the AR5-NCHR probe manufactured by Nano World Imaging Technologies. The tip in this case has a height of 2 μ with an aspect ratio of 7:1.

The initial conditions for the beam are set to zero and Z_c is set to -15nm. For the purpose of this contribution $f_x = 0$ and $f_y = 10$ Hz. The sample surface in this case is a rectangular grid with pitch of $100\mu m$ and a height of 5 nm. The image covers an area of $250 \times 500\mu m$. Given the scan rate f_y, 0.1 seconds are required for recording the image.

If conventional AFM imaging is used with 512 trace retrace line pairs and a scan rate $f_x = 10$ Hz, 51.2 seconds would be needed to complete the scan. The full beam scanning in this case is therefore 512 times faster. It is clear that while the speed-up factor may vary for different values of f_x and for a different number of trace retrace line pairs, the full beam method will always be faster than conventional AFM imaging. This is because in conventional AFM a single tip must be moved sequentially over the sample where as in this method a large number of tips are moved over the surface in parallel.

Table 5.2: Simulation Parameters for Full Beam Scan Simulation

Sample Parameters	
Pitch	100 μm
Height	5 nm
Hammaker Constant (H)	7.1 $\times 10^{-20}$ J
Inter-atomic Separation (a_o)	0.16 nm
Young's Modulus (E_s)	1.2 GPa
Poisson's Ratio (ν_s)	0.3
Beam Parameters	
Tip Radius (R)	20 nm
Young's Modulus (E_b)	130 GPa
Poisson's Ratio (ν_b)	0.3
Beam Length (L)	250 μm
Beam Width (w)	35 μm
Beam Thickness (t)	3 μm
Number of Tips (N_T)	100
Number of Spots (N_S)	100
Simulation Parameters	
Sample Time (T_s)	0.2 μs
Number of Beam Elements (N_E)	100

Figure 5.15 illustrates the results of the topography reconstruction. As can be seen that the sample surface is reasonably estimated, although there are distortions near the edges. This is due to the fact that since the sample topography is constantly changing, the assumption that $\ddot{\vec{X}}$ and $\dot{\vec{X}}$ are negligible is no longer true. This causes inaccuracies in estimation of the sample surface. Due to this reason, the topography estimates either exceed d_{max} or are lesser than a_o at some points. Such estimates are discarded from the final result. Despite these errors an acceptable reconstruction is still obtained.The error and controller signals are illustrated in Figure 5.16. The error signal oscillates initially (as shown in inset) and then converges to zero.

5.2.3 Conclusions

This contribution presents a possible solution for reducing imaging times in atomic force microscopes. The fundamental concept is to use the entire length of the cantilever beam rather than only the free end. The advantage is that samples can be scanned faster. There are also implementation challenges that need to be investigated to enable realization of this concept. The first one is the fabrication of a large number of tips on the entire beam length. This would require that each tip have a high aspect ratio and a base length less than a micro meter, as is the case with AFM probe tips from nanoScience Instruments. The current simulation assumes the fabrication of 100 tips on a 250 μm beam. The second implementation issue requiring investigation is the

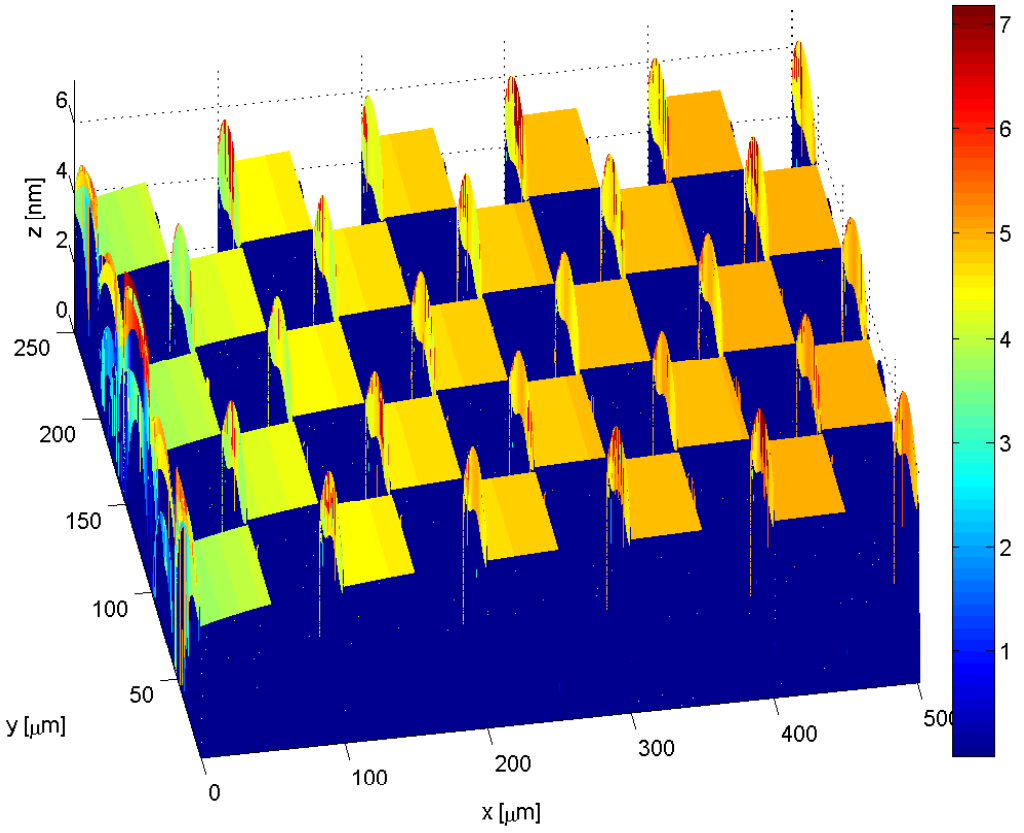

Figure 5.15: Topography reconstruction for two dimensional scan.

placement of a large number of laser spots on the beam. One possible solution is to use self sensing piezo resistive mechanisms. Instead of using a single piezo resistive element, an array could be fabricated on the entire beam length to eliminate the requirement for optical detection.

As a consequence of these limitations, the pitch of the sample imaged in this simulation is chosen be be large i.e, 250 μm. This means that only a limited amount of topography variation can be uniquely determined using the full beam method. If the pitch of the sample is too small, it is no longer possible to map the topography to the beam deflection profile in a one to one manner.

An additional implementation difficulty is the requirement of parallel sampling of large number of channels at a very high sampling rate in excess of 1×10^6 samples per second. This clearly exceeds the capabilities of a large number of low cost off the shelf data acquisition devices and would necessitate the development of analogue circuitry which can obtain the scalar error signal E from the error vector $\vec{e}$.

