UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

Control system implementation on an AFM prototype

Pedro Nuno Brunheta Pinheiro Dias Freixo

Mestrado Integrado em Engenharia Física

Dissertação orientada por:
Mário Manuel Silveira Rodrigues
Miguel Vargas Vitorino

2022

Resumo

O trabalho descrito ao longo do documento insere-se no contexto do desenvolvimento incremental do protótipo de um Microscópio de Força Atómica (AFM). As contribuições aqui expostas visam não só contribuir para o projeto de elementos fundamentais à operação deste AFM, completando-o mas também melhorar alguns aspetos que facilitam a sua utilização.

O AFM é um instrumento extremamente versátil utilizado sobretudo para obter imagens 3D com resolução subnanométrica. No que toca às amostras, é compatível com qualquer tipo de superfície, condutora ou não, e até com amostras mais sensíveis, como por exemplo células vivas. Deste modo, pode ser operado em ar, em vácuo e até mesmo em meio líquido. É uma ferramenta bastante requisitada em projetos de investigação remetentes à área da física biológica. É um dos dispositivos da indústria de semicondutores mais utilizados para realizar controlo de qualidade. De forma geral, é um instrumento chave para a investigação na ciência de materiais, pois variantes do AFM não só permitem a avaliação de propriedades físicas tais como, propriedades mecânicas e eletromagnéticas, para nomear algumas, mas também permite o estudo de fenómenos característicos que ocorrem a esta escala, como por exemplo a condensação capilar [11].

Para além da sua versatilidade, o AFM tem um custo relativamente baixo de operação. Apenas tem um componente que necessita de ser substituído com alguma frequência o *cantilever* e, salvo situações excecionais, não necessita de controlo ambiental, *refilables*, protocolos de limpeza, etc. Contudo, este instrumento acarreta um elevado custo aquisição por parte das instalações que o adquire. O preço de mercado de um instrumento de qualidade aceitável está compreendido entre os $150k - 500k$ euro.

Consequentemente, este projeto insere-se na ambição do AFMaRT do BioISI, FCUL, cuja filosofia visa o desenvolvimento de um AFM com desempenho fidedigno, funcional, obtido a uma fração do preço de aquisição dos modelos disponíveis no mercado. Pretende-se ainda que seja arquitetado e documentado de forma a permitir uma replicação célere e simples e, para além disso, desenvolvido por alunos da Faculdade de Ciências da Universidade de Lisboa, a ser operado consoante as necessidades do laboratório.

Previamente à data este projeto, já tinha sido desenvolvido e construído um protótipo, estando presente no laboratório. Tanto o hardware como o software estão operacionais para a aquisição de imagens e curvas de força. É de notar que o instrumento em causa não fica, de todo, à sombra da qualidade que se pode obter, a largos custos, no mercado e executa as tarefas para as quais está destinado.

Porém, é pouco prático. Para além de que o controlo da distância entre sonda-amostra, quer seja o grosseiro quer seja o fino, não está de todo implementado. A aproximação macroscópica é feita manualmente, o que consome tempo e paciência, e o controlo fino é feito recorrendo a outros instrumentos. Este projeto vem, assim, complementar o protótipo do AFM, tornando-o mais auto suficiente.

Há dois tipos de movimento a executar no controlo da distância entre sonda e amostra. Um fino, com uma gama dinâmica mais pequena, ordem de grandeza do micrómetro, mas com uma resolução subnanométrica e um segundo tipo de movimento mais grosseiro que realiza deslocamentos com milímetros

de amplitude cuja resolução é significativamente mais reduzida. O movimento mais fino é responsável pelo mapeamento da topografia da amostra, enquanto que o grosseiro é responsável por trazer a sonda e a amostra desde uma distância infinitamente longa, quando comparada com a escala que distam quando se está efetivamente a produzir uma imagem, até ao contacto ou quando se precisa de executar o movimento inverso para deixar o instrumento em segurança. Para além desta funcionalidade fundamental, o controlo macroscópico permite ao utilizador realizar pequenos passos de ajuste manual quando, por exemplo, o movimento fino se encontra no limite da sua gama dinâmica.

A arquitetura do sistema de controlo da distância fina entre sonda-amostra do AFM é retroalimentação negativa. Devido à natureza não-linear do instrumento, foi projetado um controlador PI (Proporcional-Integral) analógico, cujos ganhos se ambicionou serem alterados através de uma interface num computador. Para a sua realização é necessário ter um conhecimento alargado do instrumento, de modo a otimizar a *performance* do instrumento. Dito isto, uma parte fulcral do desenvolvimento do sistema de controlo foi a identificação e a modelação dos vários subsistemas que compõem o AFM para se realizar simulações do funcionamento do instrumento e, inclusivamente foram realizadas medidas experimentais da resposta dinâmica da direção vertical do *Scanner* de modo a se representar o sistema de uma forma mais realista.

O AFM em questão pode ser dividido em duas partes: a Cabeça do instrumento composta pela sonda, pelo laser e pelo sistema sensível à posição que mede a interação com a amostra. Por baixo da Cabeça, está a Base. Na Base está presente o *scanner* que, com o auxílio de três elementos piezoelétricos, um por cada eixo de movimento, que aumentam o seu comprimento quando é aplicada uma diferença de tensão aos seus terminais, fazem mover a amostra tanto no plano da mesa onde é colocada a amostra, *X0Y*, como na direção vertical. É através do elemento piezoelétrico da direção vertical que o sistema de controlo adequa a distância entre sonda e amostra durante a formação da imagem. A Base suporta a Cabeça através de três parafusos micrométricos que servem a função de pilar. Um destes pilares está acoplado a um motor DC através de uma correia.

Posteriormente, foi caracterizado experimentalmente este sistema motorizado responsável pela distância macroscópica entre sonda e amostra, de modo a quantificar o tamanho dos passos que o motor dá quando se lhe aplica um sinal durante um determinado período de tempo. Deste modo, estuda-se o comportamento regular deste mecanismo e estabelece-se uma métrica para que o utilizador possa ter um controlo direto sobre a distância sonda-amostra e ainda permite que se implemente um algoritmo altamente fiável que, de forma automática, execute a tarefa de trazer a sonda e a amostra de uma distância arbitrária entre si até que se estabeleça o contacto.

De modo a integrar o controlo da distância assim como implementar uma série de outras funções foi necessário recorrer à programação de um microcontrolador. Através do software desenvolvido será possível realizar tarefas tais como a alteração dos ganhos do controlador digitalmente, executar o algoritmo de aproximação automático, a tarefa de exibir informação quantitativa e qualitativa pertinente acerca dos processos de alinhamento do laser com o sistema sensível à posição, a escolha da magnitude de interação através da especificação do setpoint e a leitura do sinal que está a ser aplicado ao elemento piezoelétrico da direção vertical.

Para a implementação prática deste circuito integrado, recorreu-se à plataforma Arduino que se determinou ser adequado para a execução das tarefas. O Arduino aglomera tanto o suporte físico para o funcionamento do microcontrolador como disponibiliza software e um ambiente intuitivo e simples para a programação do dispositivo, que abstrai o programador de algumas tarefas menos evidentes.

Em relação ao microcontrolador, configurou-se os temporizadores para a execução dos passos do motor, que requer um intervalo de tempo rigoroso. Configurou-se um sinal PWM de 16 bits que, em

conjunto com um filtro passa baixo, serve de fonte de tensão para a definição do setpoint por parte do utilizador. Para além disso, programou-se um algoritmo de Auto-Approach bastante fiável e rápido. Utilizou-se os 6 pins do microcontrolador que têm ligação física a um conversor analógico para digital de modo a se poder recolher informação sobre os sinais mais importante provenientes do AFM, assim como se implementou um protocolo de comunicação com a interface, de modo a que o utilizador possa enviar comandos para serem executados pelo microcontrolador e este possa transmitir a informação recolhida. É também através do microcontrolador que se altera os ganhos do controlador PI. Este comunica com um potenciómetro digital, que altera a resistência dentre 2 terminais conforme desejado visto que é composto por resistências discretas, e ainda com um multiplexador analógico que altera a ordem de grandeza dos ganhos, conforme o canal que fecha o circuito.

De modo a adaptar os sinais provenientes da eletrónica exterior aos valores suportados pelo microcontrolador e vice versa, desenvolveram-se circuitos de interface. Estes circuitos adaptam o nível da tensão que está compreendido entre ± 10 V ou de 0 a 10 V para o intervalo de 0 a 5 V, ou vice versa.

Por último, também se desenvolveu uma GUI intuitiva para que o utilizador possa interagir, no conforto do computador, com o todos os elementos do instrumento descritos previamente e onde é exibida a informação recolhida pelo microcontrolador. A interface permite a visualização dos sinais do sistema sensível à posição e ao sinal de saída do controlador através de "widgets" especializados, permite ao utilizador selecionar o valor do setpoint, também contém botões que executam a aproximação automática e fazem o deslocamento vertical e, por último, permite que sejam alterados os ganhos do controlador PI.

Palavras-chave: Microscopia de Força Atómica, Controlador PI Analógico com ganhos gigitais, Instrumentação, Arduino, Algoritmo Auto-Approach.

Abstract

This work deals with the implementation of a fine and coarse tip-sample distance control as well as with the tuning of several other features that will make one AFM prototype more user friendly. The main goal was to design and integrate a PI (Proportional-Integral) Analog Controller with digitally controllable gains.

The development of the controller started by identifying and characterizing the system, with emphasis on the Z-axis *Scanner's* response, which in turn allowed to build models for all the different components that make up the AFM.

The PI Controller's gains were arranged to be independently tuned via a digital potentiometer in conjunction with an analog multiplexer. The digital potentiometer provides a fine gain adjustment while the analog multiplexer increments the gains by an order of magnitude. These devices receive instructions from a microcontroller.

In parallel, several other important enhancements were carried out, which include an implementation of an Auto-Approach functionality that automatically approaches the probe and sample without crashing onto each other. In order to achieve this, it was conducted an experimental study of the instrument's motorized coarse motion structure.

All the new features developed here were integrated in the existing prototype via the Arduino platform.

To interface the signals outputted by the AFM circuitry and the microcontroller, as well as providing robust tolerance against faulty use, additional circuitry was included. This allows the reading of important signals within the instrument's context, such as the deflection signal, amplitude signal and controller output. By taking advantage of the microcontroller's features, it was designed a voltage source that serves as an adjustable setpoint via the PWM outputs from the Arduino.

Finally, it was design and developed a GUI providing the user direct control of the tasks mentioned above and also displaying some quantitative and qualitative data, acquired by the microcontroller, about the state of the AFM.

Keywords: Analog PI Digital Gains, Atomic Force Microscopy, Arduino, Auto-Approach Algorithm, Interface, Software, Hardware, Instrumentation, Z-Axis control, Z-Scanner, Microcontroller.

Agradecimentos

Não poderia entregar este documento sem deixar umas palavras de agradecimento.

Em primeiro lugar quero agradecer aos meus orientadores, professor Mário Rodrigues e ao Miguel Vitorino por terem aceitado trabalhar comigo e me terem envolvido num meio que cultivou a minha aprendizagem e o meu crescimento pessoal durante o último ano. Ao professor Mário que sempre me prestou conselhos, nunca me negou ajuda e que esteve sempre ao meu lado durante os momentos mais difíceis da tese. Ao Miguel por me ter apoiado a inicializar o projeto quando é tudo desconhecido e tudo aterrador por ser novo.

Quero também agradecer ao meu grupo de amigos. Àqueles que conheci durante a faculdade e aos que me acompanham desde tempos mais jovens. É-me difícil expressar por palavras a minha gratidão pelo companheirismo e os momentos juntos que, no fundo, é aquilo que chamados de vida.

Por último, agradeço do fundo do coração à minha família pelo amor incondicional e por permitirem que tudo isto tenha sido possível.

Table of Contents

List of Figures viii

List of Tables x

1 Introduction 1
- 1.1 Motivation . . . 1
  - 1.1.1 Objectives . . . 1

2 Atomic Force Microscopy 3
- 2.1 AFM's Operation Broad Overview . . . 4
- 2.2 Physical Working principle . . . 5
  - 2.2.1 Intermolecular Forces . . . 5
  - 2.2.2 Probe . . . 7
  - 2.2.3 Signal Detection . . . 10
  - 2.2.4 Piezoelectric Transducers . . . 15
  - 2.2.5 PID Controller . . . 17
  - 2.2.6 Software . . . 18
  - 2.2.7 Amplitude Modulation Mode . . . 19
  - 2.2.8 Lock-in Amplifier . . . 24
- 2.3 Atomic Force Microscope Prototype Setup . . . 27
  - 2.3.1 Characterization of the AFM prototype . . . 29

3 Control System 34
- 3.1 Insights on control . . . 34
  - 3.1.1 Important Concepts of AFM control . . . 39
- 3.2 System identification . . . 40
- 3.3 PI tuning . . . 45
- 3.4 Why Analog? . . . 47

4 Improvements and their Practical Integration into the Instrument 48
- 4.1 Coarse Motion . . . 48
  - 4.1.1 Motor Study . . . 48
  - 4.1.2 Auto-Approach . . . 50
- 4.2 Z-Axis Control Integration via Arduino . . . 52
  - 4.2.1 Arduino Overview . . . 54
  - 4.2.2 Interface Circuitry . . . 60
  - 4.2.3 PI implementation . . . 63

4.3 Software 74
4.3.1 Firmware 75
4.3.2 User Interface Software 78

5 Conclusion **80**

List of Figures

2.1 Van der Walls interaction experimental data 6
2.2 Lennard-Jones Potential 7
2.3 Cantilevers 8
2.4 Schematic representation of the change in the incident plane angle 11
2.5 Cantilever deflection 12
2.6 Optical lever technique 12
2.7 PIN junction 13
2.8 Reverse Bias PIN junction 13
2.9 Transimpedence amplifier 14
2.10 Hysteresis 16
2.11 Illustration of the creep behaviour of piezoelectric materials 16
2.12 PID control system block diagram 18
2.13 Cantilever vibration modes 20
2.14 Damped Oscillatory Motion 22
2.15 Oscillation Regime in Dynamic Mode AFM 23
2.16 Vibration Amplitude Detunning Illustrations 25
2.17 Resonant Curve Shift 25
2.18 Lock-in amplifier block diagram 27
2.19 Mechanical vibration protection 28
2.20 Illustration of the Scanner's working principle 28
2.21 AFM prototype 29
2.22 AFM Head 29
2.23 AFM prototype Scanner 30
2.24 Cantilever holder structure 30
2.25 Schematic of the summing amplifier circuit. Adapted from [22] 31
2.26 Transimpedence amplifier 31
2.27 H-Bridge circuit schematic 32
2.28 ADC and DAC module 33

3.1 Generic Open Loop block diagram 35
3.2 Generic Closed Loop block diagram with a feedback element and a controller 35
3.3 Positive Gain and Phase Margin System 37
3.4 Negative Gain and Phase Margin System 37
3.5 Amplitude plot of different controllers 39
3.6 Phase plot of different controllers 39

3.7 Error signal saturation illustration . . . 40
3.8 AFM block diagram . . . 41
3.9 Scanner amplitude plot . . . 45
3.10 Scanner phase plot . . . 45
3.11 Simulated Step response of the AFM . . . 46
3.12 AFM's loop gain bode diagram . . . 46
3.13 Image of the tuning process . . . 47
3.14 3D Render . . . 47
3.15 Tuning image with alternate scanner . . . 47

4.1 Series of trials with 10 ms time duration step sizes. . . . 49
4.2 Series of trials with 20 and 40 ms time duration step sizes. . . . 49
4.3 10 ms step histogram . . . 50
4.4 40 ms step histogram . . . 50
4.5 Auto-Approach algorithm . . . 52
4.6 Analog to digital converter block diagram . . . 55
4.7 PWM signal . . . 57
4.8 Multiplexer switch . . . 57
4.9 Arduino Uno Rev3 Board Schematic . . . 58
4.10 Arduino Uno Rev3 Board Pinout Schematic . . . 59
4.11 $-10-10$ V to $0-5$ V interface circuit . . . 61
4.12 Real interface circuitry implementation . . . 61
4.13 Schematic of a current source . . . 63
4.14 Schematic of a current sink . . . 63
4.15 PI controller implementation Schematic . . . 64
4.16 Error Amplifier . . . 64
4.17 Schematic of Analog Integrator and Proportional circuits . . . 64
4.18 Integrator Schematic . . . 67
4.19 Proportional Schematic . . . 67
4.20 Illustration of the Digital Potentiometer . . . 67
4.21 Summing Amplifier with Bounded Outputs . . . 69
4.22 PWM cycle . . . 69
4.23 PWM ripple voltage . . . 74
4.24 First order low pass filter transient behavior . . . 74
4.25 Adressing Registers . . . 75
4.26 Arduino's blocking functions . . . 75
4.27 Interface . . . 79

5.1 Final PCB . . . 80
5.2 Interface . . . 80

List of Tables

2.1 Cantilever specification table 10

3.1 Gain and Phase margin relationship 38

Chapter 1

Introduction

1.1 Motivation

The presented work emerged as the result of a shared interest. The author was seeking a dissertation topic within the instrumentation field, and the other counterpart, the laboratory, were contemplating the upgrade of the existent AFM prototype that was depended on third party electronics, whose practical usability was a nuisance.

The AFMaRT laboratory is a research group based in Faculdade de Ciências da Universidade de Lisboa whose topics of interest fall within condensed matter, biological physics and instrumentation.

Reaserch at the AFMaRT laboratory is heavily reliant on the Atomic Force Microscopy core instrument and variants, not only due to it's image formation capabilities and highly sensitive measurement device, but also as a research topic itself. Commercially available Modern Atomic Force Microscopes have huge price points (150k-500k euros) that are only affordable with the aid of massive funds. Besides their incredibly high acquisition cost, the AFMaRTLab has extensive knowledge and experience in the AFM operation and promptly are at the forefront of the technology which leads to the devise and manufacture of their own equipment, suiting specific needs.

At the AFMaRT lab, interested students have actively contributed with ideas and gained expertise. Consequently, the AFMaRT's philosophy has expanded beyond research, into actively engaging in the students' education which, currently, resulted in the development of a fully functional, reliable AFM prototype designed by the young engineers and obtained at a fraction of the market cost.

This is where this dissertation is included. It follows the incremental development of an AFM prototype that was initiated in Arthur Vieira's master thesis [22].

This prototype consists of the fully operational hardware and software components of the instrument and it's performance is on par with the fields standards. However, it partially relies on third party components and electronics and, still requires some usability upgrades as the handling of this prototype may be clunky and time consuming.

1.1.1 Objectives

At the heart of this project are the incremental upgrades that a prototype must be subjected to. Firstly, the primary objective is the implementation of a modified analog PI controller whose gains are digitally controlled via a computer Interface. The implementation not only requires an adequate characterization and modelling of the prototype, for simulation purposes, but also the design of all the circuitry.

Also, there are some much needed improvements to be devised onto the instrument with regards to

the Z-axis open loop control. Namely, there is only available a manual vertical coarse motion where the user would, carefully, turn a micrometric screw in order to distance the probe and sample. This process was very tiresomely and archaic. It was studied the motorized structure coupled to the micrometric screw in order to implement a predicable and regulated step in both vertical directions. Furthermore, it was implemented a much needed Auto-Approach algorithm. This feature automates the probe to sample approach, eliminating the user from this responsibility.

Additionally, an effort was devoted to aid the user with alignment process of the laser spot in the position sensitive detector. The alignment process is made by turning two screws in the instrument while paying attention to the signal that is being outputted. Previous to this work, the visualization was made with the aid of an oscilloscope, which made the process clunky and uncomfortable. The process was revamped by having this information displayed on a computer screen. For this end, it was used an Arduino in order to convert the analog signals into a digital representation. This required the design of an interface circuitry that shifts the voltage level outputted from the position sensitive detector into one that is compatible with the Arduino's maximum ratings. The Arduino doubles down as the responsible element for the mechanism of digital gain regulation of the PI controller. All this circuits were manufactured into a printed circuit board (PCB).

And finally, the last objective was the development of a user interface through which the user, aided by the complementary circuitry, is able to interact with the hardware as well as receive quantitative and qualitative information about certain crucial tasks in the handling of the instrument that is displayed appropriately. Ultimately, this project allows for the user to control these important aspects of the instrument from the comfort of a computer screen.

Chapter 2

Atomic Force Microscopy

The AFM, short for Atomic Force Microscope, is a modern-era instrument that allows for the study of a sample's surface with nanoscopic resolution. From Ancient Greek, Mikrós, standing for small, and Skopéõ, to look at, examine, the Atomic Force *Microscope's* core physical principle does not revolve around light or any type mechanism to guide particles that is usually associated with "traditional" seeing. Its ability to produce an image of the sample does not have any resemblance with the optic/electronic instrument counterparts and it is not limited to thin or conductive samples. The AFM surpasses resolution limit of conventional optic microscopes of approximately half a wavelength - which at the limit of the visible light spectrum, corresponds to about 200 nm - and also it is a far more versatile instrument. In fact, it is used to ascertain several physical properties of a sample such as electric and magnetic polarization, nanomechanical properties such as Young's modulus and hardness as well as measure frictional forces at the nanometer scale.