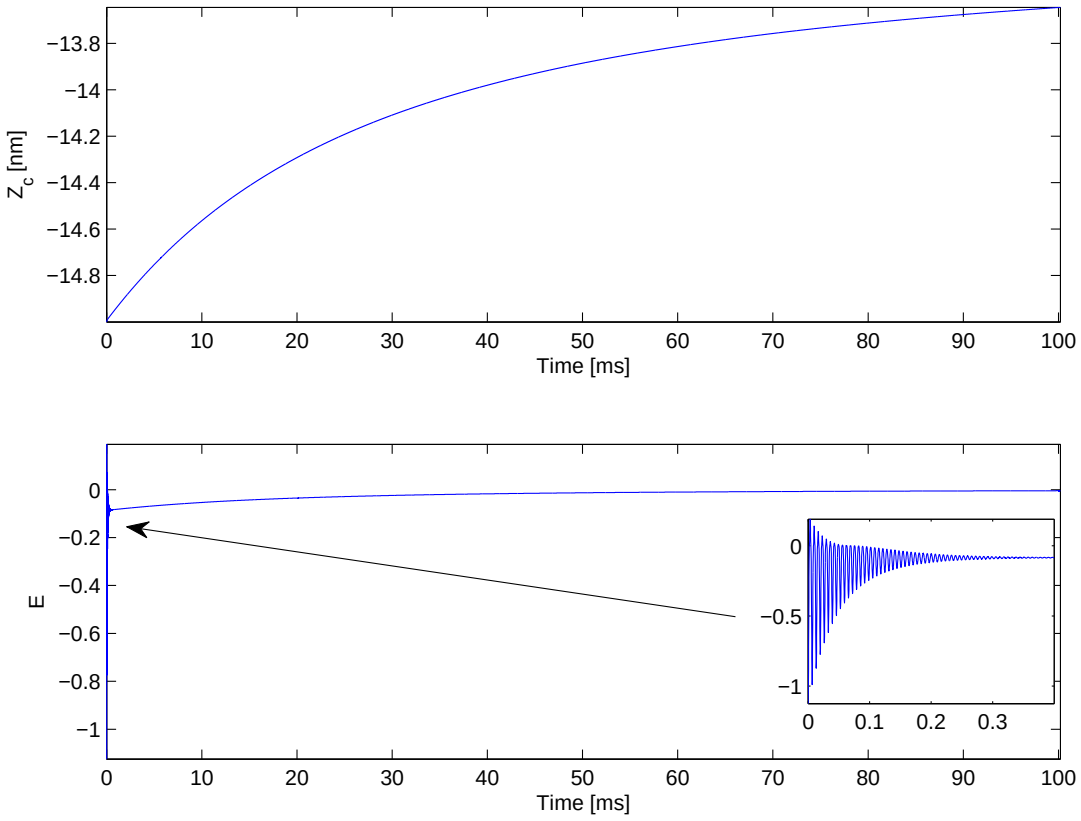

Figure 5.16: Control signal Z_c (top). Error signal E (bottom).

Lastly, the constraints assumed for maximum sample topography variation may appear restrictive, however a number of samples of practical interest exhibit topography variations that fall well within this range. One example is the viewing of biological samples for instance DNA at a high frame rate. This is important since this can lead to a better understanding of biological processes at a molecular scale. Earlier Ando *et.al* observed the motion of a Myosin V protein molecule using a high speed AFM, which improved the understanding of such motion mechanisms. In this context the development of a mechanism that uses the complete length of the cantilever beam for interacting with the surface will lead to even faster imaging and a better study of biological samples.

Chapter 6

Conclusions and Further Work

The work done so far in this project has resulted in the identification of three promising directions of research. These are application of the EMMSAC control algorithm for automated AFM controller generation, usage of compressive sensing and full beam atomic force microscopy for faster imaging. Given below are the contributions in each of these directions and possible future work.

6.1 Contributions of this Research

This research addresses the two key problems associated with AFMs namely usage difficulty due to the necessity of manual tuning and long imaging times. This has led to the following contributions,

• It is demonstrated in this research through simulated and experimental results, that the EMMSAC algorithm can generate the image without any need for manual tuning. This is verified though experimental results. In these results a benchmark calibration grating (TGZ02 from μmasch) is imaged using the automatically tuned controller and manual tuning. As illustrated in the experimental results, both methods generate comparable images.

The ultimate benefit of the entire automated tuning approach is that it can take the tuning burden off the human user. Clearly the image quality can be improved, however the basic objective of making AFMs more accessible has been demonstrated in this instance.

• Two novel methods namely Compressive Sensing and Full Beam Atomic Force Microscopy are investigated for reduction in AFM imaging times. The feasibility of the former is demonstrated though simulation and experimental results. It is shown that the compressive sensing method using a spiral trajectory is more than ten times faster, and the image is better than the one reconstructed using simple interpolation.

The speed up advantage offered by the full beam method is demonstrated through simulation results. It is observed that using this method an entire line can be imaged at once, which otherwise must be imaged point by point sequentially. Due to this reason a speed up factor in excess of five hundred is observed. While such a high speed-up might not be obtained in a physical implementation, it is reasonable to expect that imaging the sample surface using multiple tips in parallel will be much faster than sequentially imaging it using a single tip.

6.2 EMMSAC based Automatic Controller Generation

The fact that EMMSAC provides a promising solution for AFM control is evident from the simulation and experimental results reported in this dissertation. The algorithm however needs to be substantially improved to reduce the computation time and improve the plant controller set to include a greater number of possible AFM experimental setups. In this context the following enhancements can be made,

Reduction in Computation Time: Currently it is observed that given the true AFM plant input output data, a substantial amount of time is needed to evaluate the residual for a single plant. One possibility of reducing the computation time is to use a *Dynamic Plant Set* (Buchstaller (2010)). A dynamic plant set would be one that starts from default group of plants and then gradually discards plants which have a higher residual. Furthermore, new plant models close to those with the lower residuals can be introduced to find plants with even lower residuals. It is expected, that while the computation time may be high initially, it will reduced as the EMMSAC algorithm progresses.

Another advantage of a dynamic plant set is that, it offers the possibility of including a greater number of plants corresponding to a greater number of AFM experimental setups. For instance, the plant set could contain plant models corresponding to all possible combinations of different cantilevers, samples, mediums and operating modes. The plant set could start from a coarse representation of all these possibilities and then gradually discard the irrelevant plants with high residual and generate more plants in the *vicinity* of plants with lower residuals.

The ultimate advantage offered by such a mechanism is that the auto tuning mechanism can be used with a larger number of real experimental set ups. As the irrelevant plants are discarded, the computation time would reduce and consecutive lines can be scanned faster.

Another possible method for reducing the computation time is to reduce the plant set so that only the plants closer to the cluster centre are used. While useful, this mechanism will need an appropriate metric that can measure the distance between the plants.

Robust Control: The problem of the high computation time can also be reduced by using the principle of robust control theory. This means that it might be possible to replace an exhaustive plant set with a large number of plants, with a smaller set of robust plants. Each plant could have a corresponding robust controller of higher order. While the current implementation has been done entirely for PI controllers, there is no restriction in the EMMSAC framework that prohibits the inclusion of more sophisticated controllers.

In addition, as mentioned previously the concept of automated controller synthesis is also needed for image generation methods which require non-raster trajectories.

6.3 Compressive AFM Imaging

The preliminary simulation and experimental results provided previously clearly demonstrate that compressive sensing has the potential for reduction in AFM imaging times. At this point the following directions of research need to be perused,

Determination of a Better Transform: The entire compressive sensing method relies upon the sparseness of the signal that needs to be sampled. While the current research has focused on the DCT transform to obtain a sparse representation of an image, it may not necessarily be the best one. For instance while the DCT is sensitive to the frequency content in a signal, it might not be able to detect local artefacts. Therefore it might be better to explore wavelet transforms which can include information about local artefacts and might provide a better alternate.

Video Generation: Another aspect that needs to be looked into is the usage of compressive sensing for generating videos of physical processes in real time. One area that might possibly benefit from this research is the observation of crystal growth in real time (The Sheffield SPM Group). While compressive sensing has the potential to reduce scan times on any scanning probe microscope, the actual reduction in scan times is dependent on the limitations of the instruments piezo scanners. This research is motivated by the fact that crystal growth process have a time scale in minutes which makes them suitable for viewing using the Nanonics AFMs. Furthermore, improvements in frame rates can help enhance the understanding of crystal growth processes.

Lastly, compressive AFM imaging still relies on a properly tuned controller. Therefore it could be beneficial to investigate integration of automated tuning and compressive imaging.

6.4 Full Beam Atomic Force Microscopy

The concept of full beam atomic force microscopy suggests that instead of using a single point on an cantilever for measurement, the entire beam be used for the purpose. This is motivated by the possibility that a unique force profile acting along the beam length might lead to a unique deflection profile. It needs to be investigated if a one to one mapping between force profiles and image profiles exists, and if so how sensitive is the deflection profile to the variation in sample topography. This idea is similar to the one originally proposed by Kac (1966) in his contribution *Can you hear the shape of a drum?*. It is clearly interesting to investigate if the knowledge of the a cantilever's geometry, material and the deflection profile can reveal the sample topography beneath it. If successful, this method can provide a large speed advantage over the sequential scanning employed currently.

The main challenge that needs to be overcome is the large number of laser spots on the beam. This could be solved by scanning a single laser spot along the length of the beam. Alternatively it might be feasible to investigate the use of *active cantilevers*. These cantilevers are coated with a piezo resistive material which eliminates the need for an optical detection mechanism. Another limitation is the large number of signals that need to be monitored for generating an image. This will increase the cost of the data acquisition hardware needed for image generation. However the potential speed up provided by the method justifies research in this direction.

6.5 *"There is plenty of room at the bottom"*

The title of this section was originally the title of a talk given by Richard Feynman on 29 December 1959, during the annual meeting of the American Physical Society Feynman (1960). While the talk was given more than fifty years ago, the vision is as applicable today as it was when it was first delivered. This talk explains the potential for imaging and manipulation of nano structures. It also alludes to the possibility of instruments that can image and manipulate material at the nano scale. Five decades later, nano science is an established and expanding field. The instruments predicted in the lecture, now exist and have even been referred to as the *"eyes and fingers required for nano structure manipulation and measurement"* National Science and Technology Council (NSTC) (2000).

The motivation behind reproducing the above mentioned statements is to elucidate the fact that the impact of the automated controller synthesis research is by no means limited to a single AFM mode of operation. The same can be extended to other modes that include for instance Multi-frequency and Frequency Modulated AFM. As described in a recent contribution in Nature (Garcia and Herruzo (2012)), in the case of multi-frequency

AFM the cantilever is excited with two sinusoidal frequencies. As mentioned in the same contribution, this enables the imaging of sub-surface structures in a sample for instance in a biological cell. In the second method, namely the frequency modulated AFM the objective is to regulate both the amplitude and frequency of cantilever vibration. This methods has been used for generating atomic resolution images (Gross et al. (2009)). As highlighted by Kilpatrick et al. (2009), gain tuning remains a hurdle in this mode as well. To alleviate the same problem, the authors suggest a gain determination method for optimal performance.

Furthermore it must be noted that Atomic force microscopy is not just an imaging method, it is instead a generic technique that has applications extending well beyond the range of the traditional imaging applications. As an example the concept of AFM imaging has now been extended to observe magnetic dipoles using a method referred to as *Magnetic Force Microscopy*. This method images the magnetic field above a surface using a specially magnetized tip. This has found use in high density data storage applications. One application that makes use of this concept is the Centipede project developed by IBM. This project investigates the use of an array of cantilevers for generating data storage capabilities up to 1Tb/in^2. In addition AFM imaging is not restricted to topography or magnetic dipole measurements but number of other surface properties including surface potential and friction. The study of these has led to the development of *Kelvin Force Microscopy* and *Torsional Force Microscopy*. The first one studies the surface potential by charging the cantilever tip, whereas the second one observes the cantilever's torsional deflection to determine the surface friction properties.

To conclude therefore, the control tools and the speed optimization methods being developed in this research are clearly enabling technologies that will facilitate the evolution and advancement of this branch of science for a substantial period of time in the future.

Appendix A

The Dispersion Relationship

A.1 Dispersion Relationship

$$\left|\begin{bmatrix} \sin\alpha_i + \sinh\alpha_i & \cos\alpha_i + \cosh\alpha_i \\ (\cos\alpha_i + \cosh\alpha_i) - \alpha\bar{m}_{tip}(\sin\alpha_i - \sinh\alpha_i) & -\sin\alpha_i + \sinh\alpha_i - \alpha_i\bar{m}_{tip}(\cos\alpha_i - \cosh\alpha_i) \end{bmatrix}\right| = 0 \tag{A.1}$$

Let ,

$$\begin{aligned} c &= \cos\alpha_i + \cosh\alpha_i \\ \bar{c} &= \cos\alpha - \cosh\alpha_i \\ s &= \sin\alpha_i + \sinh\alpha_i \\ \bar{s} &= \sin\alpha_i - \sinh\alpha_i \end{aligned} \tag{A.2}$$

The matrix thus becomes,

$$\left|\begin{bmatrix} s & c \\ c - \alpha\bar{m}_{tip}\bar{s} & -\bar{s} - \alpha\bar{m}_{tip}\bar{c} \end{bmatrix}\right| = 0 \tag{A.3}$$

The determinant is,

$$\begin{aligned} &= -s(\bar{s} + \alpha\bar{m}_{tip}\bar{c}) - c(c - \alpha\bar{m}_{tip}\bar{c}) \\ &= -(s\bar{s} + c^2) + \alpha\bar{m}_{tip}(c\bar{s} - s\bar{c}) \end{aligned} \tag{A.4}$$

Now,

$$
\begin{aligned}
s\bar{s} &= sin^2\alpha - sinh^2\alpha \\
c^2 &= cos^2\alpha + cosh^2\alpha + 2cos\alpha cosh\alpha \\
s\bar{s} + c^2 &= sin^2\alpha + cos^2\alpha + cosh^2\alpha - sinh^2\alpha + 2cos\alpha cosh\alpha \\
\because cosh^2\alpha - sinh^2\alpha &= 1 \\
s\bar{s} + c^2 &= 1 + 1 + 2cos\alpha cosh\alpha \\
s\bar{s} + c^2 &= 2 + 2cos\alpha cosh\alpha
\end{aligned} \tag{A.5}
$$

(A.6)

Similarly,

$$
\begin{aligned}
s\bar{c} &= (\sin\alpha + \sinh\alpha)(\cos\alpha - \cosh\alpha) \\
s\bar{c} &= \sin\alpha\cos\alpha - \sin\alpha\cosh\alpha + \sinh\alpha\cos\alpha - \sinh\alpha\cosh\alpha \\
c\bar{s} &= (\cos\alpha + \cosh\alpha)(\sin\alpha - \sinh\alpha) \\
c\bar{s} &= \cos\alpha\sin\alpha - \cos\alpha\sinh\alpha + \cosh\alpha\sin\alpha - \cosh\alpha\sinh\alpha \\
c\bar{s} - s\bar{c} &= 2\sin\alpha\cosh\alpha - 2\cos\alpha\sinh\alpha
\end{aligned} \tag{A.7}
$$

Substituting these into A.4,

$$
\begin{aligned}
-(s\bar{s} + c^2) + \alpha\bar{m}_{tip}(c\bar{s} - s\bar{c}) &= 0 \\
-2(1 + \cos\alpha\cosh\alpha) + 2\bar{m}_{tip}\alpha(\sin\alpha\cosh\alpha - \cos\alpha\sinh\alpha) &= 0 \\
1 + \cos\alpha\cosh\alpha + \bar{m}_{tip}\alpha(\cos\alpha\sinh\alpha - \sin\alpha\cosh\alpha) &= 0
\end{aligned} \tag{A.8}
$$

A.2 Equivalent Stiffness k_i

The integral $\int_0^{L_c} \bar{\phi}_i(x)\bar{\phi}_j''''(x)dx$ can be solved using integration by parts by using the identity,

$$
\int udv = uv - \int vdu \tag{A.9}
$$

and letting,

$$
\begin{aligned}
v &= \bar{\phi}_j''' \implies dv = \bar{\phi}_j''''dx \\
u &= \bar{\phi}_i \implies du = \bar{\phi}_i'dx
\end{aligned} \tag{A.10}
$$

we get,

$$\int_0^{L_c} \bar{\phi}_i(x)\bar{\phi}_j''''(x)dx = \bar{\phi}_i(x)\bar{\phi}_j'''(x) - \int_0^{L_c} \bar{\phi}_j'''(x)\bar{\phi}_i'dx \tag{A.11}$$

Integrating the second term on the right by parts again and letting,

$$\begin{aligned} v &= \bar{\phi}_j'' \implies dv = \bar{\phi}_j'''dx \\ u &= \bar{\phi}_i' \implies du = \bar{\phi}_i''dx \end{aligned} \tag{A.12}$$

we get

$$\int_0^{L_c} \bar{\phi}_i'(x)\bar{\phi}_j'''(x)dx = \bar{\phi}_i'(x)\bar{\phi}_j''(x) - \int_0^{L_c} \bar{\phi}_i''(x)\bar{\phi}_j''(x)dx \tag{A.13}$$