In the Nanotechnology space, this instrument is usually located in the inspection category meaning that its use is mainly towards probing some desired feature at the nanoscale. Moreover this is a non-destructive process on solid surfaces, however, in practice, given that the underlying physical principle is a repulsive/attractive interaction between the sample and a probe with a very sharp tip, it ultimately falls upon the user's experience and knowledge to minimize large disruptions and impressed dents on the softer samples. Typically, at this scale, forces between surfaces range from hundred's of pico-Newton to micro-Newton but they are applied on very small surface areas, for instance, for a contact radius of 1 nm $P = F/\pi R^2 \approx 0.3$ GPa.

It is a highly sought after instrument in the biological physics field since it can be used in liquid environment and does no harm to cells, making possible experiments with living cells. In the semiconductor industry, it's used as a quality control tool and, of course, is a key instrument for research in materials science since variants of the AFM are used to study unprecedented phenomena that occurs at the nano scale.

Atomic force Microscopy is a 40 year old field and, since it's conception, there have been developed many experimental methodologies leading to new AFM variants. However, there are basic core concepts that are universal in all this family of techniques.

In this chapter it is presented the working principle of the most basic operation of the AFM called *contact* or *static* mode and the more advanced, yet the most prevalent, operating technique is called *dynamic* or *tapping* mode, more specifically Amplitude Modulation mode.

2.1 AFM's Operation Broad Overview

The Atomic Force Microscope's main function is to provide 3 dimensional pictures of the topography of some sample's surface. The AFM relies on the repulsive force between two objects- the sample and a probe - that arises when brought very close to each other. This interaction grows in magnitude as the distance between this objects gets smaller. The probe usually is a small, thin, long and narrow rectangular beam with a very sharp tip (few atoms wide) located at one end of its extremities, while the other end is fixed. The probe is placed with the tip facing the sample. In this configuration, the probe will bend when a force is applied at its free end. In engineering, when a change in a physical property induces a change in another property we call it a transducer. Transducers are very important measuring tools because they allow us to convert between different quantities. In this case, we convert the interaction force, that is applied to both the surface and probe tip, into a change in the vertical position of the tip. The longitudinal dimension of the probe is in the micrometer order of magnitude and the vertical displacement physically cannot surpass this scale and is typically in the tens of nanometers range. In order to measure and monitor the position of the probe position we make use of the optical lever technique [5]. A laser beam is shinned at the top face of the probe's free end, which is often coated with a layer of highly reflective material such as gold. This beam is then reflected into a position sensor photodetector composed of four individual photoactive areas placed in pairs, two on the top and two in the bottom like a two dimension Cartesian system. The reflected beam's spot is centered at the origin of the Cartesian system. Once centered, each individual photoactive quadrant's is being lit with the same amount of light. When there is deviation, the sensor's circuitry produces a voltage proportional to the imbalance of light in the four quadrants. A Force that causes a change in the tip's vertical position leads to a change in the angle of deflection of the laser beam causing the laser spot to move vertically. The photodetector will then generate a voltage in proportion to how much more lit area there is in the upper quadrant pairs relative to the bottom quadrant pair. Hence, this second transducing stage is able to convert the laser's deflection angle into a convenient computer readable signal.

Since the aim is to produce an image of the sample, it is needed to have some iterative process to collect samples and form pixels to show on a screen. The way this process is carried out is through a scanning motion, that is, either the sample or the probe (makes no difference who is moving) is moved in a two dimensional plane $x0y$, the x motion performs a left to right and then right to left sweep before the y direction is incremented. This motion is smoothly provided by piezoelectric elements. Very briefly, these elements alter their physical size when a voltage is applied. At maximum applied voltage, the variation in their length is usually around the tens to hundreds of micrometers. These elements are at the heart of the technique for they allow for very fine, precise and smooth nanometric scale movements.

Lastly, the final consideration to be made is about the control system. At the scanning motion, the tip will fluctuate its position based on the topology of the sample, whenever there is a deep the force on the tip will decrease in magnitude and contact with the surface may vanish completely or, conversely, if there are "hills" present in the sample, the force at the tip will rise considerably and may end up damaging the topography or crash and drag the tip along the scanning motion. In either scenario, the range of the Z axis is limited by the amount the cantilever can bend, which is only a few micrometers. Such a scenario is highly unwanted and is compensated with third piezoelectric element in the z-axis. The ability to control this degree of freedom not only gives the AFM a larger range of motion but allows to set a constant magnitude for the interaction force throughout the hole scan. In other words, the intended interaction force is set up by the user and every time it deviates from this value it is applied a voltage to the piezoelectric element that controls the Z axis such that it lifts or approximates the tip to the sample.

This is is called feedback control. The plot of the voltage applied to the vertical piezoelectric element at each point in the *x0y* plane is the topography of the sample.

What was just described are the most conceptually simple pieces of the puzzle that interact together to produce a topography image. The process described is called contact or static mode at constant force.

It is relevant to mention the other method of imaging very broadly used in the scientific field. As we will see later on, it provides better resolution and noise immunity. It is called dynamic mode, more specifically Amplitude Modulation mode. At the fixed end of the probe it is applied a periodic force to excite the probe near its mechanical resonant frequency. Intuitively, the force exerted on the clamped end is just making the probe change its vertical position at a special rate such that the free end's movement amplitude is at its peak value. As the tip of the probe is vertically oscillating, the laser's spot on the photodetector is mimicking this behaviour which in turn generates an electrical signal with important characteristics - a vibration amplitude, denoted simply as amplitude and a time delay with respect to the applied force at the clamped end, called phase. These are the measurable signal characteristics, which are obtained via another instrument called *Lock-in Amplifier*.

The interaction force changes the amplitude of vibration, thus, in this technique, the amplitude is the observable, hence it's the Control system's input variable, which adjust it's output value in order to keep the Amplitude constant throughout the imaging process.

This is a surface level description of the physical principles that play a role in the Atomic Microscope Imaging process. Up next a more rigorous and fitted description is presented.

2.2 Physical Working principle

In this section a more in-depth explanation on the physical phenomena underlying the AFM working principle is given. The concepts covered are the most relevant to equip the reader with a basic understanding of the fundamental AFM techniques - Contact Mode AFM Imaging and the other covered technique - Amplitude Modulation Mode - that is heavily employed.

2.2.1 Intermolecular Forces

It is hard to pin point a single type of interaction force. There are a variety of intermolecular, surface effects that play a more dominant role at certain distance ranges. These effect are, of course, very much dependent on the vacuum, air, liquid medium between the tip and sample. For instance, under ambient conditions, the surface effects that may arise are adhesion and capillary forces as water spontaneously condenses from vapor in the pores of the surface and between the cavity formed by the tip and the surface [11]. In liquids, the presence of ions contributes with additional forces. In vacuum, the interaction is solely between the atoms and molecules of both bodies.

The van der Walls force is a universal force that occurs in all objects in near proximity. It is the name given to the intermolecular forces that arise from thermal and/or quantum fluctuations in neutral atoms/molecules.

This ubiquitous force is composed of 3 well understood intermolecular interactions: The London Force, the Keesom Force and Debye Force.

The electrons in atoms and molecules have a configuration that minimizes the energy of the system. It turns out that a spacial separation between the positively charged nucleus and the negatively charged electron configuration may occur to a significant extent. A dipole is the name of two charges with opposite polarities separated by some distance. When the measure of the separation of positive and

negative atomic orbitals charges, denominated dipole moment, has net value at or close to zero it is said that the molecule is non-polar, otherwise it is called polar.

Polar molecules are permanent dipoles which interact with other permanent dipoles- Keesom Force. The local negatively charged region attracts the positively charged region of another molecule.

A second interaction that originates from permanent dipoles is induction. A polar molecule is capable of disrupting a non-polar one by repelling its electrons and altering the spacial configuration of it's electron cloud. A dipole is thus induced in the previously non-polar victim. This polarization process ultimately causes an attraction between the two molecules.

The third force listed above is the London force. In the literature it is also referred to as dispersion force. This phenomenon arises due to the random non-zero quantum fluctuations inherently present in the behaviour of the fundamental particles composing all atoms and molecules. A spontaneous change in charge density originates a dipole moment that will, in turn, attract and repel neighbouring dipoles. This instantaneous interaction is the most important in this group because all materials can be subjected to this effect [6].

All of these potentials generate an attractive force which is proportional to $1/r^6$, r being the distance between the dipoles, and scale with the size of the molecule,i.e the number of electrons it has.

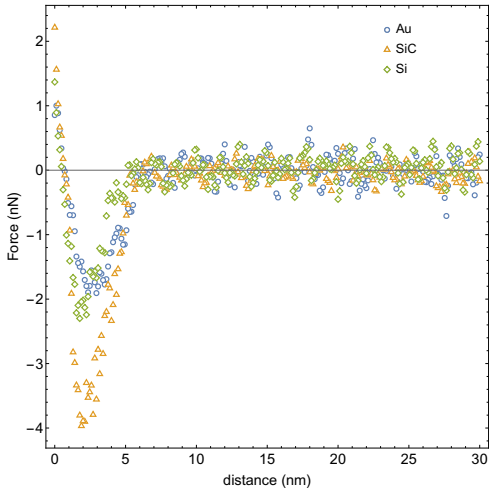

Figure 2.1: Interaction force between a silicon tip and 3 different surfaces: gold, silicon carbide, and silicon. The measurements were obtained in vacuum with a force feedback microscope.

Since sample and probe are a collection of atoms and molecules, the net force is the sum of the contributions made by each individual constituent located on these objects, hence an effective approach is to compute an integration over the idealized geometric model of the macroscopic bodies, such as a plane and a sphere.

At very short distances, the force regime is repulsive. Theoretical models of two charge neutral atoms/ molecules provide a clear picture of the interaction. These forces arise from Pauli's principle, which is the repulsion caused by the overlap of the electron clouds of two atoms, which is normally approximated by the famous Lennard-Jones potential model with a $1/r^{12}$ dependence[6].

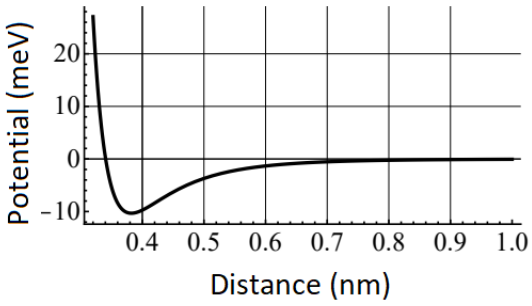

Figure 2.2: Lennard-Jones Potential model based on the interaction of two Argon atoms. Adapted from [22]

When the two bodies are in such a close proximity that the dominant interaction is the repulsion regime they are considered ”in contact” with one another. It’s this operating regime that the AFM takes advantage. Since we deal with a collection of atoms, the best description of the repulsive effects are provided by contact mechanical models. In light of continuum elasticity theories,there are descriptive models that quantize the surface deformation of the two bodies when brought into mechanical contact, references [4] and [16] present an overview of the Hertz model, Derjaguin-Muller-Toporov model, DMT for short, and Johnson Kendall Roberts model (JKR), the conditions for the their validity and the factors that may lead to the preference of one over the other. Nonetheless, they are used to ascertain mechanical properties of the sample’s material.

2.2.2 Probe

Contact models are used to predict the interaction effect on the tip of the probe. In Atomic Force Microscopy, the probe is commonly referred to as *cantilever*. In reality the probe is the beam plus the tip structure. Ultimately, the cantilever bends when in contact and a displacement may be measured.

The main idea, of course, is that it’s held fixed in one end and free to move on the other, *cantilevers* need to provide a measurable deflection with forces on the order of magnitude of nN applied to them and so these tools are fabricated to meet some criteria:

1. relatively low stiffness - spring constants of 10^{-2} to 10^{2} N/m;

2. resonance frequency as high as possible, the low stiffness and it’s physical dimensions set the limit that can be achieved. Typical values in the range of 10^{4} for contact mode operation and 10^{5} Hz order of magnitude for tapping mode. The most modern levers have frequencies above the MHz;

3. sharp tip on one of it’s faces with some curvature radius.

These are typical values. Other solutions exist for situations that require different specifications. Cantilevers are typically made out of a silicon-based material and can be fabricated with many geometries such as a triangular (V-shaped) (2.3), double V-shaped, double X-shaped but the most simple shape it can take is a rectangular one. The probes fixed end is conveniently connect to a much larger ”chip” with a few millimeters in length that greatly aids handling as it is often replaced.

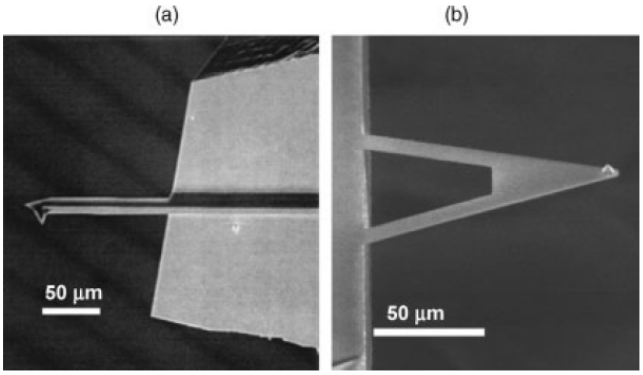

Figure 2.3: Scanning Electron Microscope images from a) rectangular and b) triangular (V-shaped) cantilevers. Taken from [16]

When providing an explanation of this transducer it is always stated that the cantilever is oriented horizontally with respect to the sample's surface to help the reader envision a more simple picture. In reality this is not the case. Usually cantilevers are mounted with certain tilt with respect to the horizontal as it is necessary to ensure that the tip, and not the chip or the chip holder onto which the cantilever is attached, touches the sample first.

Empirically, a force at the tip of the cantilever will make its body bend as it travels a certain vertical distance. Imagine the force is made by our finger placed near end of the beam, when it is raising we feel some resistance as we complete the motion, some pressure in the finger. The cantilever is in fact pushing our finger downwards, in other words, it is exerting a force back at the finger. After releasing the tip, it will move down and up, down and up, again and again.

Systems that exhibit oscillatory movements are very well known in physics.

For small displacements, we may state that the force scales linearly with the displacement from its original position, i.e, the magnitude of the force is proportional to the displacement. This is known as Hook's Law and it is expressed by :

$$\vec{F} = -k\Delta\vec{x} \tag{2.1}$$

where $\vec{F}$ is the restoring force, k is called spring constant and $\Delta\vec{x}$ is the distance between the body's rest position and its current position. $\vec{F}$ is a vector whose it's magnitude is given by the product of the displacement and a constant that depends on the expanding ability of the material, it's direction is always opposite of the displacement and its the reason for the minus sign.

As stated in Newton's second law, the sum of all forces applied to the system is equal to the acceleration experienced by the body : $m\ddot{x} = -kx$. The solution to this differential equation yields a periodic motion described by a sinusoidal function: $x = A\cos(\omega t)$ where A is the amplitude of the motion with respect to its equilibrium position, ω is the angular frequency. Substituting this solution into the equation that describes the system it is obtained:

$$-\omega^2 x + \frac{k}{m}x = 0 \tag{2.2}$$

where we conclude:

$$\omega_0 = \sqrt{\frac{k}{m}} \tag{2.3}$$

This result is very important because it tell us the system's preferred oscillation frequency. It is called natural frequency due to the fact the systems will always tend oscillate at this frequency no matter how hard the cantilever is smacked.

The motion will, however, be determined by the initial position and velocity imposed on the system at the start of the motion.

In the real world, the movement does not go on forever, it stops after some time. This is due to dissipative forces that remove energy from the system, equation 2.1 does not take that into consideration.

A rectangular shaped cantilever is a thin beam clamped at one of its ends and so the force applied to its free end deforms it. A beam is regarded as a long and narrow shape and under small deformation it is elastic, i.e, it recoups its shape. As a result of a force applied perpendicular to the extremities of beam it will bend and come under stress. Visually the beam elongates in one side and compresses on the other. At each cross section of the beam there will be a force pulling at one edge and pushing in the other, whether the edge is in the elongating or compressing side, these are called shear forces. Thus, the beam exhibits some curvature. Curvature is defined as the rate of change in the beam's slope. Supposing displacements occurring are small and that the deformation happens under the elastic regime of the material then the cross section plane remains rectangular, this assumption is valid if the length, L, of the *cantilever* is much longer than its thickness, t. To calculate this curvature we need to sum all the contribution from the internal shear forces along the beam. The derivation of this thought process is the Euler-Bernoulli's beam theory and it results in the following expression that models the height of the cantilever, z along its length x:

$$EI\frac{d^4z(x)}{d^4x} = 0 \tag{2.4}$$

where E is the Young's modulus of the material the cantilever is made of, that quantifies is the amount the given material stretches/compresses when pressure is applied to it and I is the area moment of inertia that expresses the resistance the section has against angular acceleration, this quantity is dependent on the spacial distribution of body's mass. In this particular geometry, $I = wt^3/12$, w , t are, respectively, the width and thickness of the rectangular cantilever.

To solve this differential equation some additional constrains need to be mathematically specified, called boundary conditions. The first constrains have been mentioned throughout the description when we said the the clamped end of the cantilever is immovable. The mathematical descriptions say that the vertical position in the fixed stub is zero, i.e $Z_c(x = 0) = 0$, and, consequently, this point in space cannot physically change position hence, its slope is also null, i.e $dy/dx(x = 0) = 0$. Also, the curvature of the beam at the free end vanishes : $d^2y/d^2x(x = L) = 0$. And finally, we need to specify that the shear forces experienced along the length of the beam are due to the applied perpendicular force : $d^3y/d^3x(x = L) = -F/EI$.

With all of the above considerations, one obtains the following expression for the cantilevers vertical displacement:

$$Z_c(x) = \frac{12F}{Ewt^3}\left(L\frac{x^2}{2} - \frac{x^3}{6}\right) \tag{2.5}$$

And so at the extreme end of the its displacement will be

$$Z_c(x = L) = \frac{4FL^3}{wt^3} \tag{2.6}$$

Comparing equation 2.1 with 2.6 it is established that the cantilever's spring constant is given by :

$$k_c = \frac{Ewt^3}{4L^3} \tag{2.7}$$

Table 2.1: Cantilever specification table

Physical dimensions of a commercially available(*MicroMasch*) Cantilever set with 3 probes in one chip made of silicon. Note cantilevers are complaint with the restrains that the shape of the beam must follow, i.e long thin and not wide. The smaller cantilever is stiff and also the long cantilever is softer

HQ:NSC35					
Cantilever	L (μm)	w (μm)	t (μm)	f_0 (kHz)	k_c (N \m)
A	110	35	2	205	8.9
B	90			300	16
C	130			150	5.4

Note that the physical dimensions of the cantilever and the Young's Modulus of the material set the spring constant. As said previously, cantilevers are usually made out of some silicon blend material. The Young's modulus for Silicon is 150 GPa. General purpose cantilevers are then manufactured to have a spring constant within the specifications presented earlier. As an example, in Table 2.1 it is presented the dimensions and physical properties that commercially available rectangular cantilevers have:

2.2.3 Signal Detection

Given the relationship between forces applied to the cantilever and its deflection, the next natural step is to measure it.

There are few strategies that could be employed such as:

1. Optical Lever - A laser beam is reflected onto a photodetector. A Change in the angle of deflection will offset the reflected beam's position and, consequently, a proportional signal is generated.

2. Piezoresistive detection - Certain materials change their resistivity in reaction to an applied stress. This variable resistor is then placed into a configuration named Wheatstone Bridge. This electrical circuit consists of two pairs of resistors placed in parallel while each pair in configured in series where 3 out of the 4 have known values. A voltage difference is measured in between the parallel pairs. As the cantilever bends the piezoresistive element changes impedance and a change in the measured voltage is picked up.

3. Capacitive detection - Two parallel plates form a capacitor. In this detection method, the inverse relationship between the plates' distance and it's capacitance is leveraged. There are multiple ways of measuring the change in capacitance due to a movement of the cantilever, a simple one is to form a resonating circuit with a resistor, an inductor and the capacitor. By providing an AC signal whose frequency sightly is at the resonant frequency of the circuit, changes in capacitance maybe detected with a change in phase of the output AC signal.

4. Optical Fiber Interferometry - An important property of light is its superposition principle. When two or more waves overlap in space, the resultant disturbance is equal to the algebraic sum of the individual disturbances. Two waves are said to be coherent if they are an exact copy of each other. The interference of two coherent waves results in the point by point sum of the waves amplitudes, as a result of their matched frequency. Only their spacial distribution, i.e, the value of their combined amplitudes point by point, will dictate the resulting pattern. If two waves are exactly symmetrical, their amplitudes will add up to 0, conversely, two waves will interfere constructively and double their amplitude if their space distribution matched perfectly. Following this principle, it can be seen that the interference pattern of two waves repeats every half-wavelength. Thus, it can

be used to measure displacements smaller than half of the light's wavelength. In a practical setup, a laser beam split in two such that one part is guided into a photodetector and the other is guided into the back-side of the cantilever. This is done with the use of optical fiber. At the transition region between fiber and the ambient medium, light is partially reflected back and partially transmitted into said medium. The transmitted beam is then reflected in the back-side of the cantilever and made to interfere with its reflected counterpart inside the optical fiber. It is worth noting that the photodetector produces a signal proportional to the light's Intensity. During the interference process, there will be some intensity losses because the amount of light transmitted and reflected will differ, however this is unimportant because the photodetector's signal will still exhibit the periodic variation but at some offset value. Hence, the cantilever's position can be monitored [23].

By far the most popular method is the optical lever technique (figure 2.6) due to its simplicity and great sensitivity. It will be covered straightway in more detail.

The beam from a laser diode is focused onto the cantilever's free end and reflected into a position sensitive detector. Usually, the back-side of the cantilever is coated with a layer of a highly reflective material to aid this process, aluminium being the most common.

It is important to note that the optical lever technique is sensible to the change in the angle of deflection of the cantilever, i.e it's inclination. To better see this it is useful to recall the reflection law: the angle of the reflected ray, θ_r, is equal to the angle of the incident ray, θ_i, measured from a line perpendicular to the plane of incidence. If the plane of incidence is turned some angle θ_{plane} its new normal(perpendicular) line will also have changed by the same θ_{plane} quantity. Since the position of the incident ray has not changed, the new angle of of incidence measured from the new normal line will be $\theta_i' = \theta_i + \theta_{plane}$. Intuitively, because the incident angle changes we can assume the reflected angle changes as well. To calculate the new angle of reflection we turn again to reflection law's definition and Write $\theta_r' = \theta_i$.

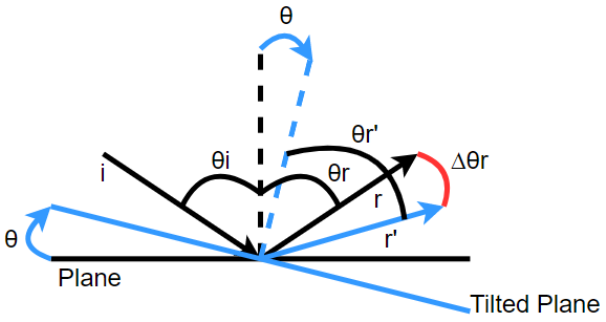

Figure 2.4: Schematic representation of the change in the incident plane angle

The point of this is to relate the change in the angle of the plane, since the plane is the cantilever, with the reflected angle so it is pertinent to measure the deviation from θ_r to θ_r'. From figure 2.4 the angle between the new normal and the reflected ray r is $\theta_i - \theta_{plane}$. To conclude, the final step is to figure out the angle between r and r' relative to the new normal which is done by subtracting the angle of r to r': $\Delta\theta_r = \theta_i + \theta_{plane} - (\theta_i - \theta_{plane})$, resulting in :

$$\Delta\theta_r = 2\theta_{plane} \tag{2.8}$$

However, now there is a need to establish a relationship between the cantilever's bending angle and vertical deflection.

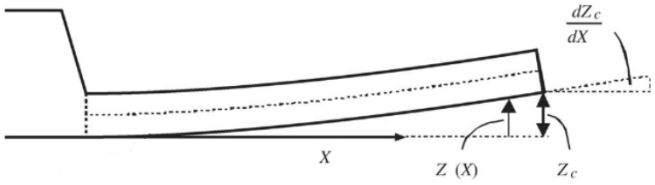

Figure 2.5: Diagram representing the deflection of the Cantilever. Adapted from [4]

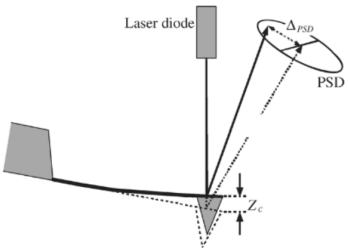

Figure 2.6: Scheme of the Optical Lever Position Detection System. Adapted from [4]

Equation 2.5 describes the height of any point along the cantilever's length.

Mathematically, the angle of the tip in relation with the horizontal is the slope of the cantilever's height:

$$tan(\theta_{z_c}) \approx \theta_{z_c} = \frac{dZ_c}{dx} \tag{2.9}$$

If assumed the deflection occurs within small angles. Calculating the derivative of 2.5 with respect to the length's direction at L, yields the following:

$$\frac{dZ_c}{dx}\Big|_{x=L} = \frac{6FL^2}{Ewt^3} \tag{2.10}$$

Comparing equation 2.10 with equation 2.6, it can be seen that the inclination is proportional to the deflection:

$$\frac{dZ_c}{dx}\Big|_{x=L} = \frac{3}{2}LZ_c \tag{2.11}$$

Consequently, so is the signal provided by the position sensitive detector (PSD) as we will see right away.

The light reflected off the cantilever's back is directed to the PSD.

The most common, over the counter, instruments for light detection available for purchase are photodiodes.

Photodiodes are made of a 3 layers of semiconductive material called a PIN junction. A semiconductor is a material that, unlike a metal, does not conduct electricity at zero kelvin degrees because it does not have free electrons. In bulk materials, electrons occupy fixed levels of energy and there is a limit on number of electrons that can occupy some level. The more electrons present in the material the higher energy state some of it's electrons will occupy. In semiconductors, the last filled energy states occupied is what is called the valence band and the empty energy level immediately above define what is called is called the conduction band. The valence and conduction band are separated by an energy gap. The larger the energy gap the more energy need to be ceded to the materials in order promote electrons to the conduction band and make it conductive. Some materials are insulators because this separation is large, even at room temperatures. Any process that provides enough energy to excite the electron from the valence to the conduction band produces two charge carriers for electrical conduction - the negative charge electron and the positive charge hole. Semiconductors made of pure elements have *Intrinsic* properties.

It is standard practice to combine pure semiconducting elements with other compounds whose charges become weakly bonded to the atoms in the semiconductor material. These impurities, as are called, have the effect of providing low cost electron by promoting it to the conduction band or easily taking the electron away from the semiconductor element and generating a hole in the valence band.

These doped materials are called N-type or P-type respectfully and have drastically different properties. The photodiode is a stack of P-type material followed by a Intrinsic type material and N- type material (PIN) as seen in figure 2.7.

Keep in mind that these are not actually chunks of material stacked on top of each other, the impurities are diffused to form this structure.

In equilibrium, N-type material has higher concentration of loosely bonded electrons donated by the impurities near its conduction band whereas the P-type material has a higher concentration of holes near the valence band and the intrinsic type material has its population of electrons in energy states up to the valence band because it is not doped with impurities. In the intrinsic layer there is a lack of free charged carriers and for this reason it is called depletion region.

Due to this imbalance in the concentrations of electrons and holes in the conduction band a naturally forming potential gradient along its length is formed, as can be seen in figure 2.7.

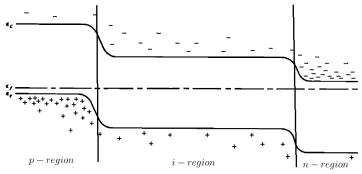

Figure 2.7: Illustration of a PIN junction with naturally formed depletion region. Adapted from [17]

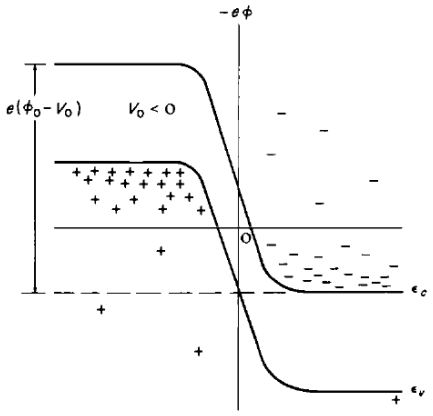

Figure 2.8: Illustration of PIN junction with an applied reserve bias. The potential barrier is increased along the Intrinsic region. Adapted from [17]

The resultant electric field effectively bars the electrons from the N side to cross over to the P-side and vice versa with the holes, yet there are still some electrons/holes that may build up enough energy to successfully cross over and recombine with their counterparts.

If a voltage is applied between the N-type and P-type terminals in reverse bias configuration, i.e, lower the P-type side potential with respect to the N-type side, the potential ”hill” is amplified. This reduces the amount of electrons in the N-type that have enough energy to surmount the potential barrier over to the P-type material and recombine with the holes on that side. However, there are still some holes on the N-side and electrons on the P-side being diffused across the structure with the aid of the potential barrier. These electron-hole pairs are generated from thermal energy in the system. In this conditions, there will be an inherent small net flow of current. This saturation current, which is the name for this phenomenon, is characterized by:

$$I_{sat} = I_{satmax}(e^{\frac{eV_{bias}}{k_B T}} - 1) \tag{2.12}$$

where e is the electron's charge, V_{bias} is the voltage applied between the P-type and N-type substrate, k_B is the Boltzmann's constant and T the temperate.

Light detection is done whenever a photon with high enough energy to promote an electron to its conduction band is absorbed in the intrinsic type material. The generated electron-hole pair is immediately collected to the side where their conduction band concentration is lower, i.e electron to the P-type

material and hole to the N-Type material. This photo-generated current is in the same direction as the saturation current and so the net current in the PIN junction is a sum of both currents:$I_{total} = I_{sat} + I_{ph}$. The photo-generated current can be expressed phenomenologically by a probability process, where it is taken into account how many photons reach the active area, the reflection coefficient for the material , the spectral absorption coefficient for the intrinsic-type layer. In practice, the manufacturer may provide a condensed expression to quantify the current:

$$I_{ph} = P_l S_{max} S_{re}(\lambda) \tag{2.13}$$

where P_l is the incident laser's power on the active area, S_{max} the maximum efficiency at a given wavelength of the the intrinsic layer and $S_{rel}(\lambda)$ is the intrinsic layer's quantum efficiency at the laser's wavelength.

It is also important to note that since the depletion region blocks the electrons on the N side from recombining with the holes on the P side, it acts as a capacitor whose capacitance is given by the geometry of the junction $C = \varepsilon_n A/d$. In addition, there is also a transit time associated with the drift of charge carriers in the intrinsic region and it is directly proportional to the distance it must travel and inversely proportional to the carriers' drift velocity. These factors impact the how fast is the photodiode's response performance. The manufacturer measures this capacitance and specifies its value in the Datasheet.

As it was just described, these mechanisms convert light into a current. The current must be converted into a voltage.

The simplest form of current to voltage conversion is the use of a resistance. Ohm's law states the voltage drop across the resistor is proportional to the current going through that resistor: $V = IR$ The proportionality factor is the value of the resistance. In practice, however, it is not so straight forward. The value of the current can be estimated with a back of the napkin calculation by knowing the power of the laser, and the parameters of equation 2.13 given by the manufacturer. As a side note, the laser's power is usually < 1 mW in order to not be harmful, especially to the human eye.

The generated current is small, microampere order of magnitude and so a big resistor is needed to boost the voltage to suitable order of magnitude. This poses some problems, namely, the usage of a high value resistor may interfere with the reading process, since voltmeters and analog-to-digital converters have also a high input impedance (a voltage divider may form), also if the circuit has some additional low value impedances in it, the value of the gain may be decreased due to the parallel equivalence. Furthermore, the system's dynamic response is directly related with the value of the resistor and the capacitance of the PIN junction since they form a low pass filter. And as a final consideration, the noise of the system is tied to the value of the resistor (Johnson noise) and since the resistor is large and the current small current noise may significantly impact the measurements.

The effective solution to this problem is through the use of a transimpedance amplifier. The schematic of the circuit is in figure 2.9.

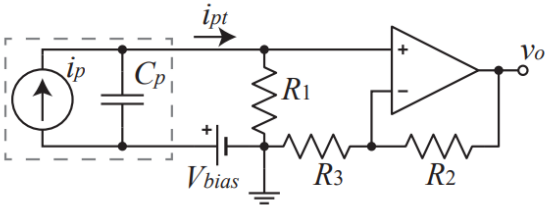

Figure 2.9: Schematic of the photodetector equivalent circuit and the transimpedance amplifier stage. Adapted from [22]

The main component in this is the Operational Amplifier, the triangular shaped symbol. The OpAmp has two inputs - an inverting input, identified as a '$-$' in it's symbol representation, and a non-inverting input '$+$' - an output and two terminals to power the device. Both inputs have an, ideally, infinite impedance so that no current flows into the terminals, whereas the output has, ideally, zero impedance. The output of the amplifier is a signal proportional to the potential difference of both inputs: $V_o = A(V_+ - V_-)$ in order to bring them to the same potential, i.e $V_+ = V_-$. The transimpedence amplifier works in a negative feedback configuration. The output terminal is connected to the inverting input. Following Kirchhoff's laws and keeping in mind that no current flows into the input terminals and they are at the same potential, the output of this particular configuration is :

$$V_o = \frac{R_2 + R_3}{R_3} R_1 I_{ph} \tag{2.14}$$

and so it is obtained a voltage proportional to the current generated by the photodiode.

The position detection system is made up of 4 of these photodiode elements making a quadrant geometry.

The reflected laser beam's spot, which is the transversal shape of the beam, will slight widen as its distance from the focus point increases- this effect is called the beam's divergence. When it reaches the position detection system the laser spot, it will cover some area in the photodetector.

It is important to make sure the laser's optical system is designed with the width of typical cantilevers in mind, in order to preserve the spot's shape because it is important to the PSD system. Given that the laser spot is approximately circular and homogeneous, when the center of the laser spot coincides with the geometric center of the quadrant structure, each photodiode has the same area being lit by the reflected light and thus the amplifier circuits are generating the same voltage. A change in vertical position is detected by adding the voltages from the upper 2 quadrants and subtracting the voltages from the lower quadrants:

$$Normal = (A + B) - (C + D) \tag{2.15}$$

As the laser's spot rises, the area covering the quadrants A and B increases and the photodiodes generate a proportionally higher current which, consequently, is transformed into a larger voltage signal. Conversely, the lit area of both of the bottom quadrants is equally decreased and the output voltage C and D is lowered and so the Normal signal goes up in value. Similarly, attractive forces may pull the cantilever toward the sample, as a result the incident spot on the PSD is lowered. In this occasion, the sum of the C and D signal is bigger than the A and B one. This results in a negative Normal signal.

Additionally, it is also useful to measure the lateral deflection of the cantilever, as it is indicative of the frictional force that arises when a line is scanned: $Lateral = A + C - (B + D)$

2.2.4 Piezoelectric Transducers

The whole system described previously depicts the mechanics through which a voltage value, indicative of the height of the topography, in a single point in space is obtained. Forming picture with some X by X pixels implies the cantilever and the sample under go a relative motion. It does not matter who moves in relation to the other, this choice is only taken int account by the architecture of the AFM. The acquisition is made by traversing a plane, *x0y*, in the sample.The motion is made in a sequential manner, the first direction, which is called the fast direction, starts at one point and is incremented to the end point to produce a line, and then it is followed by a retracement until the start position is reached. After

this the other direction is slightly increased. This is known as a scanning motion. At the end of the slow direction's range the image is formed.

It was shown that we are able to sense nanometric displacements in the z direction and now it is important to laterally move the tip with the a comparable precision in order to resolve the finer details of nanostructures. The only nanopositioning system that successfully accomplishes this feature are the piezoelectric translators. These structures operate under the inverse piezoelectric effect.

The direct piezoelectric effect is observed when a mechanical force is applied to the structure and, as consequence, a voltage difference is measured at its extremities. The mechanical stress physically compresses/elongates the crystalline structure and a net polarization arises at the surface. The reciprocal effect is also observed: the crystal being exposed to an electric field, experiences an elastic strain and deforms, either elongates or compresses according to the polarity of the field. The effect is, in concept, linear and is the desired mechanism to operate the scanning motion.

In ferro-electric materials, most commonly ceramics, the negative and positive charges of the crystalline structure of the material are displaced in a specific direction, this phenomenon is referred as polarization of the medium. In a crystal there can be multiple regions with the same polarization called domains. The domains may be not aligned with each other, however, when an sufficiently strong electric field is applied, the domains will try to align them selves with the direction of the electric field ans thus forming a single domain.

As stated, the piezoelectric effect is, naturally, an effect dependent on spacial directions, since materials are not always isotropic.

The inverse piezoelectric effect is described by the following equation:

$$\Delta L = L d_{31} V$$

where L is the piezoelectric free length, d_{31} is the directional piezo-coefficient and V the voltage applied at its electrodes.(This equation is only valid for small displacement around the equilibrium position, when large enough voltages are applied the material inevitably reaches saturation)

Underlying this equation is the fact that the displacement is not infinitely proportional to the applied voltage, the material inevitably reaches a saturation point.

A shortcoming in the behaviour of the piezoelectric elements is the hysteresis it exhibits. Hysteresis is the name applied to describe a process which is not reversible. After removing the electric field applied to the crystal to polarize its domains, they may not revert back to their previous, somewhat, randomly alignment state and thus there is a (left over) remnant polarization, added to the fact that unaligned domains exhibit some resistance in switching directions.

The hysteresis curve in figure 2.10 graphs this nonlinear behaviour.

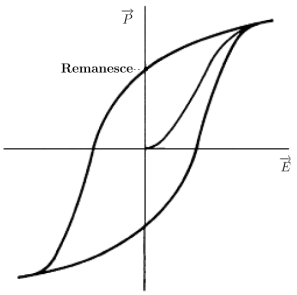

Figure 2.10: Graphical display of the hysteresis behaviour. Adapted from [17]

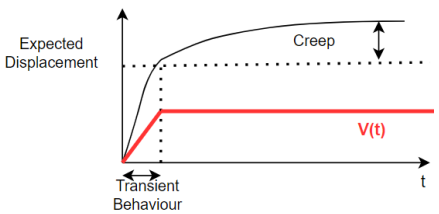

Figure 2.11: Illustration of the creep behaviour of piezoelectric materials. The red curve is the applied voltage whereas the black one is the piezoelectric displacement.

Another non-ideal behavior experienced when dealing with piezoelectric materials is the creep. Creep is the term that describes the non-instantaneous material expansion/contraction. It takes some amount of time for the piezo to fully reach it's steady state value. After the voltage has been applied, the material still continues to expand/contract.

Mainly, there are 2 types of piezoelectric actuators readily available: the piezoelectric tubes and piezoelectric stacks. The main difference between them is the degree of freedom they cover. Piezoelectric tubes are a single structure that can be made to move in all 3 spacial direction, whereas piezoelectric stacks are optimized to displace in one specific direction. Piezoelectric tubes are very sought out tool in AFMs because they're only a single structure that move in all three directions without the need to be coupled to mechanical parts, thus can easily be integrated in any architecture the instrument might have as they can be used to move both the sample or the tip. Moreover, they provide a smooth transition between the different axis, albeit there is some unwanted coupling between them. The devices used in typical AFM operation are structures with only a few centimeters long, where their range of operation goes from a 50 V to a couple hundred volts and displace from one up to a few hundreds of micrometers.

2.2.5 PID Controller

At first sight, it might seem redundant to have a piezoelectric element in the vertical z direction after all, the cantilever, by bending when repelled by the surface, presents a few micrometers of tolerance for the vertical direction. There are 3 problems with having only the length of the cantilever as tolerance for the vertical direction. The first one is that the further the cantilever is bent the more it deviates from its ideal behaviour. The physic models and the equations presented earlier, are valid only under certain approximations and so having the AFM operate under extreme deviations from the models would incur in high inaccuracy and high error of the observed images. Furthermore, the AFM use would be restricted to nanometer scale topographies, certainly, samples with larger slopes, such as cells that may reach micrometer lengths, would be inappropriate. The final reason is the unsafely behind this approach, form the sample and cantilever integrity point of view. Equation 2.1 states that the force felt by the tip and sample is proportional to the displacement and so higher structures would be subjected to higher interaction forces and so would the cantilever. It is likely that permanent damage to both objects would occur.

All of this is avoided by controlling this third degree of freedom with a piezoelectric element. In order to do so, AFM images are conducted in what is called constants force mode, i.e. it is chosen some level of interaction and it is kept constant throughout the hole scanning motion by a feedback control system. In simple terms, the interaction force is set by the user *a priori* via a fixed voltage, named Setpoint. The control system systematical compares the voltage imputed by the user and the voltage provided by the position detection system. When the voltage detected is higher than the required, recall that this means there is a greater force bending the cantilever, the control system applies a decreased voltage to the piezoelectric element in order to reduce its length and thus separate the tip and the sample, bringing the magnitude of interaction back to its desired value. Here the topography is measured in terms of the voltage provided to control the piezoelectric element.

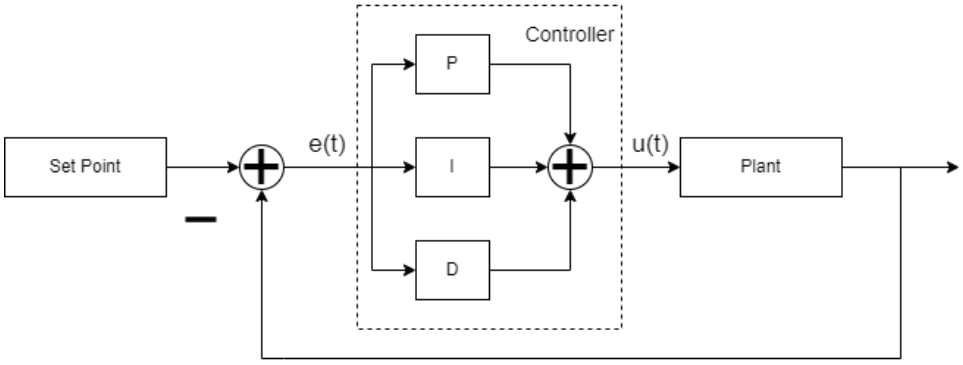

Figure 2.12: Block Diagram of a negative feedback control system, whose controller block is made up of the PID (Proportional-Integral-Derivative) stages

In figure 2.12 is presented the block diagram of a typical negative feedback control system used in AFM imaging. The core component in the block diagram is the PID controller. This is the process that does the ”heavy lifting” in order to bring the tip and sample into a desirable interaction. The input signal to the PID stage is the error signal. As the name implies, it is obtained by subtracting the user defined setpoint value to the signal provided by the PSD. The PID controller then outputs a signal which is given by the following equation:

$$u(t) = K_p e(t) + K_i \int_0^{\tau} e(\tau)\, d\tau + K_d \frac{de(t)}{dt} \tag{2.16}$$

where $u(t)$ is the output of the controller and $e(t)$ is error signal.