Finally ,

$$\int_0^{L_c} \bar{\phi}_i(x)\bar{\phi}_j''''(x)dx = \bar{\phi}_i(x)\bar{\phi}_j'''(x) - \bar{\phi}_i'(x)\bar{\phi}_j''(x) + \int_0^{L_c} \bar{\phi}_i''(x)\bar{\phi}_j''(x)dx \tag{A.14}$$

Substituting the boundary conditions we get,

$$\begin{aligned} \bar{\phi}_i'(x)\bar{\phi}_j''(x)\big|_0^{L_c} &= \bar{\phi}_i'(L_c)\bar{\phi}_j''(L_c) - \bar{\phi}_i'(0)\bar{\phi}_j''(0) = 0 \\ \bar{\phi}_i(x)\bar{\phi}_j'''(x) &= \bar{\phi}_i(L_c)\bar{\phi}_j'''(L_c) - \bar{\phi}_i(0)\bar{\phi}_j'''(0) = -\frac{\omega_j^2 m_{tip}}{EI} \end{aligned} \tag{A.15}$$

Substituting these into A.14 we get,

$$\int_0^{L_c} \bar{\phi}_i(x)\bar{\phi}_j''''(x)dx = -\frac{\omega_j^2 m_{tip}}{EI} + \int_0^{L_c} \bar{\phi}_i''(x)\bar{\phi}_j''(x)dx \tag{A.16}$$

Appendix B

The Finite Element Method

B.1 Approximation of Beam Deflection using a Cubic Polynomial

The Finite Element Method assumes that the beam is split into n elements. Figure B.1 illustrates the j^{th} beam element of length h. Here for the purpose of simplifying the derivations in the later sections the original *global coordinate* x which extends along the length of the beam is replaced with the *local coordinate* ζ (Meirovitch (2001)) . The transformation between x and ζ is given below,

$$\zeta = \frac{(jh - x)}{h} \tag{B.1}$$

In addition,

$$\frac{d}{dx} = \frac{d}{d\zeta}\frac{d\zeta}{x} = -\frac{1}{h}\frac{d}{d\zeta}, \quad dx = -hd\zeta \tag{B.2}$$

As a result $x = (j-1)h$ transforms into $\zeta = 1$ and $x = jh$ transforms into $\zeta = 0$.

Furthermore, Y_{j-1} and θ_{j-1} refer to the beam deflection and rotation respectively at the left end of the beam element, and Y_j, θ_j refer to the same quantities at the other end. The rotations here are defined as the derivative of the displacement with respect to x. Owing to the coordinate transformation mentioned above rotations at any node for instance the j^{th} node can now be written as ,

$$\theta_j = \frac{dY(x)}{dx}|_{x=jh} = -\frac{1}{h}\frac{dY(\zeta)}{d\zeta}|_{\zeta=0} \tag{B.3}$$

The complete vector of nodal displacements a_j can be given as,

$$a_j = \begin{bmatrix} Y_{j-1} \\ h\theta_{j-1} \\ Y_j \\ h\theta_j \end{bmatrix} \tag{B.4}$$

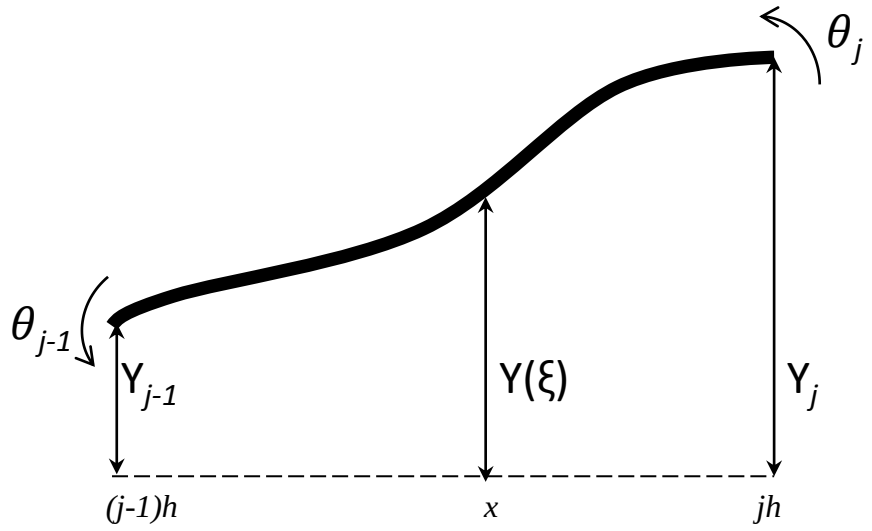

Figure B.1: Finite element for a beam in bending showing the nodal displacements (Meirovitch (2001)).

Approximating the beam deflection Y(ζ) as a cubic polynomial we get,

$$Y(\zeta) = c_1 + c_2\zeta + c_3\zeta^2 + c_4\zeta^3 \tag{B.5}$$

At the right end ζ=0

$$Y(0) = c_1 \tag{B.6}$$

$$\frac{dY(\zeta)}{d\zeta} = c_2 + 2c_3\zeta + 3c_4\zeta^2 \implies -h\theta_j = c_2 \tag{B.7}$$

At the left end ζ=1

$$Y_{j-1} = c_1 + c_2 + c_3 + c_4 \tag{B.8}$$

$$-h\theta_j = c_2 + 2c_3 + 3c_4 \tag{B.9}$$

Solving for c_1,c_2,c_3,c_4 and substituting in B.5

$$Y(\zeta) = (3\zeta^2 - 2\zeta^3)Y_{j-1} + (\zeta^2 - \zeta^3)h\theta_{j-1} + (1 - 3\zeta^2 + 2\zeta^3)Y_j + (-\zeta + 2\zeta^2 - (\zeta^3)h\theta_j \quad \text{(B.10)}$$

$$Y(\zeta) = \phi_1 Y_{j-1} + \phi_2 h\theta_{j-1} + \phi_3 Y_j + \phi_4 h\theta_j \quad \text{(B.11)}$$

Where ϕ_1,ϕ_2,ϕ_3,ϕ_4 are known as *Hermite Cubics*. Writing the hermite cubic functions as a vector ϕ^T, the beam displacement $Y(\zeta)$ can be written as,

$$Y(\zeta) = \phi^T a_j \quad \text{(B.12)}$$

B.2 Potential Energy of a Beam Element

Next the potential energy of the beam element is derived. The total potential energy of all the beam elements can be given as,

$$PE = \frac{1}{2}\sum_{j=1}^{n}\int_{(j-1)h}^{jh} EI(x)\left[\frac{d^2Y(x)}{dx^2}\right]^2 dx \quad \text{(B.13)}$$

Using (B.1) and (B.2) we get,

$$\frac{d^2Y(x)}{dx^2} = \frac{1}{h^2}\frac{d^2Y(\zeta)}{d\zeta^2} \quad \text{(B.14)}$$

Similarly the product $EI(x)$ can be written as,

$$EI(x) = EI[h(j-\zeta)] = EI_j(\zeta) \quad \text{(B.15)}$$

The integral can then be written using (B.2) as,

$$\int_{(j-1)h}^{jh} EI(x)\left[\frac{d^2Y(x)}{dx^2}\right]^2 dx = \int_1^0 EI_j(\zeta)\left(\frac{1}{h^2}\right)^2\left[\frac{d^2Y(\zeta)}{d\zeta^2}\right]^2 (-h)\,d\zeta \quad \text{(B.16)}$$

$$= \frac{1}{h^3}\int_0^1 EI_j(\zeta)a_j^T\frac{d^2\phi(\zeta)}{d\zeta^2}\frac{d^2\phi^T(\zeta)}{d\zeta^2}a_j\,d\zeta$$

$$= a_j^T K_j a_j\ ,\ j = 1, 2, \ldots, n \quad \text{(B.17)}$$

Now substituting the *hermite cubics* developed previously K_j can be written as,

$$K_j = \frac{1}{h^3}\int_0^1 EI_j(\zeta)\frac{d^2\phi(\zeta)}{d\zeta^2}\frac{d^2\phi^T(\zeta)}{d\zeta^2}d\zeta \tag{B.18}$$
$$= \frac{4}{h^3}\int_0^1 EI_j(\zeta)\begin{bmatrix}3(1-2\zeta)\\1-3\zeta\\-3(1-2\zeta)\\2-3\zeta\end{bmatrix}\begin{bmatrix}3(1-2\zeta)\\1-3\zeta\\-3(1-2\zeta)\\2-3\zeta\end{bmatrix}^T d\zeta \quad j=1,2,\ldots,n$$