The $u(t)$ signal is a weighted combination of a signal proportional, integrated and differentiated of the error term, whose weights are determined by the K_p, K_i and K_d respectfully, which are commonly denoted as the gains of the PID controller.

The negative feedback architecture is employed to enable the improvement of the performance of the system it is controlling, however, if used improperly, it may have disastrous effect.

The PID controller may be seen as general ’one size fits all’ kind of controller, however there are very compelling reasons for it to be implemented in AFMs. Firstly, although we have come up with linear models to describe the instruments operation, the AFMs is nonlinear in nature. There are nonlinear control techniques but due to the complexity of nonlinear problems the design methods are not simple. And secondly, there are too many parameters that are susceptible to change - Choice of cantilevers, sample nature, mode of operation, scanning speed and ambient conditions to name a few. It becomes unmanageable to keep track of all these variables and have a good experience operating the AFM. It would be required a very good estimation guess of all the variables to adjust the control parameters.

Due to these reasons, PID is the appropriate baseline for the control system of an AFM. The gains are adjusted during the imaging process by the user in order to obtain satisfactory performance.

2.2.6 Software

Finally, the last fundamental piece in the AFM tooling kit is the software. Most scientific instruments could not function properly without software aid and, moreover, it interfaces the hardware and user. In regards to the AFM operation, there are some basic features that must be implemented in the software, these may fall into two categories: functional features and feedback features. The functional features are ones that provide direct control over the instrument, i.e command it to do some action. The user needs to interact with image specific tasks :

- Before starting an image, specify the physical size of the image - e.g. a $1\mu m$ x $1\mu m$ square image- and the resolution of the image - 100 samples per line, for instance.
- Set Point selection- Specify the desired interaction strength.
- Line scan velocity selector. This feature provides the user the capability to change the time it takes to sweep a line. This is useful to quickly probe a region of the sample to look for the structures of interest and, along side the tuning procedure of PID gains, make the system stable.
- A start, stop and pause scan buttons. Of course, these are of extreme importance as they coordinate, behind the scenes, all of the different elements that intervene in the the execution of the scan.
- Coarse motion: It is used to move the tip macroscopic distances. This feature is necessary in every AFM instrument, as it is mainly used to bring the tip and sample into contact. Also, whenever the cantilever is replaced or the sample is swapped for another one, since it is safe practice to separate both surfaces a few millimeters so that when they're mounted in their place the user is reassured that collisions are avoided.
- PID Tuning - Adjust the PID gains.

The feedback features offer the user visual information about the ongoing processes:

- Line Trace and Retrace Display - This display plots the sampled values of given line. From here the user can extract qualitative data about the ongoing image formation. Just by looking at this Line Plot it can be seen if the scan is going too fast, are the gains are sufficiently appropriate, are there wanted features in the structure, is there noise in the signal, is the set point appropriate...
- Real Time Display of the overall picture being taken- This display plots the 3D image in real time so the user can qualitatively ascertain the quality of the picture. This is a colour coded 2D version of the 3D Image, each point is plotted in a XY coordinate system with different colors emulating different heights.
- Z piezoelectric current applied voltage - Not only signals the user that the control circuitry is operating as intended but also points out if the piezoelectric element has reached saturation of its operating limits, i.e, if it is fully stretched or at nominal length. When this happens the user need to take action.

Additionally, the software may feature calibration setting and image processing capabilities.

2.2.7 Amplitude Modulation Mode

The above physical principles explanation referrers to contact mode AFM. Contacts mode is also known as static mode AFM. AM AFM is also known as dynamic AFM, the Dynamic part usually refers to more than one technique and so AM specifies it.

There have been developed other techniques to further diversify the instrument's utility. Out of those, it is pertinent to mention AM AFM, which stands for Amplitude Modulation Atomic Force Microscopy. This imaging technique is still used to make topography images of samples, nonetheless, it has significant improvements over contact mode and so it is broadly used in scientific literature. The following is an explanation of its physical principles.

The cantilever is mechanically excited by a force applied to its clamped end. This is usually done with the aid of a piezoelectric element coupled to the cantilever-holder structure. The equation of motion for this system is a obtained through the Euler-Bernoulli theorem stated in 2.4 .

In this equation is missing the damping term. Rigorous analysis of this system is a path rather non-trivial to follow for readers without some considerable background in physics and mathematics. Although it provides a better quantitative description of the system, the overall intuitive description may quickly get obfuscated.

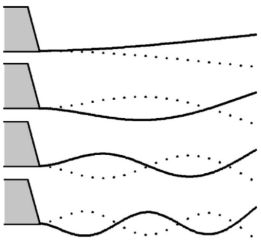

Figure 2.13: Illustration of a few vibration modes of a rectangular cantilever. Taken from [4]

Some take away points that approach this shows is that the geometry of the tip must be considered for determining the tip surface forces. Additionally, because the cantilever is a rigid body, it will have different vibration modes as illustrated in figure 2.13. If the analysis is restrained to system's first harmonic then the best way to understand the behaviour is to approximate the cantilever to a spring system with a point mass attached to it and a dashpot to insert some dissipation in the system to make it more realistic, this ways the oscillations do not continue indefinitely. By replacing the rigid body problem to a more simple one the dynamics are much more easy to navigate. The equation of motion of the system is the following second order differential equation (eq. 2.17):

$$F = -kz - \gamma\frac{dz}{dt} + F_{exc} \tag{2.17}$$

where $F_{exc} = A_0\cos(\omega t)$ is the periodic force applied to the clamped end - it has some amplitude and frequency ω that determines how many times per second it pushes the cantilever. This equation represents what is called in physics a forced harmonic oscillator and its behaviour is very well known.

Differential equations exhibit the superposition property. It is called homogeneous if the equation only has terms dependent on the unknown function and its derivatives, here is an example : $y'(t) + a(t)y(t) = 0$. An non-homogeneous equation has terms that are not explicitly dependent on the unknown function and its derivatives, such as : $y'(t) + a(t)y(t) = b(t)$. It follows that the general solution of the equation has the following form : $y(t) = y_H(t) + y_{NH}(t)$ where y_H are the solutions of the homogeneous equation and y_{NH} may be a particular solution of the non-homogeneous equation. In essence, the behavior of the system is a superposition of both an harmonic oscillator and a forced oscillator.

Lets us consider then the homogeneous equation:

$$F + kz + \gamma\frac{dz}{dt} = 0 \tag{2.18}$$

It only differs from equation 2.2 due to the dissipative term. It is included because it provides a more realistic description of the system.

The solutions to equation 2.18 reveal the tip's motion takes the form:

$$z(t) = z_0(t) + z_1(t) = C_0 e^{-\left(\frac{\gamma}{2m} + \sqrt{\frac{\gamma^2}{4m^2} - \frac{k}{m}}\right)t} + C_1 e^{-\left(\frac{\gamma}{2m} - \sqrt{\frac{\gamma^2}{4m^2} - \frac{k}{m}}\right)t} \tag{2.19}$$

where C_0 and C_1 are constants determined by the initial velocity and initial position of the tip.

Looking more closely at the exponents of 2.19:

$$\frac{\gamma}{2m} \pm \sqrt{\frac{\gamma^2}{4m^2} - \frac{k}{m}} \tag{2.20}$$

There are 3 possible behaviour regimes that are could take place based on the physical conditions of the system and all three of them can occur in AFM. The term $\gamma^2/4m^2 - k/m$ is either positive, zero or negative dependently of the magnitude of the terms inside the square root.

- Positive - the damping resistance dominates the stiffness term resulting in a heavily damped system. In this case, the system slowly returns to its equilibrium. Any new stimulation decays exponentially.
- Zero - there is a balance between the two terms. The system is said to be critically damped. This is the tipping point in the behaviour of the system as it will converge to its equilibrium position as quickly as possible.
- Negative - The system is lightly damped and the result is a oscillatory motion around the value system's equilibrium point that fades away as time moves on.

For this discussion the lightly damped case gives the most important behaviour. The other two cases do not produce oscillatory motions.

When the expression inside the square root is negative the exponent becomes a complex quantity, where it is rewritten $\sqrt{-1} = i$:

$$\pm i \left(\sqrt{\frac{k}{m} - \frac{\gamma^2}{4m^2}}\right) \tag{2.21}$$

The term inside the curly brackets must have dimension of inverse of time, i.e frequency, thus the systems gains a new frequency:

$$\omega_{new} = \sqrt{\frac{k}{m} - \frac{\gamma^2}{4m^2}} \tag{2.22}$$

As we have seen, the system oscillates around is equilibrium point with a some frequency lower than its undamped counterpart, if the system is only slightly damped then the new resonant frequency is practically the same as the natural frequency ω_0. However, due to the real exponent, this oscillatory movement decays with time and ultimately converges so its equilibrium point. Figure 2.14 illustrates this motion.

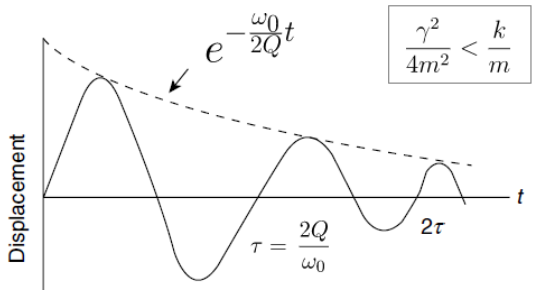

Figure 2.14: Illustration of the damped oscillatory motion as a function of time. The amplitude is decreased exponentially. Adapted from [19]

The constants C_0 and C_1 are still to be determined at the initial instance of the motion, as discussed in the simple harmonic motion case above.

This is the systems initial reaction when perturbed by the excitatory force.

Now we need to consider the case where the periodic motion is applied.

Intuitively, the result from this equation will provide the steady state behavior of the system because system is being driven by the excitation and it has no alternative but to follow it. Since the applied force is oscillatory we expect to find results in the same form: $z = A\cos(\omega_{exc}t)$. From the earlier results, we can express the excitatory periodic force with a pure complex exponent: $F_{exc} = Fe^{i\omega_{exc}t-\phi}$:

$$m\frac{dz^2}{dt^2} + \gamma\frac{dz}{dt} + kz = F_{exc} = Fe^{i\omega_{exc}t} \tag{2.23}$$

$$(-A\omega_{exc}^2 m + i\omega_{exc}A\gamma + Ak)e^{i(\omega_{exc}t-\phi)} = Fe^{i\omega_{exc}t} \tag{2.24}$$

By comparing the terms in the right and left side, thought the Euler relationship: $e^{i\delta} = \cos\delta + i\sin\delta$, it is obtained :

$$A(\omega_{exc}) = \frac{F}{\sqrt{(k-m\omega_{exc}^2)^2 + \gamma^2\omega_{exc}^2}} \equiv \frac{F \setminus m}{\sqrt{(\omega_0^2 - \omega_{exc}^2)^2 + (\frac{\omega_0\omega_{exc}}{Q})^2}} \tag{2.25}$$

$$\phi(\omega_{exc}) = \arctan\left(\frac{\gamma\omega_{exc}}{k-m\omega_{exc}^2}\right) \tag{2.26}$$

Where the expression for the Amplitude was rewritten with ω_0, the undamped harmonic natural frequency 2.3, so that the behaviour the equation describes may be more intuitively recognized. Also, the is its defined term $Q = \omega_0 m/\gamma$ as it is a quantity very commonly used when speaking about harmonic oscillators. It is measure for the quality factor of the harmonic oscillator and is a dimensionless variable. If Q is high, there is low damping and the system oscillates with a frequency very closely to its natural frequency as it can be seen by rewriting equation 2.22 :$\omega_{new} = \omega_0\sqrt{1-1/2Q^2}$. The quality factor can also be though as a measure of how (much time)many oscillations does the system go until its simple harmonic amplitude is reduced by a factor of e, by analyzing the real exponent from equation 2.19:

$$e^{-1} = e^{-\frac{\omega_0}{2Q}t} \longrightarrow t = \frac{Q}{2\omega_0} \tag{2.27}$$

More importantly, the maximum amplitude of the oscillation can be written to be proportional to Q. The maximum amplitude is obtained by finding the excitation frequency at with the amplitude graph as

a maximum, i.e differentiating equation 2.25 with respect to the excitatory frequency:

$$\frac{dA}{d\omega_{exc}} = 0 \longrightarrow \omega_{max} = \omega_0\sqrt{1-\frac{1}{2Q^2}} \tag{2.28}$$

and thus $A_{max}(\omega_{exc} = \omega_{max}) = FQ^2/k\sqrt{Q^2 - 1/4}$.

In the case Q is high, one can approximate $\omega_{new} \approx \omega_0$ and so the maximum amplitude the system oscillates is when it is excited at its natural frequency : $A_{max} = QF/k$.

As we can see from the above results, the state of the system, i.e its Amplitude and Phase, is completely defined since there is no initial value constants left to be determined. One is able to say that this motion has no memory of its initial state.

Qualitatively, when the system is at rest and it is disturbed with the excitation force its first reaction is to oscillate at its natural frequency, in case there is no damping, but as time moves on this initial reaction is transitory and exponentially decreases until it reaches its steady state motion which is forcibly imposed by the excitation force. The system's steady state motion is a sinusoidal function that has the same frequency of the excitation force ω and has a phase lag, ϕ, with respect to it. The phase is an angle and is a measure of the delay of the response of the system has in relation to its stimuli. In this case, it measures the response's delay to the excitatory force.

The previous considerations about the system's motion did not include the interaction force at the tip of the cantilever. The analysis was done with only with the excitatory force in consideration. Now it is time to introduce the tip-sample interaction forces, however some difficulties arise due to the non-linearity nature of it. During the tip motion, as illustrated in figure 2.15, the interaction takes distinct forms at various lengths and time instances (tips position changes with time).

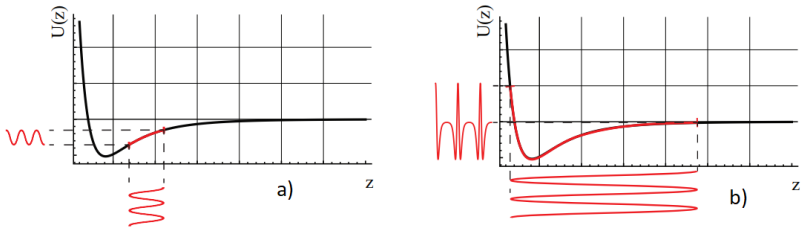

Figure 2.15: Illustration of the different oscillation regimes present in Dynamic Mode AFM. In a) the vibration amplitude is small enough to allow the consideration of a linear interaction between tip and sample. In b) the excitation amplitude is too large and so the interaction is no longer linear.

For instance, the jump to contact phenomena won't be avoided in softer cantilevers. To further develop the model, it is paramount to impose small amplitude oscillations so that a linear relationship may be built. At small displacements, we may use the first term of Taylor's expansion to draw some more meaningful physical associations:

$$F_{int} = F_{int}(0) + k_{int}(z - z_0) \tag{2.29}$$

where $k_{int} \equiv -F_{int}/\partial z$. Under the assumptions that Taylor's expansion is valid, the functional behaviour of $\partial F_{int}/\partial z$ is very identical to that of an elastic response

The combination of equation 2.17 with 2.29 produces an effective spring constant:

$$k_{eff} = k_c + k_{int} \tag{2.30}$$

The dynamics of the system are then influenced accordingly:

$$\omega_{int} = \sqrt{\frac{k + k_{int}}{m} - \frac{\gamma^2}{4m^2}} \tag{2.31}$$

The Amplitude and Phase, looking at equations 2.25 and 2.26, it can be established that there is an inverse relationship with these quantities, and so when the interaction gets stronger, if the systems is excited at the resonant frequency, the amplitude decreases and it is no longer $\pi/2$ radians out of phase with the excitation. This presents a mechanism that alters the amplitude and phase of harmonic oscillator with the interaction force.

Is is important to note AM AFM is sensitive to the gradients of the interaction force, or in other words, the slope of the force curve shown in figure 2.2, but not to the force. The force only changes the equilibrium position around which the cantilever oscillates.

There is another important remark to be made about this imaging mode. Looking back at equation2.28, it shows that when the tip and sample do not interact, i.e they are *infinitely* apart, the system will be excited very closely to its natural frequency, where its amplitude is greatest. As the approach begins the tip and sample interact and therefore the system's resonant frequency will be modified according to equation 2.31. Since the excitation frequency is made constant, the actual oscillation amplitude will have a new value determined by the new resonant curve, as illustrated by image 2.17. The resonant curve curve shifts based on the slope of the force curve. Taking into account figure 2.2, when not in contact the force is a horizontal curve and so its slope is 0. As the distance shortens - from right to left in the picture - the slope begins to increase and since $k_{int} = -\partial F_{int}/\partial z$ the effective spring constant is lower than the free one and so the resonant curve will shift to the left. The amplitude will be smaller than the previously free of interaction and, most importantly, it will be located to the right of the resonate peak. Once contact is established, the repulsion part of the curve has negative slope and thus the effective spring constant will increase, shifting the resonant curve to the to the right and so the amplitude increases until it the peak of the curve is reached and then begins to decrease.

This poses a serious problem to the control system as it is only capable of controlling signals that strictly decrease or increase, as per the error equation $e(t) = V_1 - V_2$, depending on the order of the terms.

To overcome this pitfall, user predicts this behaviour and must set the frequency of the excitation signal to some value sightly smaller than the free resonant frequency. When approaching, the amplitude starts at below its maximum value, increases subsequently in the attractive part of the interaction and finally, in contact decreases until the desired set point value is reached.

2.2.8 Lock-in Amplifier

The reason for establishing dependencies with the Amplitude and Phase is because they are the measurable quantities in this mode. To carry out these measurements, it is used an instrument called a *Lock-in* Amplifier. This instrument is used to detect and measure very small AC signals with extremely good accuracy, even when the signal is obscured by noise.

In essence, the *Lock-in* Amplifier generates a sinusoidal signal with a certain frequency, a reference, used to excite some system. In this case it is used to excite the piezoelectric element that applies a force to the base of the cantilever. Subsequently, it receives, as an input, the response of the system and it extracts the response of the system at that reference frequency. When the system is excited with a periodic disturbance, it's response may be decomposed as a sum of sines and cosines with different

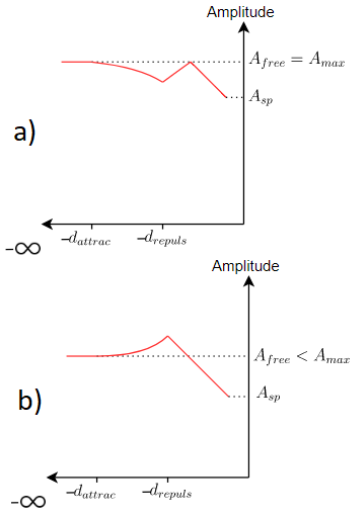

Figure 2.16: Illustration of the vibration amplitude's variation as a function of the distance between sample and probe. The distance was arbitrarily chosen to be negative for the illustration purposes, d_{attrac} is meant to represent the distance at which the attractive part of the interaction is dominant and d_{repuls} the distance at which the interaction is repulsive. In a) it is chosen a excitation frequency that matches the resonant frequency of the cantilever. In b) the excitation frequency is chosen below the resonant frequency.

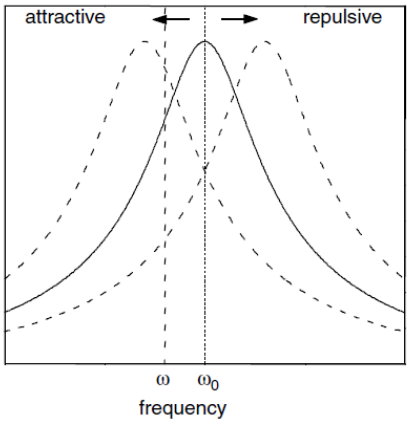

Figure 2.17: Illustration of the shift in the Resonant curve of the cantilever as it is being subjected to either an attractive force, which shifts the resonance to lower frequencies or repulsive force. Taken from [16]

frequencies, each with its own amplitude, this is what the Fourier Theorem states:

$$F(\omega) = \mathscr{F}\{f(t)\} = \frac{1}{2\pi i}\int_{-\infty}^{+\infty} f(t)\cdot e^{-i\omega t}dt \tag{2.32}$$

The *Lock-in* amplifier then takes advantage of some properties encountered in the frequency domain analysis. It performs point by point multiplication of the reference signal that excites the system by the response of the system it is receiving as input. The result may appear to be an odd signal but in the frequency domain it is defined as having two frequencies, one as being the algebraic sum of frequencies of both frequencies and the other as being the difference of them. In case of a pure sinusoidal excitation, $V_{ref} = V_0\cos\omega_1 t$ and a pure sinusoidal response from the system, $A\cos(\omega_2 t + \phi)$, where ϕ is the phase difference with respect to the reference signal, ω_1 and ω_2 are some arbitrary value of frequency.