If E and I_j are constant over all the elements the integration results in,

$$K_j = \frac{EI_jn^3}{L^3}\begin{bmatrix}12&6&-12&6\\6&4&-6&2\\-12&-6&12&-6\\6&2&-6&4\end{bmatrix} \quad j=1,2,\ldots,n \tag{B.19}$$

B.3 Kinetic Energy of a Beam Element

The element mass matrices may be derived using the kinetic energy of the system which can be given as,

$$KE = \frac{1}{2}\sum_{j=1}^{n}\int_{(j-1)h}^{jh} m(x)Y^2(x)dx \tag{B.20}$$
$$= \frac{1}{2}\sum_{j=1}^{n}\int_1^0 m_j(\zeta)Y^2(\zeta)(-h)d\zeta$$
$$= \frac{1}{2}\sum_{j=1}^{n} h\int_0^1 a_j^T\phi(\zeta)\phi^T(\zeta)a_jd\zeta$$
$$= \frac{1}{2}\sum_{j=1}^{n} a_j^TM_ja_j \tag{B.21}$$

Substituting the *hermite cubics* we get,

$$M_j = h\int_0^1 m_j(\zeta)\begin{bmatrix}3(1-2\zeta)\\1-3\zeta\\-3(1-2\zeta)\\2-3\zeta\end{bmatrix}\begin{bmatrix}3(1-2\zeta)\\1-3\zeta\\-3(1-2\zeta)\\2-3\zeta\end{bmatrix}^T d\zeta \quad j=1,2,\ldots,n \tag{B.22}$$

For constant m_j, integration results in the following mass matrix,

$$M_j = \frac{m_j L}{420n} \begin{bmatrix} 156 & 22 & 54 & -13 \\ 22 & 4 & 13 & -3 \\ 54 & 13 & 156 & -22 \\ -13 & -3 & -22 & 4 \end{bmatrix} \quad j = 1, 2, \ldots, n \tag{B.23}$$

B.4 The Nodal Force Vector

The last step of the final element modelling procedure is to replace the continuous force $f(x,t)$ acting on the beam element and replace it with an equivalent vector of forces and moments acting only on the element nodes. The method of doing this is to determine the work done by the external force $f(x,t)$ acting on the element under the assumption that the displacement $y(x,t)$ is small. The displacement is then written in terms of the hermite cubics. This leads to an equivalent force vector that acts on the element nodes. This is illustrated as follows,

$$\delta W = \int_{(j-1)h}^{jh} f(x,t)\delta y(x,t)dx \tag{B.24}$$

$$= h\left[\int_0^1 f_j(\zeta,t)\phi^T(\zeta)d\zeta\right]\delta a_j(t)$$

$$= F_j^T(t)\delta a_j(t) \tag{B.25}$$

where,

$$F_j(t) = h\int_0^1 f_j(\zeta,t)\phi(\zeta)d\zeta = h\int_0^1 f_j(\zeta,t)\begin{bmatrix} 3\zeta^2-\zeta^3 \\ \zeta^2-\zeta^3 \\ 1-3\zeta^2+2\zeta^3 \\ -\zeta+2\zeta^2-\zeta^3 \end{bmatrix} d\zeta \quad j = 1, 2, \ldots, n \tag{B.26}$$

$F_j(t)$ is the j^{th} nodal force vector. The effect of the discretization process thus is to replace the distributed force $f(x,t)$ by the concentrated forces and moments acting on the nodes. The global force vector can be obtained by adding the bottom two components of the j^{th} node to the top two components of the $j+1$ node. Reddy (2006) refers to this vector as the "statically equivalent" forces and moments at the nodes.

B.5 Equations of Motion for a Beam Element

The equations describing the dynamics of the beam element can now be obtained by inserting the kinetic, potential energy and the nodal force vector into Lagrange equations of motion Meirovitch (2001). The final result can be written as,

$$M\ddot{a}(t) + Ka(t) = F(t) \tag{B.27}$$

B.6 Global Matrix Assembly and the Finite Element Model

The solution obtained for the case of single beam element can be extended to n elements by constructing the global mass and stiffness matrices and the global nodal force vector. These global matrices can be constructed by observing that the overlapping nodes between two adjacent beam elements will have the same displacement and rotation. Consider for instance the element $i-1$ and it's adjacent element i. The displacements and rotations of right node of element $i-1$ and the left node of element i will be the same. Therefore once the equations for the two elements need to be combined, the variables corresponding to the displacements and rotations for the overlapping nodes need to be replaced by a single pair of displacement and rotation variables. After this substitution two additional steps need to be carried out. First, the equation terms corresponding to the displacement and rotation of the overlapping node need to be added. Secondly, the forces acting on the overlapping nodes must be added.

These substitutions can be achieved by addition of the bottom right 2×2 elements of the elements of the stiffness and mass matrices for the $i-1$ element with the top left 2×2 elements of the i^{th} element. This procedure is illustrated illustrated in the Figure B.2.

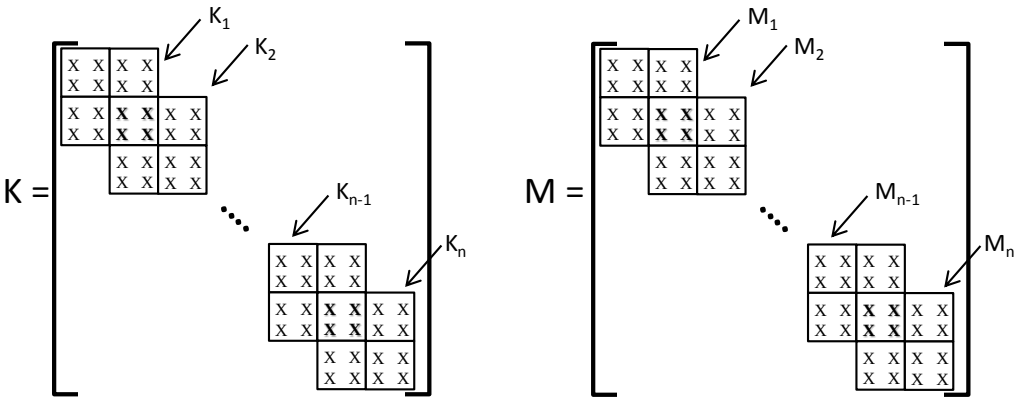

Figure B.2: Global mass and stiffness matrices (Meirovitch (2001)).

The global nodal force vector can also be constructed in a similar manner by adding the bottom two elements of the $i-1$ nodal force vector with the top two elements of the i^{th} nodal force vector.

B.7 Application of Boundary Conditions

Finally the four boundary conditions need to be applied to the finite element model. The first two boundary conditions can be applied by setting the displacement and rotation of the first node to zero. Let these degrees of freedom (DOF) be referred to as the constrained DOFs U_c, and the DOFs for the remaining free nodes be U_f.

The finite element matrices can then be partitioned in the following way (Craig and Kurdila (2006)),

$$\begin{bmatrix} M_{cc} & M_{cf} \\ M_{fc} & M_{ff} \end{bmatrix} \begin{bmatrix} \ddot{U}_c \\ \ddot{U}_f \end{bmatrix} + \begin{bmatrix} K_{cc} & K_{cf} \\ K_{fc} & K_{ff} \end{bmatrix} \begin{bmatrix} U_c \\ U_f \end{bmatrix} = \begin{bmatrix} P_c \\ P_f \end{bmatrix} \tag{B.28}$$

Where $M_{cc}, M_{cf}, M_{fc}, M_{ff}$ and $K_{cc}, K_{cf}, K_{fc}, K_{ff}$ are block matrices of suitable dimensions to permit matrix multiplication.