The output of the multiplication state, called mixing stage, in the time domain is:

$$v_{mix}(t) = v_{ref}(t)\cdot v_{sys}(t) = V_0\cos(\omega_1 t)A\cos(\omega_2 t + \phi)$$

and its Fourier Transform is:

$$V_{mix}(\omega) = \frac{1}{2\pi i}\int_{-\infty}^{+\infty} V_0\cos(\omega_1 t)\cdot A\cos(\omega_2 t)\cdot e^{-i\omega t}dt$$

From the Euler relationship, we can rewrite a cosine in the complex representation as:

$$\cos\delta = \frac{e^{i\delta} + e^{-i\delta}}{2}$$

resulting in :

$$V_{mix}(\omega) = \frac{V_0 A}{2\pi i}\int_{-\infty}^{+\infty}\frac{1}{2}\left[(e^{i\omega_1 t}+e^{-i\omega_1 t})\cdot(e^{i\omega_2 t+\phi}+e^{-i\omega_2 t+\phi})\right]\cdot e^{-i\omega t}dt$$

$$= \frac{V_0 A}{2\pi i}\int_{-\infty}^{+\infty}\frac{1}{2}\left[e^{i(\omega_1+\omega_2 t+\phi}+e^{-i(\omega_1+\omega_2)t+\phi}+e^{i(\omega_1-\omega_2)t+\phi}+e^{-i\omega_1-\omega_2 t+\phi}\right]\cdot e^{-i\omega t}dt$$

$$= \frac{V_0 A}{2\pi i}\int_{-\infty}^{+\infty}(\cos((\omega_1+\omega_2)t+\phi)+\cos((\omega_1-\omega_2)t+\phi))\cdot e^{-i\omega t}dt$$

$$\longrightarrow V_0 A\left[\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\cos((\omega_1+\omega_2)t+\phi)\cdot e^{-i\omega t}dt+\frac{1}{2\pi i}\int_{-\infty}^{+\infty}\cos((\omega_1-\omega_2)t+\phi)\cdot e^{-i\omega t}dt\right]$$

$$\longrightarrow V_0 A(\mathscr{F}\{\cos((\omega_1+\omega_2)t+\phi)\}+\mathscr{F}\{\cos((\omega_1-\omega_2)t+\phi)\})$$

and so it is shown that a multiplication of two sinusoidal signals in the time domain is the sum of the Fourier Transforms of those signals at $(\omega_1+\omega_2)$ and $(\omega_1-\omega_2)$ in the frequency domain.

The result of the Fourier Transform of the mixed signal is stated below:

$$V_{mix}(\omega) = \frac{AV_0}{2}\sqrt{\frac{\pi}{2}}(e^{-i\phi}\delta((\omega_1+\omega_2)-\omega)+e^{i\phi}\delta((\omega_1+\omega_2)+\omega)+e^{-i\phi}\delta((\omega_1-\omega_2)+\omega)+e^{i\phi}\delta((-\omega_1+\omega_2)+\omega)) \tag{2.33}$$

where $\delta(x)$ is the Dirac's Delta Function which, without going into detail, is null everywhere except when its argument is zero, in which case it is equal to one. (The result of the Fourier Transform of a pure sinusoidal function such as $\cos(\omega t)$ are two points in the real plane which are symmetrical with respect to the ordinate axis)

If the system is linear the system's response signal will have the same frequency as the excitation and this is the central argument of the Lock-in because the output of the system may be buried in noise that contains a continuum spectrum of frequencies however, as we shall see, the Lock-in extracts the wanted signal as a result of it being present at 0 Hz. The mixed signal is fed to a low pass filter to extract this DC component, hence signals containing frequencies other than the reference frequency, are rejected and do not affect the measurement. The value outputted from the low pass filter stage is however, sensitive to the phase, since the real part of 2.33, where $\omega_1=\omega_2$ is $\propto A_{LP}\cos\phi$, where A_{LP} is a gain that the Low Pass filter may have.

As for the phase, the output of the low pass filter stage, now has no information about it. In order to recoup the phase of the signal, an additional mixing stage is added, in parallel to the amplitude demodulation one, where the reference signal is 90° shifted in relation with the adjusted reference signal phase, as can be seen in figure 2.18. With this configuration, the output of the this low pass filter stage is $V_{out\,ph}\propto A_{LP}\cos(\phi+\frac{\pi}{2})\leftrightarrow A_{LP}\sin(\phi)$

Denominating the output of the fist stage $X=V_{sig}\cos\phi$, which is called the *in-phase* signal, and $Y=V_{sig}\sin\phi$, called *quadrature* component of the signal. These two quantities define a vector representative of the Amplitude and Phase of the system's response relative to the reference signal. Hence, it is defined the signals R and Θ that represent the Amplitude and Phase difference, respectfully :

$$R = \sqrt{X^2 + Y^2} = V_{Amplitude} \quad , \quad \Theta = \arctan\left(\frac{Y}{X}\right) = V_{Phase} \tag{2.34}$$

Most often, images are made in constant amplitude, so it is the variable being controlled in the feedback loop.

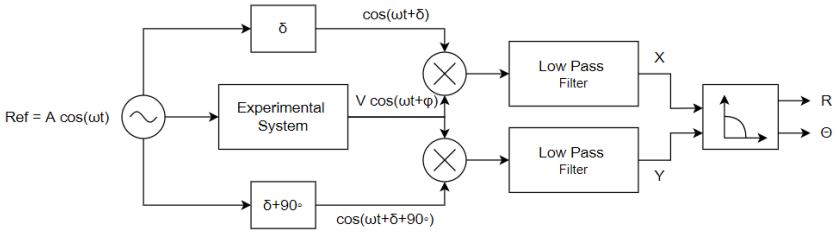

Figure 2.18: Simplified Block Diagram of the operation principle of the *Lock-in Amplifier*

2.3 Atomic Force Microscope Prototype Setup

In the previous general considerations it was only addressed the immediate necessary elements that make up the AFM. In spite of this, a lot of complementary equipment was left out of the discussion because they do not play a primary role in the core working principle but are, nonetheless, fundamental in a practical setup of the instrument. These are only taken into account at the transition from concept to practical implementation and, their understanding is vital not only because they are needed but also because the performance of the AFM is dependent of them.

This section will specify those extra tools. The specifics of each component selection/feature will, naturally, vary from setup to setup. The goal of this discussion is to showcase what other tools are in the background of an AFM laboratory since they might not be obvious from the description of the working principle.

In a successful AFM implementation, it is essential to shield the instrument from mechanical vibrations and acoustic noise. The simple act of walking or talking inside the lab or, even worse, a knock on the table or workbench where the AFM rests is enough to disturb the measurements. For proper mechanical isolation, the AFM must be in contact with either a highly damping structure, which is the case of the commercial AFM available in the laboratory - the AFM is suspended by four elastic ropes. Or it should rest above a non-continuous medium. This was the solution proposed by professor Mário Silveira and PhD Miguel Vitorino to shield the Force Feedback Microscope (FFM) from mechanical noise [23]: A wood box is filled with sand and the AFM is placed above its lid. The lid is cut in order to reduce it's dimensions so that it does not make any contact with the edges of the box, only the sand, as can be seen from picture 2.19.

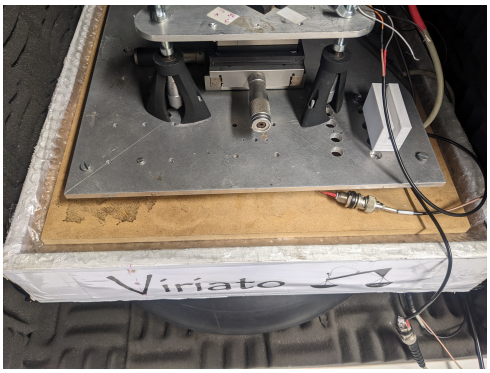

Figure 2.19: Wood box filled with sand in order to protect the AFM from mechanical noise

Then it is customary to place it inside an enclosure with sound damping foam covering its walls.

An important device that could not go without mention is what is known as the *Scanner*. It is a mechanical structure where piezoelectric element stacks are mounted to allow for the nanopositioning of the sample. It consists of an enclosure with a mechanical flexure in one side and a hard wall on the other. When a voltage is applied the piezoelectric element will expand and exert a force on flexure that is elastically deformed. This not only allows for the displacement of the sample is the direction the piezoelectric element is pushing but will also displace the sample in the opposite direction when the voltage is reduced and the piezo's length is brought closer to its equilibrium size. Figure 2.20 illustrates it's working principle.

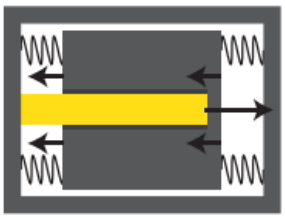

Figure 2.20: Illustration of the Scanner's working principle. The yellow figure represents the piezoelectric stack. Taken from [22]

The *scanner* can be made to operate in all the 3 direction or in just the *x0y* plane. When the AFM employs a Piezoelectric Tube it, usually, does not need this device.

This leads to voltage amplifiers. As stated in the previous section, the piezoelectric Elements typically used operate under the $0-150$V range. This might not be considered high voltage operation but it is an order of magnitude greater than the standard voltage range that common laboratory electronic equipment is prepared to handle - the standard is -10 to 10V input and output. These are specialized amplifiers due to not only providing a suitable gain at the standard -10 to 10V voltages but also because they are designed to drive a high capacitance load. Also, the amplifier must have have 3 independent channels, one for each spacial direction.

Another must have tool in the AFM is a micrometric table under the sample holder. This allows a repositioning either the sample or the cantilever, in the *x*0*y* plane distances of a few centimeters with sufficient resolution. This is useful for the user to place the probe over region of interest. In addition, the positioning is aided with a camera/optic microscope placed above the sample, providing the user a direct view of the sample, in order to facilitate this process.

Besides the *x*0*y* plane, it is also indispensable a micrometric motion actuator in the vertical direction

as it provides the coarse motion of the tip-sample approach. This vertical coarse motion is crucial for the handling of the instrument. It is used to distance the tip and sample a large amount, relative to typical micrometer length scales, as it eliminates the possibility of a tip-sample crash, hence it leave the instrument at a safe state. This feature is used whenever the user needs to replace either the sample or the probe, or simply when turning the instrument off.

And the last pieces of equipment that are a necessity in the AFM is an Analog to Digital converter, that is utilized to sample the control system output voltage and transfer them to a computer, and a Digital to Analog converter in order to provide a voltage source to the piezoelectric element's Amplifier, responsible for the scanning motion, with the aid of the instrument's software.

2.3.1 Characterization of the AFM prototype

The AFM prototype this work builds upon, began its development in Arthur's master thesis [22]. The Microscope can be divided into two mechanical segments: the *Base* and the *Head* which are detachable from one another.

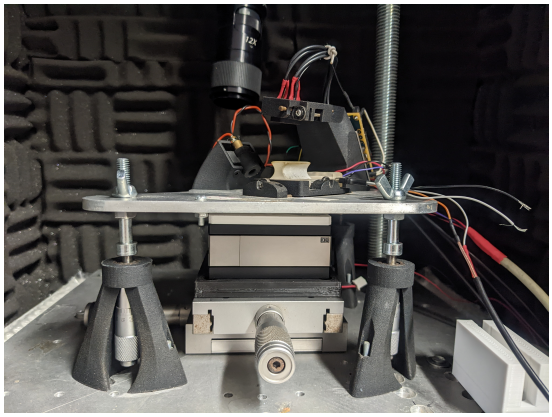

Figure 2.21: AFM prototype this work is based on.

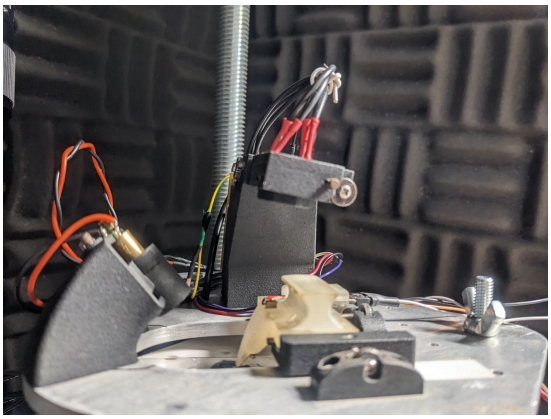

Figure 2.22: Close up of the head of the AFM. It can be seen the laser diode in the left and at the center is shown both the cantilever holder (light colored structure) and the PSD tower.

The whole structure is fixed on a metal plate. The *Base* section is composed of 2 micrometric tables stacked on top of each other and, on top of them is fixed the *Scanner*. The tables allow the user to macroscopically adjust the position of the *scanner*, and hence the sample, in the x and y direction. The *Scanner* consists in a plastic piece where 2 piezoelectric elements stacks are inserted in order to perform the scanning motion in the x and y direction.

On top of the *XY Scanner* it is mounted the sample holder. In this piece is where the z axis piezoelectric element stack is inserted. Similarly to the $x0y$ mechanism, the *piezo* pushes against a solid surface (the *XY scanner*) and a flexure, which in this case is the sample holder. As can be seen from figure 2.23, the sample holder as the shape of a table which is screwed in place. With this type of architecture, the sample is the one who moves relatively to the probe.

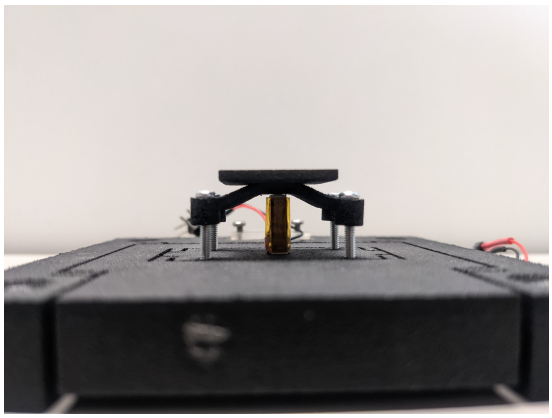

Figure 2.23: The picture shows the scanner of the AFM prototype. The black bed is the *XY scanner* where it is coupled the vertical axis. In the picture it can be seen the Z piezoelectric element stack that sits below the sample holder.

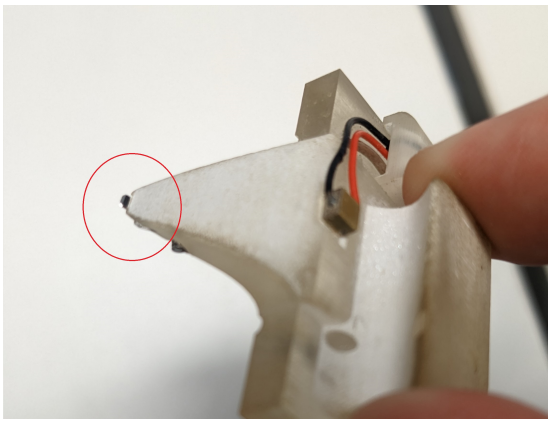

Figure 2.24: Close up of the cantilever holder. It is highlighted the cantilever place. At the top of the structure it can be seen the piezoelectric element (yellow square) that produces the excitatory force in order to drive the cantilever at it's mechanical resonant frequency

Currently, there are in place *XY* piezoelectric element stacks that cover an effective $30 \times 30 \mu m^2$ scanning area, whereas the *z-piezo* only has an effective expansion of $7,7 \mu m$. All of them are meant to be operated in the $0-75$V range. Also, the manufacturer specifies a 115 kHz resonant frequency for the *z-piezo*.

The *Head* is a triangular shaped piece of aluminium where the cantilever holder, the laser and the PSD are held in place. The cantilever holder is also a detachable structure in order to aid the replacement of the probe. This piece has a tetrahedral shape as can be seen from figure 2.24 and at the vertex there is a a spring that hold the chip of the cantilever in place. At the top of it can be seen the small piezoelectric stack that is responsible for the mechanical excitation of the cantilever in the dynamic mode. This structure has an alignment mechanism. This alignment stage allows a lateral adjustment of the cantilever so that it's position relative to the laser is regulated and the laser beam spot is reflected.

At the opposite side, is where the laser and its optics are located. The laser is located above the cantilever holder and is pointed downwards, in order to hit the back side of the cantilevers. The laser emits a red beam and it has less than 1mW of power so that it's operation does not require the usage of special eye equipment.

Above, is located the Position Detection System which is supported by a tower structure that conveniently stores the amplifier electronics. It has a four quadrant geometry made of PIN diodes. The active area of the PSD is also mounted in a adjustable system with two degrees of freedom. This serves to align the laser spot, that has been reflected off the cantilever, in the geometric center of the PSD. All of adjustment stages have two screws each for the user to move the plane in the two dimensions and thus, align the cantilever with the laser beam or the reflected laser beam spot with the center of the PSD's active area, prior to the start of an image acquisition.

There are a total of four transimpedance amplifiers, one for each quadrant. The photodetector and amplifier equivalent circuits is shown in figure 2.9.

The output of this stage is given by equation 2.14. The selected gains were calculated taking by estimating the maximum current generated by the PIN junction, with the help of the laser's power and the reverse bias voltage.

The resistor values selected were: $R_1 = 16.2k\Omega$ $R_2 = 9,09k\Omega$ $R_3 = 1k\Omega$. Totaling a transimpedance gain of $G = 163.5k\Omega$.

v_V

The voltages provided by these pre-amplifiers are then fed to a summing circuit as shown in figure 2.25.

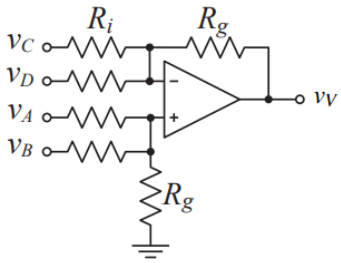

Figure 2.25: Schematic of the summing amplifier circuit. Adapted from [22]

This is to compute the Vertical and Horizontal signals.

The output of this circuit is given by :

$$V_O = \frac{R_g}{R_i}(v_1 + v_2 - v_3 - v_4) \tag{2.35}$$

The Vertical and Horizontal signals are obtained by placing the preamplifier output voltage from each photodetector quadrant into its corresponding place.

The DC gain of this circuit is approximately 2.13.

The PSD outputs a voltage proportional to the deflection angle as long as the spot beam is near the geometric center of the four quadrants. The sensor remains under linear operation for small deflections, this condition is met while imaging. The OpAmp will saturate if the laser spot vertical or horizontal displacement covers only 2 of the four quadrants as is depicted in figure 2.26.

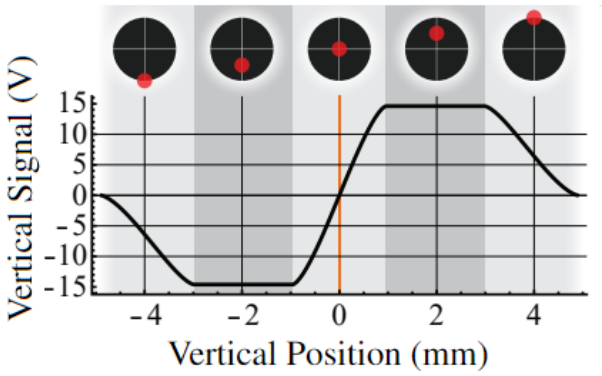

Figure 2.26: [
Illustration of the Position Sensitive Detection's Vertical Signal behaviour with respect to the position of the laser's spot]Illustration of the Position Sensitive Detection's Vertical Signal behaviour with respect to the position of the laser's spot. Adapted from [22]

The *Base* has 3 vertical pillars that support the *Head*. Each pillar is embedded with a micrometric screw. Their purpose is to raise or lower the *Head*. It it through them that the approach of the tip-sample is made. Up until now, the user would need to manually turn the screws and monitor the signal of deflection/ Amplitude in order to establish contact, disengage or simply adjust the height. Having a human perform these actions, is time consuming and makes the process highly error prone. At one of the pillars, there is a DC motor coupled via a transmission belt (made out of a rubber band). The motor is driven by a typical *H-bridge* circuit as shown in figure 2.27.

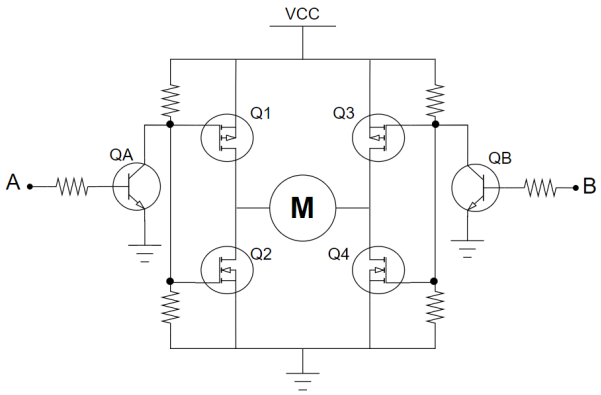

Figure 2.27: Schematic of a typical H-Bridge circuit that allows for a bidirectional control of the motor's rotation. There a total of 6 transistors, 2 npn bipolar transistors, 2 N-type and 2 P-type MOSFETs. V_{CC} is the power supply voltage, which is typically about 12 V. Terminals A and B are where the HIGH or LOW signals are applied.