Setting $U_c = 0$, we get,

$$M_{cf}\ddot{U}_f + K_{aa}U_f = P_c \tag{B.29}$$

$$M_{ff}\ddot{U}_f + K_{ff}U_f = P_f \tag{B.30}$$

To obtain the displacement of the free nodes only the second of the above two equations needs to solved. The first one can be solved if necessary to obtain the reaction forces at constraints. Therefore for the free nodes only the block matrices M_{ff}, K_{ff} and P_f are needed. Alternatively the same result can be achieved by assembling the system matrices as described previously and removing the first two columns and rows respectively.

The remaining two boundary conditions correspond to the shear force and bending moment at the last node on the free end. If these were not equal to zero they would be added to the last two elements of the nodal force vector. Since these are zero no additions need to be made.

B.8 The Damping Matrix

Finally the damping matrix can be included in the system as a linear combination the stiffness and mass matrices. This form of damping is referred to as the *Rayleigh damping* (Craig and Kurdila (2006)). The damping matrix D can thus be given as,

$$C = \alpha_o M + \alpha_1 K \tag{B.31}$$

The constants α_o and α_1 depend upon the material properties of the beam and need to be derived experimentally.

B.9 Simulation Results

This section presents the simulation results obtained for *modal* analysis of the beam. Modal analysis simply means the determination of the eigen-frequencies and eigenfunctions of the beam.

For the present case the force acts perpendicular to the beam surface at the free end and the deflection of the free end is measured as the response. This is illustrated in the Figure B.3

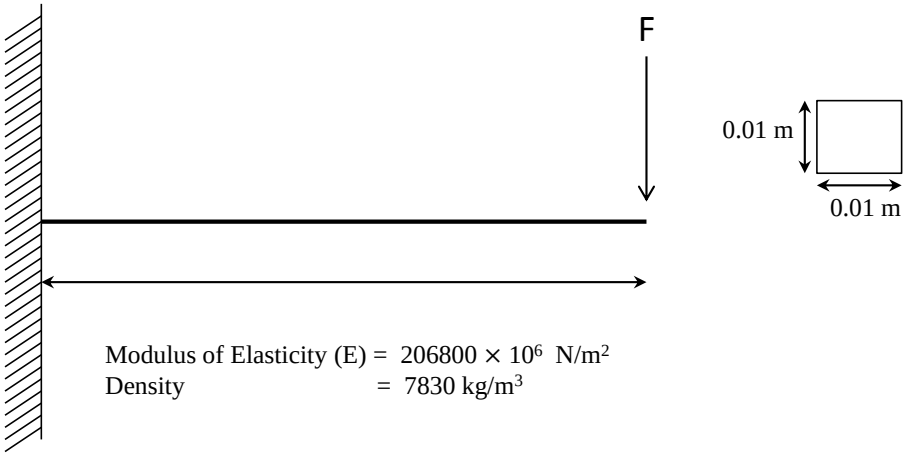

Figure B.3: The cantilever beam.

Two methods are used for the purpose mentioned above. The first one is the *Analytical* method which uses the expressions for the eigenfunctions derived in the second chapter. The second method is Finite Element Modelling. The finite element model for the present method consists of ten elements. The global matrices are generated using the

method described previously and the simulation is carried out using Matlab and ANSYS. The remaining sections provide the simulation results.

Figure B.4 provides the first three eigenmodes obtained using the analytical method. Figure B.5 provides the same modes generated using FEM simulation using Matlab and ANSYS.

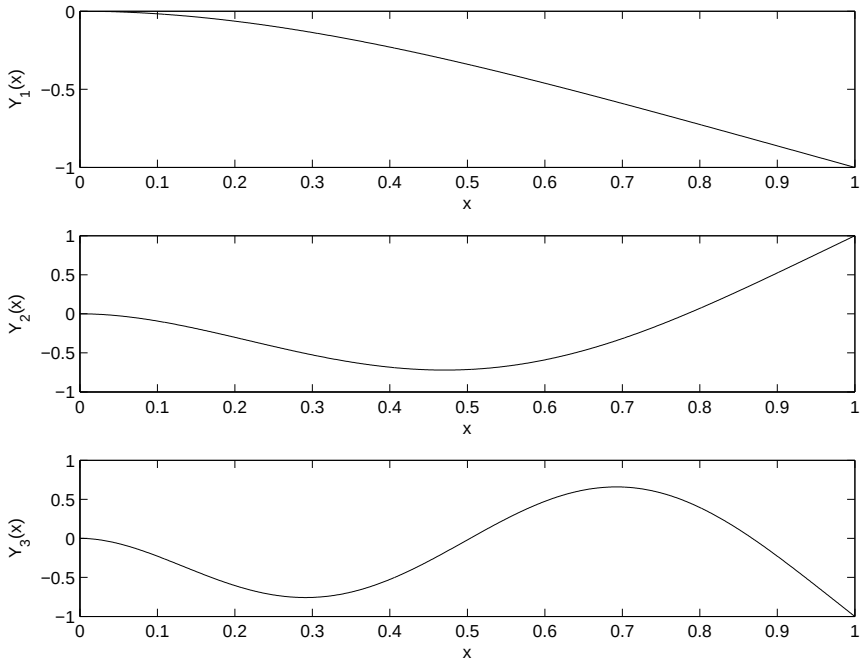

Figure B.4: Natural modes of vibration using analytical method.

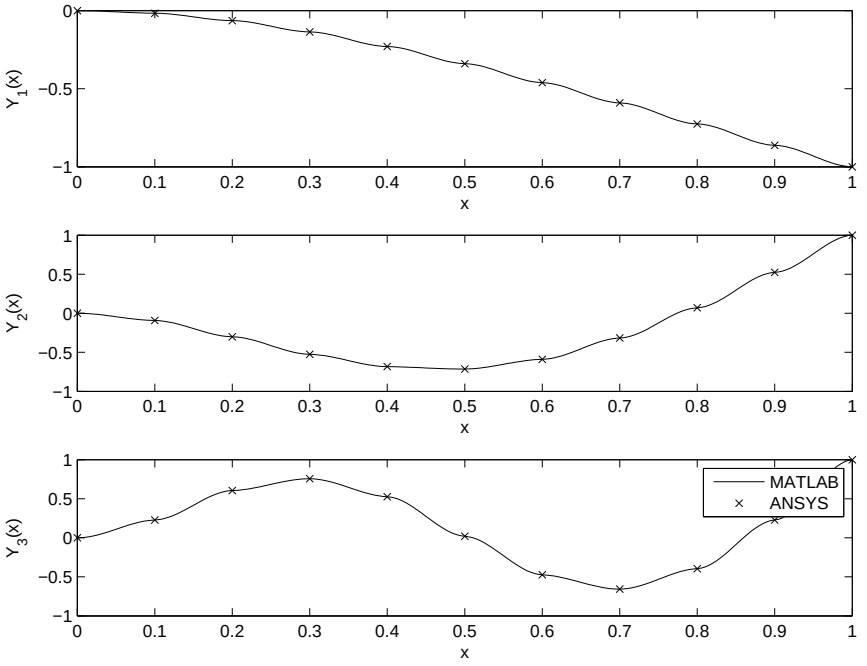

Figure B.5: Natural modes of vibration using FEM.

Table B.1 provides the resonance frequencies for these modes.

Method	ω_1 [Hz]	ω_2[Hz]	ω_3[Hz]
Analytical	8.3018	52.0268	145.6765
FEM(Matlab)	8.3000	52.0110	145.6380
FEM(ANSYS)	8.3019	52.0285	145.7136

Table B.1: Resonance frequencies.

References

D.Y. Abramovitch, S.B. Andersson, L.Y. Pao, and G. Schitter. A tutorial on the mechanisms, dynamics, and control of atomic force microscopes. In *American Control Conference, 2007.*, pages 3488 –3502, July 2007.