This enables the bidirectional control of the coarse motion with the use of a digital signal, here's how it works:

There are a total of 6 transistors labeled Q_A, Q_1, Q_2 on the A terminal and Q_B, Q_3, Q_4 on the B side, that are used as a switch. There are only two regions of operation that are of interest - the transistor saturation region, where a maximum current flows through the Drain to it's Source, and the cut-off region, where only a minuscule amount of current flows through it, it is in practice zero. Transistor Q_A and Q_B are NPN bipolar transistors whereas,Q_2 and Q_4 are N-channel MOSFETs and, in turn, Q_1 and Q_3 are P-channel MOSFETs . N-type transistors enter their saturation region when the voltage between it's Gate and Source is greater then some threshold value (usually 0.7V) and enter their cut off region when the voltage is much lower than this threshold. P-type transistor on the other hand, enter their saturation region when the voltage between their gate and source is lower than the threshold, which is usually around -0.7V and are turned off when the voltage is greater than it.

The circuit works by providing a High level one of the terminal and a Low level to the other terminal.

Notice that the The P-channel transistors have their Source at the V_{CC} potential, thus the condition for the Saturation region is $V_g < VDD + V_{th}$. Applying a High voltage to the A side will open the Q_A transistor. Since its Source is at ground potential the gate of Q_1 will also be at 0V. This satisfies its saturation condition and this transistor conducts current. On the other hand, Q_2 has its Source and Gate at the same potential and so it is in its cut off region.

On the B terminal a low level voltage is applied to Q_B meaning this transistor is on its cut-off region. The Voltage at the gate of Q_3 is set by the resistor connected to V_{CC}. The voltage drop at the resistor is chosen so that it ensures Q_3 remains within the cut of region, i.e small voltage drop. The gate of Q_4 is close to the V_{CC} level and so the condition for its channel being open $V_g - V_s > V_{th}$ is satisfied. A current will flow from V_{CC} into the motor and into ground and the motor spins in some direction. Conversely, applying a High level voltage at the B and a Low level voltage at the A side opens a path for the current to flow from V_{CC}, through Q_3, into the motor and through Q_2 into ground, making the motor spin in the opposite direction.

With this configuration, if the same voltage is applied to either side then the motor will not spin. If a High voltage is applied both Q_A and Q_B are closed, the gates of Q_1 and Q_3 are at the V_{CC} potential and so are also closed, whereas Q_2 and Q_4 are open and so the motor its at ground potential. If a Low voltage is applied all the transistors reserve their states and both terminal of the motor are at V_{CC} potential. Since no difference in potential is applied there is no current flowing and the motor stays still.

The circuit also has the benefit of isolating the digital signal source from having to provide the current

in order to spin the motor.

Besides the physical portion of the AFM, there is also the external complementary circuitry as mentioned above.

At the Lab, there is exists piezoelectric driving Amplifiers that were commercially acquired. These amplifiers were designed especially for this purpose, to drive piezoelectric actuators. It receives as input voltages from the standard -10 to 10V range and outputs a fixed $\times 15$ DC gain signal.

For the Amplitude Modulation operation it is used a digital *Lock-in Amplifier* (PLL) that has several useful features. Firstly, it provides a frequency sweep used to find out the resonant frequency of the cantilever. This provides an immediate feedback to the user, since not only allows for a intuitive selection of the excitatory force frequency, but also, the shape of the curve is indicative of the laser alignment with the cantilever and also of the overall physical conditions of the cantilever. The other feature of use is the ability to change the low pass cut off frequency and the DC gain of the output filter stage. It is always a trade off when adjusting this setting. The lower the filter's cut off frequency the better noise attenuation is achieved however, the output voltage will take more time to converge to it's final value, thus slowing the measurement.

Lastly, the AFM tour is completed at the data acquisition system, which encompasses the DAC and ADC. The data acquisition hardware used for this prototype is a *National Instruments chassis* cDAQ-9177. This is a hardware-computer interface that fit up to four modules sold by *National Instruments*. The modules that are of use in this application are the DAC, named NI 9263, and an ADC, NI 9215. Both modules have 4 channels with 16 bits resolution on the standard voltage range with read/write speeds of up to 100 thousand samples per second, which equates to one sample each 10 μs. These are high performance devices and, perhaps, the main reason for the use of this systems is because of the software support provided by *National Instruments*. It not only provides a graphical programming environment called LabVIEW, which has disseminated through the scientific community, but also because it provides a simple API- short for Application Programming Interface that maybe thought of as a library of functions and protocols that abstract away low level code - for easy integration in a custom software, which is the case for this project. This *chassis* connected to the computer via a USB port. It is through this connection the commands for the *chassis* are transmitted and also by which it sends the collected data.

This is, of course, managed by the software trough which the user interacts. It was custom made for this specific prototype and it was named *Vegrandis*.

Figure 2.28: Photograph of the *National Instruments chassis* cDAQ-9177 that contains both an ADC and DAC module.

Chapter 3

Control System

3.1 Insights on control

The design of control systems can be divided into two distinct parts. One is concerned with the design of individual components, the other with the design of the overall systems by utilizing existing components [15]. This requires expertise on both control and instrumentation fields.

In this chapter, it will be covered, albeit briefly, important aspects regarding control theory in general and some particular to the AFM instrument. Later we shall discuss the individual elements that make up the Control System.

The chain of operation of the system can be described by the functioning principle of it's constituent subsystems, which occupy a specific place in the logic of the broader system. Here every subsystem will be represented as a block and each block represented by a transfer function. The block usually has one input and one output. The transition from one block to the other is represented by an arrow, as usual.

Physical systems are described by differential equations. One useful method of solving the differential equations, given some perturbation to the system, is with the Laplace Transform. With the Laplace Transform, a linear differential equation may be rewritten as a polynomial equation whose variable has both real and imaginary parts, it is usually denoted as $s = \sigma + i\omega$.

A very useful insight is that the Fourier Transform is a special case of the Laplace Transform, with $s = i\omega$. The former represents a function by a sum of sines and cosines and the later does the exactly same but also with adds exponential terms akin to transient behaviour.

In simple terms, the control system is concerned with the interconnection of the individual components so that the output of the overall system follows, as closely as possible, a desired signal, usually termed setpoint [15]. The system to be controlled is named the "Plant".

There are two types of control system: The Open Loop and Feedback or Closed Loop. In the Open Loop control system, A Reference signal is transformed and applied to the actuator subsystem of the Plant so that it has a predetermined response from it. In other words, The transformation of the reference signal is independent of the Plant's output. In Block Diagram Representation, this control method may be represented as a single pipeline.

In a Closed Loop control system, the output of the Plant is measured and compared with the reference signal. The result of this comparison is then used to drive a Controller Block, whose function is to properly transform the actuating signal in order to drive the Plant's output to the desired level, thus the output follows the desired referenced signal. The Block Diagram of this control system is shown in Figure 3.2. In case of negative feedback control, a sensor measures the output value of the variable and compares the result with the reference value (setpoint). An error term is computed and fed to the

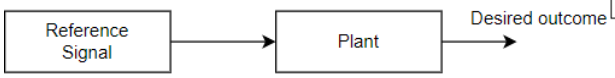

Figure 3.1: Generic Open Loop block diagram

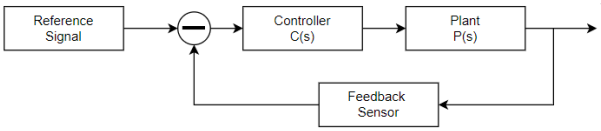

Figure 3.2: Generic Closed Loop block diagram with a feedback element and a controller

Controller block that converts it into a suitable input to the Plant which, in turn, adjusts it's output and thus match the desired value. The Close Loop control system is the appropriate architecture for the Atomic Force Microscope due to the time variant nature of the disturbances, after all the cantilever is being dragged across a surface and so, by design, the probing nature of the instrument requires a constant motorisation of the output.

In the design and analysis of control systems, the response of a subsystem (block), $y(t)$ relative to some input, $u(t)$ is given by: $Y(s) = T(s)U(s)$, where $Y(s)$, $U(s)$ are the Laplace Transforms of the output and input, respectfully, and $T(s)$ is the system's transfer function in the Laplace representation. Note that this equation is a simple product. The advantage of this representation is the simple relation between blocks along a pipeline, which results that the output relative to the input is a multiplication of the transfer functions along that path. From this property, it is possible to demonstrate that the transfer function of a feedback loop, as the one in figure 3.2, is:

$$\frac{Y(s)}{U(s)} = \frac{C(s)F(s)P(s)}{1+(C(s))F(s)P(s)} \tag{3.1}$$

where $P(s)$, $F(s)$ and $C(s)$ are the transfer functions of the plant, the feedback process and the controller, respectfully.

An important part of control design is the evaluation of the control loop's stability. Every control system must be stable, otherwise the system will malfunction, break, burn or simply oscillate with an amplitude larger than desired.

By stability it is meant that if a disturbance with a finite value is presented then the output of the control system must also have a finite magnitude, otherwise if the response grows with time it may reach unattainable values. Also, stability means that if an impulse is applied to the system, it's output must converge to 0 as time progresses. A special case of stability is when an excited system's response to an impulse is a oscillation between two finite values.

Equation 3.1 is the starting point for the stability criteria. There are a few methods, both analytical and graphical, that allow to predict the stability of the system.

In the Laplace representation, all the differential equation are transformed into polynomials. One of those methods is based on the the fact that all transfer functions are a quotient of polynomials as represented in equation 3.2. The roots of the numerator are called zeros and the roots of the denominator are called poles. As a first stability criteria, the number of zeros cannot be larger than the number of poles. Also, another stability condition is determined for the roots of the denominator. It may be shown that if the poles of the transfer function lie on the right side of the Argand-plane (s-plane) the system will have terms that do not decay with time, and so the system's response will rise to infinity. Thus, a system is stable only if every pole of it's transfer function lies on the left side of the Argand-plane, i.e their real part is negative.

$$H(s) = \frac{N(s)}{D(s)} = \frac{b_n s^{n-1} + ... + b_2 s^2 + b_1 s^1 + b_0}{a_n s^{n-1} + ... + a_2 s^2 + a_1 s^1 + a_0} \tag{3.2}$$

This method works wonders when the system has 2 or 3 or even 4 roots. It becomes incredibly complicated to compute the roots of a system with, for example, 8th or 10th order polynomials. For this reason other stability criteria can be used.

There are tools, other than analytical ones, that simplify the design of the controller process, since it may not be possible to model the system by a mathematical model and may be only inferred from measured data.

Additionally, when a controller is designed it is with some performance criteria in mind. The performance of a control system may be divided into two parts: The steady state performance and the Transient response. The steady state performance is concerned with the accuracy of the the output compared with the reference signal.

When a reference signal is applied to the system in the form of a step function with magnitude M , the final value of the output is given by:

$$y(t) = \lim_{t\to\infty} y(t) = \lim_{s\to 0} sY(s) = \lim_{s\to 0} sH(s) \times \frac{M}{s} = M\frac{b_0}{a_0} \tag{3.3}$$

The deviation from the reference value M is the position error of the system.

On the other hand, the Transient response of the system is the analysis of the system's output path until the steady state is attained. The characterization of the Transient response is generally specified in the terms of Rise Time, which is defined as the time it takes for the system to reach 90% of it's final value; the Overshoot that is the difference between the steady state value and the response's maximum value during it's come up, and the Settling Time that specifies the total time it takes for the system to be within some range of it's final value. The rise time and the overshoot specifications are usually in conflict with one another, in the sense that, in some cases optimization for one of them leads to a deterioration of the other, thus, there must be a compromise between the two.

Let us consider again the feedback equation but this time let's assume the transfer function of the feedback element is incorporated in the plant in order to show how the controller block changes the performance of the plant. The system's transfer function is now equation 3.4.

$$G(s) = \frac{C(s)P(s)}{1 + C(s)P(s)} \tag{3.4}$$

Another way of checking stability is by finding if there are values of $s = \sigma + i\omega$ that make the product $C(s)P(s)$ equal to -1. This approach requires the mapping of all the s values that belong in the left side of the Argand-plane through the $C(s)P(s)$ function. The result of this mapping is called Nyquist plot. Stability is checked if, for all values of s bounded to complex number with negative real part, does not cross the point $(-1,0)$ of the Argand-Plane.

The main point of this method, however, is not so to check stability but to measure the degree of stability of a system, by computing the distance between the Nyquist plot of $C(s)P(s)$ and the point $(-1,0)$.

Instead of measuring the distance of the curve to the $(-1,0)$ point in a Nyquist plot, a more practical method, whose criteria is equivalent, is to evaluate the Gain Margin and Phase Margin in a Bode Plot .

Given the Transfer function of a closed loop system, as shown in equation 3.4, an important point of view is obtained when s= $i\omega > 0$, i.e it is looked at the how the system reacts to a spectrum of frequencies as input.

At each frequency ω, the system will react accordingly with some amplitude and phase difference. Mathematically, the amplitude is given by $|C(i\omega)P(i\omega)|$ and the Phase by

$\arctan\left(Im(C(i\omega)P(i\omega))/Re(C(i\omega)P(i\omega))\right)$.

The Bode diagrams are plots of the amplitude vs frequency in a logarithmic scale and the phase vs frequency plotted in a linear scale for the phase and logarithmic scale for the frequency, which correspond to measurable quantities.

Since these plots take advantage of the logarithmic scale, the effect of a zero or pole has a small effect on the amplitude of the signal at frequencies much smaller than the frequency at which the pole/zero occurs. For that reason it may be approximated an asymptotic behaviour and generalized that a pole has a -20dB per decade (decade means a frequency that is a factor of 10 larger/smaller relative to the pole/zero frequency, two decades means a frequency 100 times larger/smaller and so on) on the amplitude, and a total of -90° phase shift starting from a decade prior and ending a decade later. Similarly, a zero increases the amplitude by +20dB per decade, and has a $+90^\circ$ phase shift starting from a decade prior and ending a decade later.

A feature of the Logarithmic scale is that the effect of each individual zero or pole can be obtained separately and then added to make up the whole system's frequency response.

Tying it all back to the conditional stability criteria, two special points are defined. The frequency at which the phase crosses below the -180° line is called cross-over frequency. The frequency at which the gain crosses below the 0dB line is called gain cross over frequency. It's at this points that are defined the Gain margin and Phase Margin. Gain Margin is the gain at the phase cross over frequency measured from 0dB. If at the ω_p the point in the Amplitude Plot is below the 0dB line, the gain margin is positive. And conversely, Phase Margin is the phase available at the gain cross over frequency measured from the -180 degree line. If at the ω_g, the point at the Phase plot is above the the -180 degree line, i.e the phase is greater then -180, the phase margin is positive. The stability criteria, not shown here, dictates that a system with both positive margins is stable.

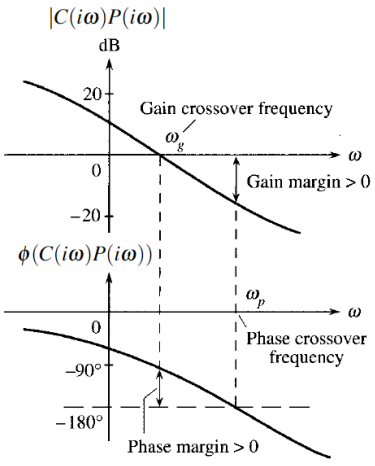

Figure 3.3: Amplitude and Phase diagrams of a stable system where is represented the gain and phase crossover frequency. Adapted from [15]

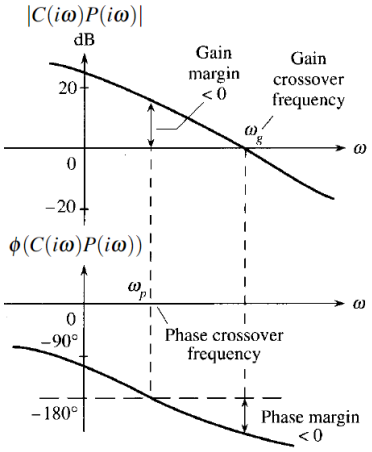

Figure 3.4: Amplitude and Phase diagrams of a unstable system. Adapted from [15]

Another way of wording this stability criteria, that may provide better intuition for this subject, is by writing $C(s)P(s)$ as their Amplitude and Phase representation :

$$C(i\omega)P(i\omega) = |C(i\omega)P(i\omega)|e^{i\arctan\left(\frac{Im(C(i\omega)P(i\omega))}{Re(C(i\omega)P(i\omega))}\right)} \tag{3.5}$$

At the phase cross-over frequency, the polarity of the signal is shifted and if at the particular frequency the amplitude of $|C(i\omega)P(i\omega)|$ is not below 0 dB, then the loop is no longer in negative feedback

and crosses into positive feedback, as can be seen by equation 3.4. In order to main stability, at that frequency the gain of the system should be smaller than 1. The same train of thought applies to the gain cross over frequency, there should be a 180° phase shift for frequencies whose gain is below 0 dB threshold.

If both phase and gain margins are positive then the system is stable.

Performance wise, gain and phase Margins do not have a direct relationship with the time domain criteria presented earlier (rise time, overshoot and settling time). Only some approximate figures may used as a rule of thumb in the design stage. This is the reason for the absolute need of a system simulation. The bandwidth of the system is related with the speed of the response and is given by the gain cross-over frequency whereas the overshoot may be related to the gain or phase Margins according to the table 3.1:

Table 3.1: Approximate relationship between the system's response Overshoot (OS) with the Gain Margin(GM) and Phase Margin (PM)

GM (dB)	PM(°)	OS
≥ 12	≥ 60	≈ 1
≥ 10	≥ 45	≈ 1.3
≥ 15	≥ 30	≈ 1.9

The performance may also evaluated using the full closed loop function of the system, i.e equation 3.4.

The steady state value is found in the Amplitude Plot at the lower frequency range. This is much like at the previous definition. The Transient response criteria has a close relationship with the resonance peak and the bandwidth of the system. The resonance peak is the maximum amplitude value reached by the frequency response. Graphically, it is determined by the largest Amplitude/Gain (amplitude and gain are used interchangeably) value. Here, the Bandwidth of the system is defined as the frequency at which the amplitude decreases to 70.7% of the steady state value. i.e $0.707G(0)$, or to put it another way, if the low frequency gain is 0dB, the cutoff frequency happens at the point where the amplitude reaches $20\log(0.707) = -3$dB.

With this criteria, it is up to the design of the controller to tune it's parameters in order to reach some Gain/Phase Margin specification.

As it was stated in the previous chapter, the PID controller is the general architecture used in AFMs. The PID is the term that defines the mathematical operation this controller applies to the error signal, as shown in equation 2.16. The Proportional block multiplies the error term by some constant amount K_p.

In the frequency domain point of view, the Proportional Gain is simply a straight horizontal line in the Amplitude plot that crosses the vertical axis at the $20\log(K_P)$ point. This means the Proportional Gains raises or lowers the Amplitude Plot of $C(s)P(s)$ in the vertical direction. And it provides no phase difference to the loop gain. The Proportional term alone is not suitable for a steady state tracking of a reference signal.

The Integral Block outputs a signal that is proportional to the integral of the error term. The integral operation continuously sums the error term as it progresses with time. The output of this stage is then the cumulative error value over time, scaled by the integral gain, K_I. A prolonged error term will be met with a increasingly higher output from this block. The Integral action is the most important term for our application since it is actively removing steady state errors in the output. In the Frequency Domain, the Integral Block is represented in the Amplitude plot as a line with -20dB slope across all the frequencies, starting at the origin of the frequency spectrum with magnitude $20\log(K_I)$. Also, in the Phase plot, it

sets a vertical offset of -90° for all frequencies.

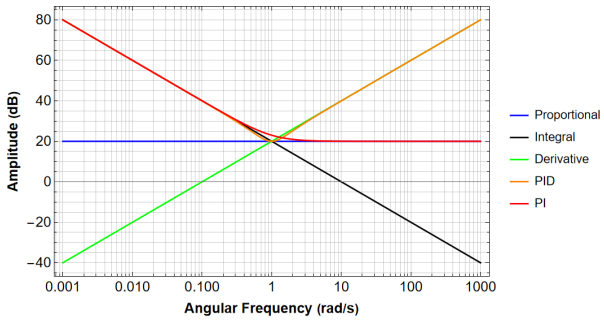

Figure 3.5: Amplitude plot of the different controller blocks. The Gains of each Block was set to 10.