P. Agarwal and M. V. Salapaka. Control and systems approaches to atomic force microscopy. In *IFAC 17th World Congress*, pages 10456–10467, 6-11 July 2008.

Agilent. Atomic force microscopy - what is it?. Available Online: http://www.home.agilent.com.

S. B. Anderson. Curve tracking for rapid imaging in afm. In *IEEE Transactions on Nanobioscience*, volume 6, pages 354–361, December 2007.

R.G. Baranuik. Compressive sensing. In *IEEE Signal Processing Magazine*, pages 119–124, July 2007.

G. Binnig, C. Gerber, E. Stoll, T. R. Albrecht, and C. F. Quate. Atomic resolution with atomic force microscope. In *Europhysics Letters*, volume 3, pages 1281–1286, June 1987.

G. Binnig and C. F. Quate. Atomic force microscope. In *Physics Review Letters*, volume 56, pages 930 – 934, March 1986.

D. Buchstaller. *Robust Stability and Performance for Multiple Model Switched Adaptive Control*. PhD thesis, University of Southampton, January 2010.

D. J. Burns. *A System Dynamics Approach to User Independence in High Speed Atomic Force Microscopy*. PhD thesis, Department of Mechanical Engineering, May 2010.

E. Candes and J. Romberg. l_1 magic recovery of sparse signals via convex programming. Available Online http://users.ece.gatech.edu/~justin/l1magic/.

E. Candes, J. Romberg, and T. Tao. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. In *IEEE Transactions on Information Theory*, volume 52, pages 489–509, Feb 2006.

A. Caron, U. Rabe, M. Reinstadtler, J. A. Turner, and W. Arnold. Imaging using lateral bending modes of atomic force microscope cantilevers. In *Applied Physics Letters*, volume 85, pages 6398–6400, December 2004.

G. M. Clayton, S. Tien, K. K. Leang, Q. Zou, and S. Devasia. A review of feedforward control approaches in nano positioning for high speed spm. In *Journal of Dynamic Systems, Measurement and Control*, volume 131, pages 061101–1 061101–19, November 2009.

Cornell University - Department of Animal Science. Mutations. Available Online http://www.ansci.cornell.edu/usdagen/translocation.html.

R. R. Craig and A. J. Kurdila. Fundamentals of structural dynamics, second edition, 2006.

W. J. Dauksher, K. J. Nordquist, N. V. Le, K. A. Gehoski, D. P. Mancini, D. J. Resnick, L. Casoose, R. Bozak, R. White, J. Csuy, and D. Lee. Repair of step and flash imprint templates. In *Journal of Vacuum Science and Technology B: Microelectronics and Nanometer Structures*, volume 22, pages 3306 – 3311, November 2004.

S. Devasia, E. Eleftheriou, and S. O. R. Moheimani. A survey of control issues in nano positioning. In *IEEE Transactions on Control Systems Technology*, volume 15, pages 802–823, September 2007.

O. Dos, S. Ferreiraa, E. Gelinckb, D. de Graafb, and H. Fischer. Adhesion experiments using an afmparameters of influence. In *Applied Surface Science*, volume 257, pages 48 – 55, October 2010.

R. Feynman. There's plenty of room at the bottom an invitation to enter a new field of physics. In *Caltech Engineering and Science*, volume 23, pages 22–36, February 1960.

A. J. Fleming. High-speed vertical positioning for contact-mode atomic force microscopy. In *2009 IEEE/ASME International Conference on Advanced Intelligent Mechatronics*, 14-17 July 2009.

A. J. Fleming and A. G. Wills. Optimal periodic trajectories for band-limited systems. In *IEEE Transactions on Control Systems Technology*, volume 17, pages 552–562, May 2009.

H. Fujimoto and T. Oshima. Contact-mode afm control with modified surface topography learning observer and ptc. In *34th Annual Conference of IEEE Industrial Electronics, 2008. IECON 2008.*, pages 2515–2520, 10-13 November 2008a.

H. Fujimoto and T. Oshima. Nanoscale servo control of contact-mode afm with surface topography learning observer. In *10th IEEE International Workshop on Advanced Motion Control, 2008. AMC '08.*, pages 568–573, 26-28 March 2008b.

R. Garcia and E. T. Herruzo. The emergence of multifrequency force microscopy. In *Nature Nanotechnology*, volume 7, April 2012.

R. Garcia and A.S Paulo. Attractive and repulsive tip sample interaction regimes in tapping-mode atomic force microscopy. In *Physical Review B*, volume 60, pages 4961–4967, August 1999.

M. Grodzicki, S. Smolareka, P. Mazura, S. Zubera, and A. Ciszewski. Characterization of cr/6h-sic(0 0 0 1) nano-contacts by current-sensing afm. In *Applied Surface Science*, volume 256, pages 1014 – 1018, November 2009.

L. Gross, F. Mohn, N. Moll, P. Liljeroth, and G. Meyer. The chemical structure of a molecule resolved by atomic force microscopy. In *Science*, volume 325, pages 1110–1114, August 2009.

M. Guthold, M. Bezanilla, D. A. Erie, B. Jenkins, H. G. Hansama, and C. Bustamante. Following the assembly of rna polymerase-dna complexes in aqueous solutions with the scanning force microscope. In *PNAS*, volume 91, pages 12927 – 12931, December 1994.

J. P. Hespanha, D. Liberzon, and A. S. Morse. Overcoming the limitations of adaptive control by means of logic-based switching. In *System and Control Letters*, volume 49, pages 49–65, 2003.

H. Hoelscher and U. D. Schwarz. Theory of amplitude modulation atomic force microscopy with and without q-control. In *International Journal of Non-Linear Mechanics*, volume 42, pages 608–625, 2007.

J. N. Israelachvili. *Intermolecular and Surface Forces*. Elsevier, 1985.

S. Iyer, R. M. Gaikwad, V. Subba-Rao, C. D. Woodworth, and I. Sokolov. Atomic force microscopy detects differences in the surface brush of normal and cancerous cells. In *Nature Nanotechnology*, volume 4, pages 389 – 393, April 2009.

Y. Jeong, G. R. Jayanth, S. M. Jhiang, and C.H Menq. Direct tip-sample interaction force control for the dynamic mode atomic force microscopy. In *Applied Physics Letters*, volume 88, pages 204102 – 204102–3, May 2006.

M. Kac. Can you hear the shape of a drum? In *American Mathematical Monthly*, volume 73, pages 1–23, April 1966.

S. Kasas, N. H. Thomson, B.L. Smith, H.G. Hansama, C. Bustamante X. Zhu, M. Guthold, E. T. Kool, M. Kashlev, and P. K. Hansama. Escherichi coli rna polymerase activity observed using atomic force microscopy. In *Biochemistry*, volume 36, pages 461 – 468, January 1997.

U. Khan, H. Chong, and M. French. Automated controller tuning for atomic force microscopes using estimation based multiple model switched adaptive control. In *IEEE Conference on Decision and Control*, December 2013a.

U. Khan, H. Chong, and M. French. Full beam atomic force microscopy. In *IEEE International Conference on Automation and Computing*, September 2013b.

U. Khan, H. Chong, and M. French. Real time controller implementation for an atomic force microscope using a digital signal processor. In *IEEE International Conference on Applied Electronics*, September 2013c.

J. I. Kilpatrick, A. Gannepalli, J. P. Cleveland, and S. P. Jarvis. Frequency modulation atomic force microscopy in ambient environments utilizing robust feedback tuning. In *Review of Scientific Instruments*, volume 80, 2009.

N. Kodera, D. Yamamoto, R. Ishikawa, and T. Ando. Video imaging of walking myosin v by high-speed atomic force microscopy. In *Nature*, volume 468, pages 72 – 76, June 2010.

M. Krstic, A. Balogh, and A. Smyshlyaev. Backstepping boundary controllers for tip-force induced flexible beam instabilities arising in afm. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 2430–2435, 13-15 December 2006.

C. Lee. *Control Systems Based Analysis and Design Methods for Scanning Probe Microscopy*. PhD thesis, University of Illinois at Urbana-Champaign, 2010.

C. Lee and S. M. Salapaka. Two degree of freedom control for nanopositioning systems fundamental limitations, control, design and related trade-offs. In *American Control Conference*, pages 1664–1669, 10-12 June 2009.

Z. Li, E. Lee, and F. B. Amara. Performance enhancement in high-speed contact - mode atomic force microscopy. In *IEEE Transactions on Control Systems Technology*, volume 17, pages 1193–1201, September 2009.

J. Ma. Compressed sensing for surface characterization and metrology. In *IEEE Transactions on Instrumentation and Measurement*, volume 59, pages 1600–1614, June 2010.

D. Martnez-Martnez, L. Kolodziejczyk, J.C. Snchez-Lpeza, and A. Fernndez. Tribological carbon-based coatings: An afm and lfm study. In *Surface Science*, volume 603, pages 973 – 979, April 2009.

L. Meirovitch. *Fundamentals of Vibrations*. McGraw Hill, 2001.

J. Melcher, S. Hu, S. Johnson, and A. Raman. Equivalent point-mass models of continuous atomic force microscope probes. In *Applied Physics Letters*, volume 91, pages 053101–1 053101–3, 2007.

J. Melcher, S. Hu, S. Johnson, and A. Raman. Invited article: Veda: A web-based virtual environment for dynamic atomic force force microscopy. In *Review of Scientific Instruments*, volume 79, January 2008a.

J. Melcher, S. Hu, and A. Raman. Invited article: Veda a web-based virtual environment for dynamic atomic force microscopy. In *Review of Scientific Instruments*, volume 79, pages 061301–1 061301–11, 2008b.

J. Melcher, D. Kiracofe, S. Hu, S. Johnson, and A. Raman. Virtual environment for dynamic AFM version 2.0. Available: www.nanohub.org. volume 2.0, April 2011.

S. Misra, H. Dankowicz, and M. R. Paul. Event-driven feedback tracking and control of tapping-mode atomic force microscopy. In *Proceedings of the Royal Society A*, volume 464, pages 2113–2133, August 2008.

Y. Morita, S. Migitaa, W. Mizubayashia, and H. Ota. Afm measurement of atomic-scale si surface etching by active oxidation. In *Surface Science*, volume 604, pages 1432 – 1437, August 2010.

Nanonics Imaging. *Personal Communications*.

National Science and Technology Council (NSTC). National nanotechnology initiative: The initiative and its implementation plan. Available Online: http://www.wtec.org/loyola/nano/IWGN.Implementation.Plan/nni.implementation.plan.pdf. July 2000.

S. Necipoglu, S. A. Cebeci, Y. E. Has, L. Guvenc, and C. Basdogan. A robust repetitive controller for fast afm imaging. In *IEEE Transactions on Nanotechnology*, volume 10, pages 1074–1082, September 2010.

L. M. Picco, L. Bozec, A. Ulcinas, D. J. Engledew, M. Antognozzi, M. A. Horton, and M. J. Miles. Breaking the speed limit with atomic force microscopy. In *Nanotechnology*, volume 18, December 2007.

G. Prakash, S. Hu, A. Raman, , and R. Reifenberger. Theoretical basis of parametric-resonance-based atomic force microscopy. In *Physical Review B*, volume 79, March 2009.

J.N. Reddy. An introduction to the finite element method, third edition, 2006.

O. M. E. Rifai. *Modelling and Control of Undesirable Dynamics in Atomic Force Microscopes*. PhD thesis, Massachusetts Institute of Technology (MIT), 2007.

O. M. E. Rifai and K. Y. Toumi. Adaptive control of uncertain dynamics at the nanoscale. In *Proceedings of the 44^{th} IEEE Conference on Decision and Control*, pages 1180–1184, 12-15 December 2005.

O. M. E. Rifai and K. Y. Toumi. On automating atomic force microscopes: An adaptive control approach. In *Control Engineering Practice*, volume 15, pages 349–361, March 2007.

J. E. Sader, J. W. M. Chon, and P. Mulvaney. Calibration of rectangular atomic force microscope cantilevers. In *Review of Scientific Instruments*, volume 70, pages 3967–3969, October 1999.

G. Schitter, K. J. Astrom, B. DeMartini, and G. E. Fantner. Design and modelling of a high-speed scanner for atomic force microscopy. In *American Control Conference*, pages 502–507, 14 - 16 June 2006.

G. Schitter and M. J. Rost. Scanning probe microscopy at video rate. In *Materials Today*, volume 11, pages 40 – 48, 2008.

A. Sebastian. *Nanotechnology: A Systems and Control Approach*. PhD thesis, Iowa State University, 2004.

H. Seelert, A. Poetsch, N. A. Dencher, A. Engel, H. Stahlberg, and D. J. Mueller. Proton-powered turbine of a plant motor. In *Nature*, volume 405, pages 418–419, May 2000.

T. Shiraishi and H. Fujimoto. Anti-saturation surface topography observer for amplitude modulation atomic force microscope. In *ICROS-SICE International Joint Conference*, pages 23–28, 18-21 August 2009.

B. Song, N. Xi, R. Yang, K. Wai, C. Lai, and C. Qu. Video rate atomic force microscopy (afm) imaging using compressive sensing. In *IEEE Conference on Nanotechnology*, pages 1056–1059, 15-18 August 2011.

R. W. Stark, G. Schitter, M. Stark, R. Guckenberger, and A. Stemmer. State-space model of freely vibrating and surface-coupled cantilever dynamics in atomic force microscopy. In *Physical Review B*, volume 69, pages 085412–1 085412–9, 2004.

D. S. Taubman and M. W. Marcellin. *JPEG 2000: Image Compression Fundamentals, Standards and Practice*. Kluwer International Series in Engineering and Computer Science., 2002.

Texas Instruments. Dual 500kSPS 16 bit 2 + 2 channel simultaneous sampling analog to digital converter ADS 8361. Available Online: http://www.ti.com/lit/ds/symlink/ads8361.pdf. a.

Texas Instruments. High speed FET input operational amplifier OPA 132. Available Online: http://www.ti.com.cn/cn/lit/ds/symlink/opa132.pdf. b.

The Sheffield SPM Group. Available Online: http://spm.group.shef.ac.uk/index.php/multimedia/videos.

R. Vazquez, F. J. R. Sierra, and R. W. Stark. Transfer function analysis of a surface coupled atomic force microscope cantilever system. In *Proceedings of the 2006 American Control Conference*, pages 532–537, 14-16 June 2006.

Veeco. *SPM Instruction Manual*. VeecoInstruments, 2006.

Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: From error visibility to structural similarity. In *IEEE Transactions on Image Processing*, volume 13, April 2004.

Y. Wu, Q. Zou, and C. Su. A current cycle feedback iterative learning control approach for afm imaging. In *IEEE Transactions of Nanotechnology*, volume 8, pages 515–527, July 2009.

K. Yagasaki. New control methodology of microcantilevers in atomic force microscopy. In *Physics Letters A*, volume 375, pages 23–28, November 2010.

K. Yamasue and T. Hikihara. Control of microcantilevers in dynamic force microscopy using time delayed feedback. In *Review of Scientific Instruments*, volume 77, pages 053703 – 053703–6, June 2009.

Q. Yang and S. Jagannathan. Atomic force microscope based nanomanipulation with drift compensation. In *International Journal of Nanotechnology*, volume 3, pages 527 –544, October 2006.

R. G. Yang, N. Xi, K. W. Lai, B. Zhong, C. K. Fung, C. Qu, and D. H. Wang. Nanomechanical analysis of insulinoma cells after glucose and capsaicin stimulation using atomic force microscopy. In *Acta Pharmacologica Sinica*, volume 32, pages 853 – 860, June 2011.

Y. Zhang, Y. Fang, and X. Dong. Output feedback robust adaptive controller design for dynamic atomic force microscopy. In *IEEE International Conference on Control Applications*, pages 1666–1671, 8 -10 September 2010.