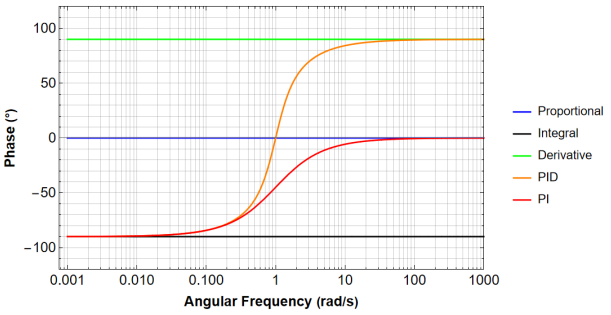

Figure 3.6: Phase difference plot of the different controller blocks. The Gains of each Block was set to 10.

In the Derivative Block, the contributor is the rate of change of the error signal. This block is sensitive to the how quickly the error term varies. In the Frequency Domain, the Derivative Block has an inverse relationship with the Integral block. In the Bode Plots, it shifts the Phase by $+90^\circ$ Amplitude graph, it is a line with $+20$ dB slope that starts, again, at K_D Hz.

The three outputs are then summed and sent to the Plant. The gains from these stages are adjusted or, more commonly, tuned in order to increase of decrease the system sensitivity to that term, and meet the Gain/Phase Margin criteria. Figures 3.5 and 3.6 show the amplitude and phase relationships when these block are combined. For lower frequencies, the response of the system is usually dominated by the Integral block, whereas in higher frequencies the derivative term takes precedence.

Ultimately, the Design of the controller must be made with the best possible model of the Plant it is trying to control.

3.1.1 Important Concepts of AFM control

In this section we will cover some parameters of the AFM that directly impact the control System or impose limitations on it.

The image formation process requires the user to input a setpoint which is linked to the desired tip-sample interaction. Most often the setpoint relates to the deflection of the tip, as in the case of contact mode, or to the vibration amplitude in the case of amplitude modulation. The control system will permanently monitor their values via the error signal, that is defined as the actual measured value and the setpoint. The error signal also quantitatively reveals how well the control system is maintaining the desired reference value.

As previously discussed, the bandwidth of the control system is directly responsible for the quickness of it's response. From the frequency domain point of view, the bandwidth of the system defines the maximum frequency it can follow a sinusoidal input. However, since the image formation process requires the tip to be moved across the surface at a certain speed the bandwidth analysis may fall short at forecasting the control system's performance. The picture gets even worse by bringing in an extra unpredictable variable which is the sample's topography. One can imagine a sample with a sinusoidal shape as topography, in this case the control loop performance is directly associated with the spacial frequency of the sample and the scanning speed. The Control loop will be excited by a sinusoidal input whose frequency is the sample's spacial frequency times the speed of the scan. The world is not ideal and so this straight forward approach is not realistic. Additionally, the full tip-sample the interaction is neither linear nor monotonic, depends on both the tip and sample characteristics which are never the

same, making it difficult to predict ideal gains or setup some type of automation of the gain setting. One day the proportional gain is helpful the next day it must be turned off.

A more realistic way of establishing the system performance is in the time domain. The Step disturbance is the most effective way to evaluate the feedback loop. This may be simulated by numerical computation but it does not take into account the effect of the scanning speed. As the cantilever comes across a step valley, the interaction with the surface decreases. The cantilever responds to this either with a decreasing deflection or an increasing vibration amplitude, Either way, the physical change it goes through will follow an exponential term with a time constant given by $Q/2\omega_0$, as was demonstrated in the previous chapter. Cantilevers have an inherently high Q which makes this reaction slow, in the millisecond order of magnitude. If the scanning is being made in fast speeds, by the time the error signal picks up a downward step, the cantilever may have traveled a significant distance and the tip will detach from surface altogether. This is called the parachuting of the tip and this effect may be seen in the pictures.

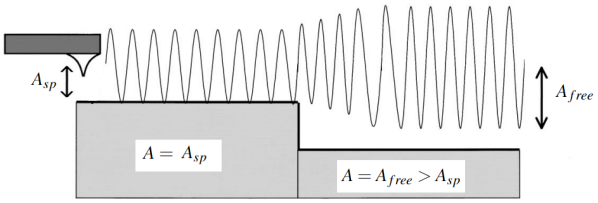

Figure 3.7: Representation of the cantilever scanning over a downward step in dynamic mode. Before the step it's vibration amplitude is equal to the Setpoint's value. It then proceeds to detach from the surface and therefore it's amplitude rises. In this case it exemplifies the extreme case where the the control system is not fast enough to prevent it reaching it's maximum value. Adapted from [12]

It is aggravated if the control system's gains are not well optimized. Also, if the set point is chosen very closely to the interaction free value, as the tip has been detached from the sample and it is no longer in contact, the control variable may in fact reach the interaction free value. When this happens the error value reaches a maximum. The error saturation causes the topography to appear linear due to the integral action, with slope related to the gains of the PID and the setpoint value [12]. This of course is an issue for very delicate surfaces and the interaction needs to be minimized. When color coding the height of the topography, the picture will show this characteristics by looking like there is a shade between the peak and the valley. When an upward step is encountered, the system will react according to it's step response. At high scanning speeds it might not be fast enough to avoid a collision wit the sample. It is desired to have a larger Integral Gain in order to increase the bandwidth of the controller. As previously mentioned, the Integrator has a large Gain for lower frequencies since it's attenuation is 20dB per decade of frequency increase starting at DC. The larger the gain the higher the bandwidth of the system and thus greater scanning speeds are achieved but also low frequency noise is amplified. However, at the phase crossover frequency, stability criteria demands the gain at this point to be negative in the logarithmic scale. At most system this will happen at the resonance frequency of a subsystem, as we will see later. The system may be brought back to stability with the Proportional gain, since it lifts the Amplitude graph vertically, but, in most situations, this action is limited.

3.2 System identification

In this section, it will be constructed a model for the components that make up the AFM. It is imperative to develop a model for the AFM in order to simulate it's dynamic response and estimate the Gains of the PID.

A detailed explanation of the constituents that make um the AFM was given in the previous chapter so that the system identification would be a much easier endeavour in this one.

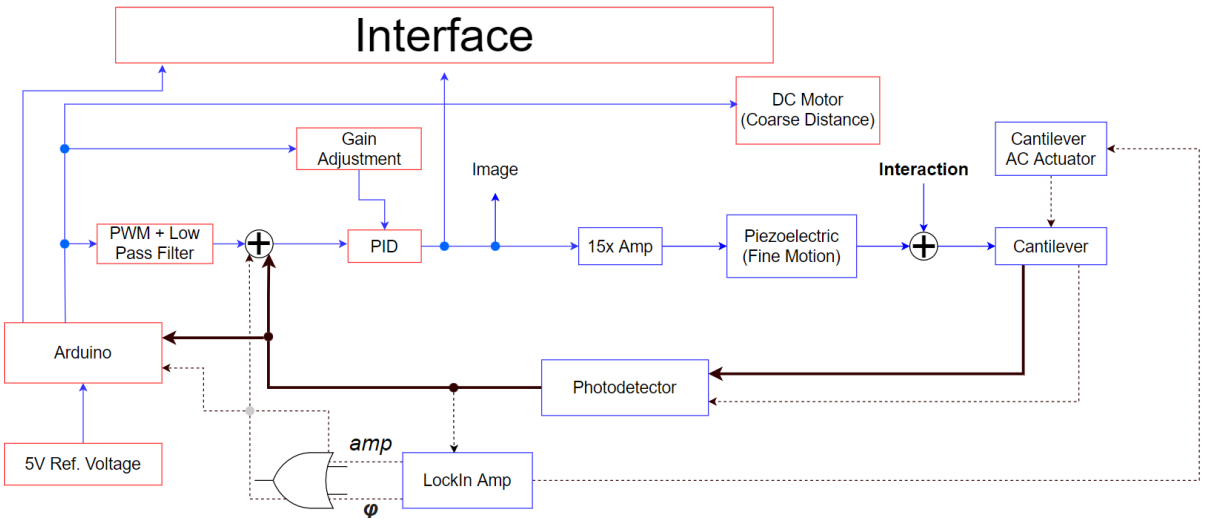

Figure 3.8: Block diagram of the whole AFM. The block in red were developed in this project. There are two paths represented in the feedback loop. The thick black line represents the contact mode path, while the dashed line is for the dynamic mode. The blue blocks are common to both modes of operation

Although the AFM is a nonlinear system, there are some components that may be approximated by linear models. The identification of the system is fundamental to understand it's dynamic behaviour in order to fruitfully establish certain performance criteria, set some estimation for the Gains of the PID or formulate some guidelines on how to improve the instrument.

This process starts at the identification of the control variable, which were clearly identified in chapter 2. In contact mode, the imaging is intended to be made with a constant interaction force between tip and sample. In this analysis, we look for the dynamic behaviour of all the actuating stages that take part in the AFM chain of operations. The overall *Plant* transfer function is obtained by the multiplication of the individual transfer functions of all the subsystems. In the case of Amplitude Modulation Mode, it must be taken into account the extra piece equipment, the *Lock-In Amplifier*.

In figure 3.8 is displayed the block diagram of the AFM. The Blocks that make up the Plant are the Cantilever, The Position Detection System, the Voltage Amplifier and the Piezoelectric Element and the *Lock-In Amplifier* in case of the Dynamic mode.

We begin the model by assuming the force and sample are at all times in contact with one another, meaning the magnitude of the interaction only is time dependent. This greatly simplifies the model since this assumption simplifies the non-linear behavior of the interaction.

Beginning at the cantilever, as previously discussed it may be modeled as an harmonic oscillator, with a undamped resonance frequency ω_0 and a quality factor Q. The Laplace Transform of this physical system wields a quadratic transfer function with a constant numerator:

$$C(s) = \frac{\omega_0^2}{s^2 + \frac{2\omega_0}{Q}s + \omega_0^2} \tag{3.6}$$

The typical cantilevers used at the laboratory have resonance frequencies specified in table 2.1, which are all in the hundreds of kHz. The quality factor is depended on the atmospheric medium the AFM is operating. The majority of the samples are probed in either air or aqueous medium, such as in the case of biological samples (cells). The Q values may have up to two orders of magnitude between these

conditions. Typical Q figures in air are about 100 while in water the values are about 5.

The bandwidth for this system is in the 10^5Hz order of magnitude. The Amplitude resonant peak may reach 40 dB higher than the low frequency value.

Now we turn our attention to the PSD system, which will be referred to as simply The Photodetector. As previously seen, the physical principle of the photodetector may be simplified schematically with a capacitance in parallel with a resistance whose signal is then amplified by and OpAmp, as is represented in figure 2.9. OpAmps are usually used in model building as ideal components, however they are real devices and so their real characteristics may matter when trying to build (somewhat) accurate models, especially when high performance is in mind. With that said, OpAmps are also a feedback loop system and have limited performance. OpAmps have a maximum open loop gain that limits the performance when a feedback loop is added to the circuitry. In their datasheet, there usually is specified the unity gain Bandwidth. The OpAmp is, like the vast majority of components, a low pass filter. With the unity gain Bandwidth information it is possible to compute how much the gain will be reduced with the Gain Bandwidth Product. It can be shown the Gain Bandwidth product for a device is approximately constant. With this it is possible to calculate how much bandwidth is lost due to gain. Of course, Operational Amplifiers are much more complex devices, but this is a good starting point. The OpAmp used in this amplifying stage is the AD713, a high bandwidth Operational Amplifier that is widely used. It's datasheet specifies a unity gain bandwidth of 4 MHz. The feedback gain obtained from it's equivalent circuit, in figure 2.9 is given by: $(R_2+R_3)/R_3 \approx 10$, thus the bandwidth of the pre-ampifier circuit is about 400 kHz

Figure 2.9 may be thought of as two individual low pass filters in series, hence the overall Transfer function for this Block is the multiplication of the to stages:

$$PH(s) = \frac{\frac{1}{R_{in}C_{ph}}}{s+\frac{1}{R_{in}C_{ph}}} \frac{1+\frac{R_2}{R_3}}{\frac{s}{\frac{G_0\omega_0}{2\pi\left(1+\frac{R_2}{R_3}\right)}}+1} \tag{3.7}$$

where $G_0\omega_0$ is the gain bandwidth product of the OpAmp in Hz

The low pass filter that is composed of the equivalent capacitance of the PIN junction $C_{ph} = 20$ pF, provided by the manufacturer, and the current-to-voltage conversion resistor R_i have a bandwidth of about 3 Mhz. The total Bandwidth of the system is limited by the lowest cut off frequency in this subsystem, as can easily be seen by the Logarithmic nature of the Amplitude Graphs. The results was confirmed via *PSpice* simulations of the Amplifier Stage that yielded a cut off frequency of about 320 kHz.

Next we shall cover the Amplifier that drives the Piezoelectric Element. It is known *a priori* that it has a fixed $\times 15$ Gain, which translates to $20\log(15) = 23.53$ dB. From this simple description alone no information about it's dynamic performance is revealed. As all real devices it must have some dynamic equivalent model. This information was available at the manufacturers website, which specified this particular model to be a second order Bessel low pass filter with 2 kHz bandwidth.

This type of filter has special characteristic Transfer Function, it's denominator is a Bessel polynomial and, given that the filter is second order it is deduced that the polynomial is quadratic, hence: s^2+3s+3. The roots of this polynomial are $s = -3/2 - i\sqrt{3}/2$ and $s = -3/2 + i\sqrt{3}/2$. These are the poles of the filter.

We start with the transfer function:

$$B(s) = \frac{k}{s^2 + 3s + 3} \tag{3.8}$$

where k is, for now, some arbitrary number that sets the 0 Hz gain. If we set $k = 3$ we get the normalized Transfer Function, due to having $B(0) = 3/3 = 1 \rightarrow 0$ dB. In this form, the cutoff frequency is located at $\omega_n = \sqrt{3}$ Hz. This is just a baseline form in order to attain the actual system Transfer function. And, to do just that, it is needed to scale the function in magnitude and, most importantly, in frequency.

The cutoff frequency shifting is made by the substitution of the s variable into s/ω_n, where ω_n symbolizes the new cutoff frequency that matches the specification: $\omega_n = 2\pi f$, where f is the cutoff frequency of 2kHz. By making the substitution and rearranging the terms in order to have the same representation as equation 3.8, it is reached:

$$B(s) = \frac{3\omega_n^2}{s^2 + 3\omega_n s + 3\omega_n^2} \tag{3.9}$$

All that's left to do is adjust the Low frequency gain so that it matches the times 15 amplification specified by the manufacturer: $B(0) = A3\omega_n^2/3\omega_n^2 = 15 \rightarrow A = 15$.

The Transfer Function of the Piezoelectric Amplifier is :

$$B(s) = \frac{15 \times 48 \times 10^6 \pi^2}{s^2 + 12 \times 10^3 \pi s + 48 \times 10^6 \pi^2} \tag{3.10}$$

The final subsystem left to analyse is the Piezoelectric element.

The usage of Piezoelectric stacks in micropositioning is to apply a force unto a an object and displace it. A voltage is applied to the piezoelectric element and it will expand and due to the mechanical contact with the Scanner it will push on it. The force generation by the piezo is found to be linear with with the amount it has expanded, the piezo's stroke. But it is dependent on the the load it is actuating. If the load exerts a constant force back, the piezo exerts a net force, and the maximum stroke is observed. With a load that also has a linear force with the displacement, the full extension range of the piezo may be reduced.

The voltage induced displacement announced by the piezoelectric manufacturer, is defined as the piezo's free stroke, which is measured in the absence of any applied force. It may be also defined another important parameter that is the maximum force the piezo is able to generate. The manufacturer determines this value by fully expanding it without any force and then applies an increasing load until the piezo compresses back to its original length [27]. With this data points, it it then defined the piezo's stiffness:

$$k_P = \frac{F_{max}}{\Delta L} \tag{3.11}$$

It is an important parameter for characterizing the full-system behavior, including it's resonant frequency.

The piezo operation is very similar to that of an harmonic oscillator whose dynamic characteristics have been previously described. Take into consideration that this is just a simple ideal model, piezoelectric elements are solid bodies and have inherent non-linearity's whose treatment is outside the scope of this text. This harmonic oscillator would have a 115 kHz resonant frequency.

Now taking into account all the subsystems previously described, we note that the lowest bandwidth elements is the Piezoeletric Amplifier with about 2 kHz. If one would swap the Amplifier for one with better specifications, that certainly exist in the market, the prototype would compete with the most

modern AFMs, discribed in the literature, that have video rate image generation [7].

With the setup described, the line imaging rate does not exceed 2 lines/second with a well tuned PID

With some AFM operating experience, one would suspect that there is some other limiting factor is being overlooked.

What was left out of the discussion was the Scanner, the structure that houses the piezoelectric element stack and counterbalances the piezo's displacement when its expansion is reduced. This piece is fundamental in the AFM performance and it is subject to a lot of study in the literature.

The prototype's vertical scanner is a simple table that is screwed into the XY scanner piece, as can be seen from image 2.23. When the piezo expands it pushes the table top upwards. The table top is attached to the 4 screws that held it in place via the beams on the side of picture 2.23. In combination with the piezoelectric stack, this system can be conceptually broken down as two spring systems in parallel, given that the screws are have a stiffness infinitely larger than table top.

The Equation of motion of this system, in the Laplace Transform are given by :

$$m_P X_P s^2 = -k_P(X_P - X_0) - K_{SH}(X_P - X_0) \tag{3.12}$$

where k_P stiffness of the piezo and k_{SH} is the spring constant of the sample holder, which is unknown, $(X_P - X_0)$ the piezoelectric displacement, m_P it's equivalent mass that can be computer in the technical document of the manufacturer [27].

An attempt at estimating the sample holder spring constant leads, again, to the Euler-Bernoulli's beam Theory, equation 2.4. The geometry of this particular system is different than the cantilever one. This beam is clamped on both ends, and so the boundary conditions needed to reach the stiffness expression are the following:

$$\begin{cases} \frac{d^4u(x)}{dx^4} = 0 \\ \left.\frac{d^3u(x)}{dx^3}\right|_{x=L} = -\frac{F}{EI} \\ \left.\frac{d^2u(x)}{dx^2}\right|_{x=L} = 0 \\ u(0) = 0 \end{cases} \tag{3.13}$$

The stiffness of this particular beam configuration is obtained by:

$$k_{beam} = \frac{12EI}{L^3} \tag{3.14}$$

where I is the area moment of Inertia of a rectangular section $wt^3/12$, L the length and E is the Young's Modulus.

The thickness, width and length of the structure we measured to be 3 mm, 3 mm and 12 mm, respectfully, whereas the Young's Modulus is material specific. This piece, much like other components of the prototype, is made out of 3D printed plastic with E= 4.9×10^9 Pa. Since the structure has 4 supporting beams, they are springs with parallel configuration and so the equivalent stiffness is given by the sum of all four individual stiffness. It was obtained $k_{SH} = 8.16 \times 10^5$ N/m.

It was also measured the weight of the whole piece $(1,028 \pm 0.001$ g) and thus, the simulated structure yields a bandwidth of about 10 kHz.

Upon further consideration, it was decided to measure the frequency response of this stage. For this task it was used the AFM, with the aid of a *Lock-In Amplifier*, to perform the measurement.

The setup was as follows: the cantilever is approached into contact with the sample holder's surface.

The PID controller is turned off. The piezoelectric element is directly excited with the Lock-in reference signal (5V peak-to-peak sinusoidal signal), bypassing the limited bandwidth of the amplifier, 5V is small in comparison with the full range of the piezo, hence the displacement will be small. The PSD system will record the oscillation of the cantilever, whose signal is fed to the Lock in. The measurement is made possible due to the cantilever and photodetector response being flat for lower than their cutoff frequencies, which are located at the hundreds of kHz. The Lock-in Amplifier's reference frequency is swept across a desired interval and a recording of the sample holder response is made.

The frequency response measurement yielded the following Bode Diagrams:

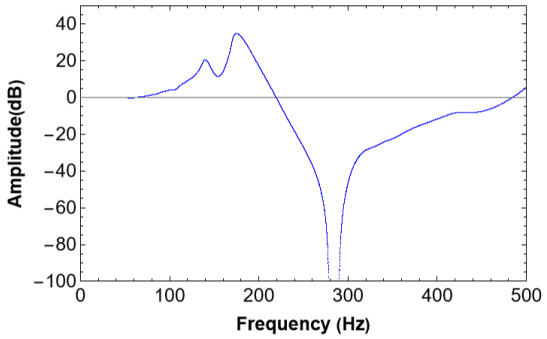

Figure 3.9: Amplitude plot resultant from the measured frequency response of the vertical direction Scanner. The frequency axis is in linear scale

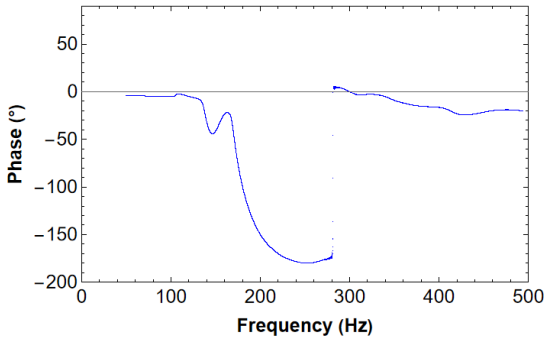

Figure 3.10: Phase difference plot resultant from the measured frequency response of the vertical direction Scanner. The frequency axis is in linear scale

It was found that the bandwidth of the vertical Scanner stage is only about 200 Hz. This is the Bandwidth limiting subsystem of the Whole AFM apparatus. At 200Hz, it may be disregarded the piezo's dynamic response.

From graphs 3.9 and 3.10, it was built a second order polynomial model with equation 3.6 as a template, were the natural frequency and Q value was adjusted to the experimental data.

3.3 PI tuning

With all the transfer functions modeling the system, the next step was to compute the value of the PID gains given certain performance parameters in mind. Several sources advise caution with the Differentiator term. Although the Differentiator term would compensate for the phase delay of the Integrator, it it is highly prone to exacerbate the resonant peaks present in the system [7]. Additionally, it amplifies the high frequency noise that is certainly present in the optical system [1]. Furthermore, in the practical setup in the lab, it was common knowledge to avoid the Differentiator block in the controller, thus, it was eliminated from the design.

Figure 3.12 shows the bode diagram of the loop gain $C(i\omega)P(i\omega)$ where the controller is a PI. Notice that, regardless of the gains, the overall Plant's transfer function is dominated by the Scanner's harmonic oscillator behaviour, which, at it's resonant frequency exhibits a -180° phase shift. Consequently, the Phase cross over frequency is at that location. In order to maintain the system's stability, the system's Amplitude response should be below the 0 dB line and so the Gain cross over frequency must be below the resonant frequency. This implies that the system always has a Phase Margin of 90° and so the Gain margin is the degree of freedom by which the design's rule should be adjusted. It was computed the gains of the PI in order to provide a Gain Margin of about 10 dB, which yielded P$\approx 10^{-3}$, I$\approx$ 500 Hz. These values are only a starting point. After, the gains were tuned via the Bode diagrams and the step

response simulation. This tuning iteration process was done solely as an academic exercise because the meaningful tuning of the PI was done experimentally.

With the estimation of the value of the gain, it was prototyped the controller on a breadboard. Special attention was given to the integral gain which is the essential component of the controller for our application. The Integrator architecture is a simple Miller Integrator made with an OpAmp, a resistor, a potentiometer and a capacitor, as shown in figure 4.17. It was chosen a 200 nF capacitor, a 1 kΩ resistor and a 10 kΩ potentiometer in order to manually tune the gain.

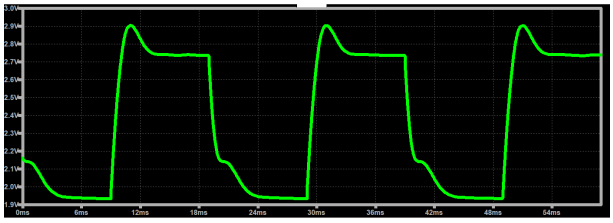

Figure 3.11: Step response simulation of the Control system using the Pspice software.

The tuning process was carried out in the same way as it would be done in the regular use of the instrument: it is started an imaging process at a certain line scan velocity and the potentiometer is set to it's maximum value. Then, if the system is stable, the topography outcome is monitored via the software. Subsequently it is adjusted the knob of the potentiometer in order to increase the integral gain and notice the impact on the topography signal (excessive noise or excessive overshoot). Finally, it is tweaked the scanning velocity in order to limit test the stability of the system and the faithful reproduction of the topography.

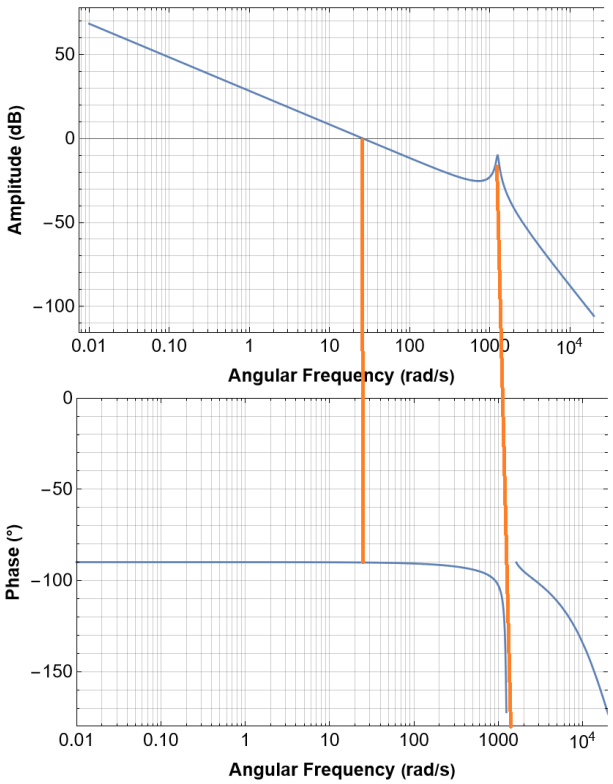

Figure 3.12: Bode Diagram of the Loop gain. It is highlighted the Gain and Phase cutoff frequency. The software is having trouble computing the phase difference of the system, nonetheless, the resonant frequency of the Scanner dominates the frequency response of the system. At the resonant frequency the system exhibits a -180 phase shift.

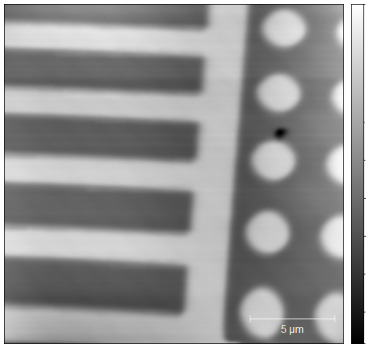

Figure 3.13: Image acquired with the analog PI at the end of the project. Image of a calibration sample. The feature's depth is 112 nm in a 20×20 μm scan.

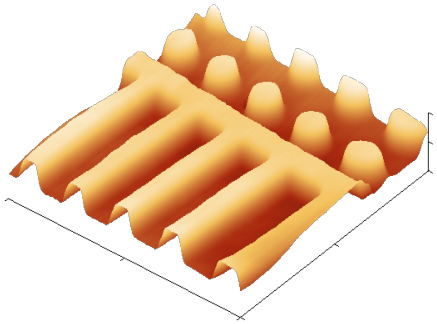

Figure 3.14: 3D render of the image in figure 3.13.

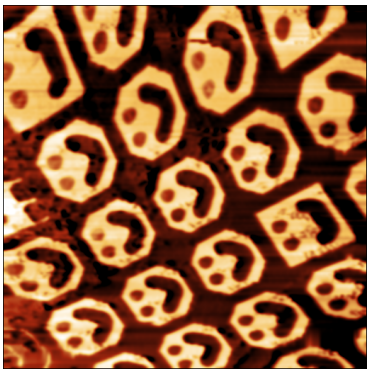

Figure 3.15: Image acquired with the analog PI in the prototype phase with a different scanner. The image has 80×80 μm dimensions. The depth of the features is 100 nm.

3.4 Why Analog?

Most commercial Scanning Probe Microscopes use digital signal processor-based PID feedback controllers [8]. These devices require an Analog-to-Digital and Digital-to-Analog Converters built in the the overall architecture. These conversion operations, as well as the actual signal processing computations, add additional delay that further limit the imaging speeds. In such implementations, the signal requires to be sampled and quantized by the ADC before being sent to the processor unit. In order for the control system to be reliable, the sampling should be 10-20 times higher than the system's closed loop bandwidth [8]. Also, before the sample stage, the signal must be filtered in order to reduce aliasing from high frequency signals. Then the digital signal must be delivered to the processor unit. If this digital signal process is not embedded into a single chip, the communication process between the peripherals (ADC and DAC) further adds delay. The processor operates on the data, according to the instructions coded in the program, which should be optimized for performance. And finally, the DAC converts the resultant control value back to a voltage in order to be applied to the Plant.

There are a lot of step in this process and that makes the design of the digital controller much more delicate, when compared to an Analog solution. Besides speed, there are inherent quantization errors induced by the ADC and DAC that may only be reduced by a high precision version of these circuits that not only add cost but also complexity to the design.

On the other hand, Analog PID controllers, by nature do not require sampling and they "simply do" the signal mathematical transformation, on a wider bandwidth and with no quantification noise.

In the rest of the discussion of this work, the trade offs between analog and digital will be further cemented and the conclusion that, besides providing a built in digital tuning of the PID Gains, that this work provides a workaround, the digital version requires a much refined design set of skills that the author was initially devoid of.

Chapter 4

Improvements and their Practical Integration into the Instrument

4.1 Coarse Motion

The Fine motion is mediated via the control system and executed by the stretching of the piezoelectric element, providing very fine motion. A coarse motion is also necessary. The coarse motion serves the purpose of a macroscopic adjustment of the tip to sample distance. This distance must be smaller than the range of the piezoelectric element expansion. It is a vital mechanism for the AFM operation as it provides a mechanism to not only engage contact between tip and sample but also separate them a wide distance that leaves the instrument in a safe state so that the user does not accidentally damage them while handling it.

As describes in section 2.3.1, the coarse motion in the prototype is preformed by turning micrometric screws that raise or lower the Head's aluminium plate. This plate is triangular shaped and has three points of contact, one contact point for each screw. Turning one of the screws is sufficient for the cantilever to be lowered, due to the geometry of the piece. Additionally there is a motor coupled to one of the screws via a rubber band.

Previous to this work, the prototype lacked both a reliable mechanism to have a uniform step action and an automatic approach mechanism that eliminates the user from this delicate process.

The following describes both the motor step uniformization process and the Auto-Approach algorithm.

4.1.1 Motor Study

As previously mentioned, the DC motor is coupled to the micrometric screw via a transmission belt (rubber band). The motor itself was salvaged from an old DVD player, whose original function was to open and close the DVD tray in a smooth movement. Although these motors are advertised as high rotations per minute, once a load is applied they have their angular velocity decreased.

In this new application, the motor will, mostly, be triggered via short burst of current in order to simulate a step. The triggering is performed by the application of a digital signal to the H-Bridge circuit as described in 2.3.1,with which it's easy to trigger with millisecond time precision.

The idea here is to study the relationship between the duration of a ON state digital signal, which generates a current pulse that flows through the motor, and the vertical displacement from the micrometric screw. And then, try to establish a short enough time duration that the motor is able to faithfully

reproduce a step with a fixed vertical length. At very short pulses the motor may not exert a sufficient torque to overcome the rotational inertia of the screw.

No theoretical models were developed for the motor + H-Bridge circuit. It was used the AFM to carry out the measurements of the steps. The design of the experiment was straightforward: The tip and some dummy sample's surface are placed at a distance such that the piezoelectric element is almost fully extended; with the help of the Arduino, the motor performs the step and reduces the tip-sample distance and, consequently, the control system reacts accordingly and reduce the applied voltage to the piezoelectric element. By multiplying the difference between the applied voltage before a step and after a step with the calibration constant the displacement in meter may be computed.

Let's denote a trial the cycle of approaching the tip and sample from the almost fully extended state of the piezo to it's voltage free length and then inverting the motion back until it's extended state.

Given the setup, the experiment is limited to step sizes that do not exceed half of the piezo range $\approx 7.7\ \mu$m.

The measurements were conducted with different "time steps" with increasingly bigger magnitude. It was found that time intervals of 5 ms or less do not produce to any meaningful vertical distance, i.e the the motor does not turn the screw with time intervals of less than 5 ms.

In figures 4.1 and 4.2 it is shown some trials with 10 and 40 ms time intervals.

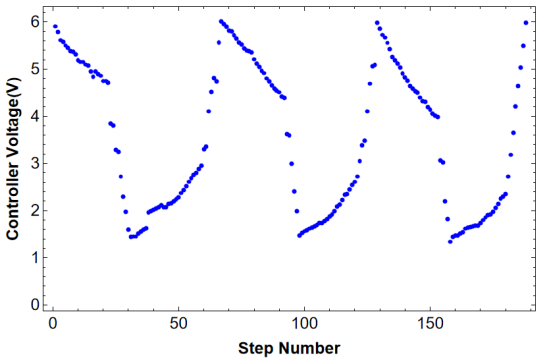

Figure 4.1: Series of trials with 10 ms time duration step sizes.

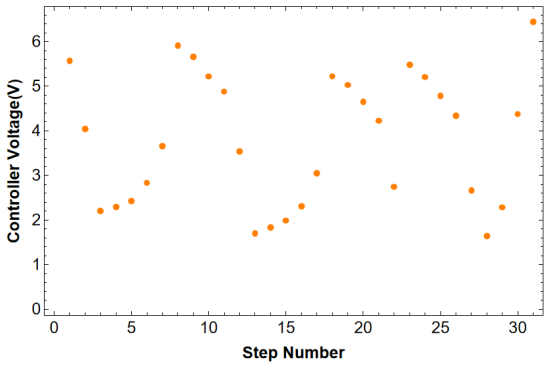

Figure 4.2: Series of trials with 20 and 40 ms time duration step sizes.

Taking a closer look at the graph 4.1, which maps the piezoelectric voltage as a function of a step increment over 3 trials and it can be seen the a uniform (systematic) behavior.

In the 10 ms step trial, it is evident some interesting behaviour. There are two distinct step regimes. One that occurs at the start of each trial, either with the piezo voltage at the maximum or at the minimum, where the points are very densely packed. And the other, occurring after the first regime, produces larger, but still uniform, step sizes. The first regime occurs at the inversion of the movement's direction and is due to the elastic rubber band compressed, whereas the other regime appears when it is fully stretched. In the 40 ms trials the different regimes are much less noticeable as the motor is performing larger steps.

With the data from the trials, the absolute distance for each step was computed and it was build histograms 4.3 and 4.4 in order to better extract the distance traveled per step. From them it is possible to infer the higher count bins are representative of the steps taken when the rubber band is not fully stretched. The highlighted region provides information about the distance traveled per step when the band is fully stretched. It is possible to estimate the an average value of the step as well as it's standard deviation. We are interested in finding a step value that is a fraction of the range of the piezoelectric element thus, both of these steps satisfy this condition with a distance traveled of about $0.5 \pm 0.15\ \mu$m

and 1.7 ± 0.6 μm for the 10 ms and 40 ms step, correspondingly.

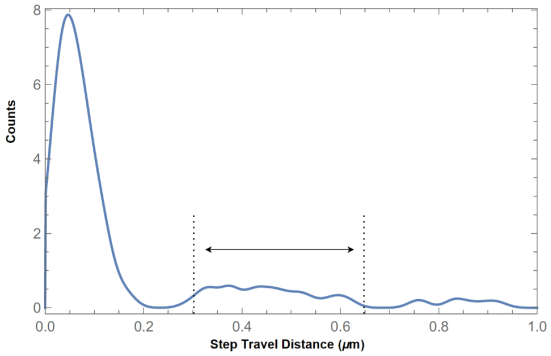

Figure 4.3: Continuous histogram obtained from the step size variation calculated from the trials is figure 4.1. The counts are in arbitrary units. It was used the Mathematica software to build this graph.

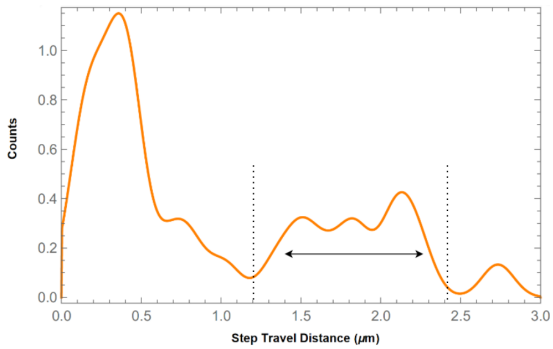

Figure 4.4: Continuous histogram obtained from the step size variation calculated from the trials is figure 4.2. The counts are in arbitrary units. It was used the Mathematica software to build this graph.

4.1.2 Auto-Approach

The Auto-Approach feature is a much needed and important improvement in the AFM prototype. It simply consists of an automatic process that approaches the tip and sample. It's implementation requires a compromise between safety and swiftness.

If the Auto-Approach feature was absent from the instrument, the user would need to perform the approach manually, i.e turning, very gently, the screw himself or endlessly pressing a button. A human can quickly make the tip travel macroscopic distances and place tip and sample very close to its eye space resolution. He can cover this distance in a matter of a few seconds with a continuous rotation, however he eventually must slow down and monitor the deflection signal as fast as the human reaction time allows, and at this stage an experience user knows that he must make small steps or turn the screw very slowly. Given that it is hard to maintain a constant rotation step he may opt to not overshoot it and take even more time. This is not nice experience, it is tiresome and makes the act of making measurements with the AFM even more unpleasant. If the user needs to switch samples frequently throughout the day, he/she needs a lot of mental fortitude and nerves of steel because, otherwise, there is a seriously high probability that the user starts to neglect the approach and, eventually, crashes the tip unto the sample.

Having an automated process to perform this task is a safer and more reliable approach and, in the case where the distance between the surfaces is only a few millimeters apart the procedure is incredibly expedite, even at a constant small step. To elucidate further on this matter, an example is provided:

Suppose a new sample is to be loaded into the AFM and it has an unusually large height profile. The user, as good practices dictate, before inserting the sample into the sample holder, retracts the tip a macroscopic distance of, for instance, about 5 cm so that there is a lot of head room. He/She lifts the aluminum plate, installs the sample onto the Scanner and mounts the microscopes head back in it's place. The user takes another look at the tip and sample profile and he can estimate that there are about 2 cm left between them. Since he/she can clearly see a macroscopic gap there are two options, either manually approximate them further until the distance is barely distinguishable and then start the auto-approach feature or, just start and let the automatic process complete the task. Out of laziness, the user decides to proceed with the second choice. Taking into account that the large step makes the tip travel, on average, 1μm it will take about $2 \times 10^{-2}/1 \times 10^{-6} = 20000$ steps in order to reach contact. The step takes 40 ms to conclude and before another step can be performed the signal related to the position of the

cantilever needs to be sampled and compared, assuming, of course the need for a digital representation. However, before the signal is sampled the Cantilever must be left idle in order to recover from the applied step stimuli. The recovery takes about $5 \times Q/2\omega_0 \approx 30$ ms in air. Without proof, for now, let's also assume the Analog to Digital conversion takes about 100 μs per sample and also we sample the signal 100 times in order to have definite certainties about the measurement. In total, the process will take about $(0.1 \times 100 + 30 + 40) \times 10^{-3} \times 20000/60 = 26,7$ minutes, which is absurdly high for a task to complete. The first choice would yield much better results as the distance between the two object is still, in the length scales relevant to the AFM, colossal, say 2 mm. In this case the amount of time needed to complete the task is lowered by an order of magnitude, which corresponds to about 2 minutes, a more admissible time interval.

A trivial, yet robust, procedure to perform the Auto-Approach would be to iterate between taking a step and expanding the piezoelectric element to check whether the desired interaction level is within reach and then, if it isn't, retracting the piezo and repeating the process. The step size needs to be appropriately sized to be about half the piezoelectric element's range. This is the basic implementation of most auto-approach algorithms and, although it results in a safe approach, is not very time efficient as the piezo's stretching process is time consuming.

After having some experience with manual approaches, it is apparent that, while in dynamic mode, the Amplitude signal sometimes is decreased as the approach is in an intermediate stage. This phenomenon occurs at distances greater than the interaction length (some few thousand μm) and it signals the progressive approximation of the two surfaces [23]. The phenomenon is called viscous cavity damping and it occurs due to a decrease in the cavity's length formed by the gap between the cantilever and the sample. As the cantilever vibrates in a viscous fluid, the fluid offers resistance to the beam displacement and as it is vibrating close to a solid surface, the behaviour of the fluid and, therefore, the cantilever are modified the the surface due to confinement [9]. Phenomenologically, there is a decrease in Q factor as a consequence of the two surfaces becoming closer and closer and so, while in amplitude modulation mode, the amplitude of vibration is decrease and thus, this phenomenon can be used to sense and signal the vicinity between tip and sample .

In order to leverage this behaviour, an improved Auto-Approach algorithm was designed. The main idea is to have two vertical step sizes, one big step of about $1,7$ μm and a small one of about 0.5 μm. It is worth mentioning that this algorithm also circumvents the stretch and recall stage of the robust alternative, since with an analog control system this operation is not so straight forward to perform and so in out algorithm the piezoelectric element is always fully engaged until the approach is reached. The algorithm developed is described in the following:

At the start of the Auto-Approach, a first set of 100 consecutive measurements are taken in order to pinpoint whether the setpoint is in reach and avoid missteps. If the user did not accidentally pressed the Auto-Approach button, it carries on with the approach by saving the value of the Amplitude at this location. Then proceeds to perform a small step and then, after waiting for the system to reach the steady state (relaxation time) another set of samples are taken and it is computed the variation in the Amplitude signal. If the value obtained is small enough to considered part of the random fluctuations, i.e a few milivolts, we may assume the tip and sample are still a great distance apart and so the step size is enlarged to the big step for the next iteration. As previously mentioned, however, if the tip is within the sample's vicinity, the variation will be meaningfully, i.e the amplitude signal will have decreased some few hundred mV, and this means that we are close enough to the sample and need to proceed with caution, i.e maintain the same small step until the step point is reached. If it is determined that the big step is suitable then the cycle is repeated by saving the voltage before a step and after a big step is taken,

until a meaningful amplitude variation is detected. This ensures a faster approach in dynamic mode while preventing damage to the tip or sample.

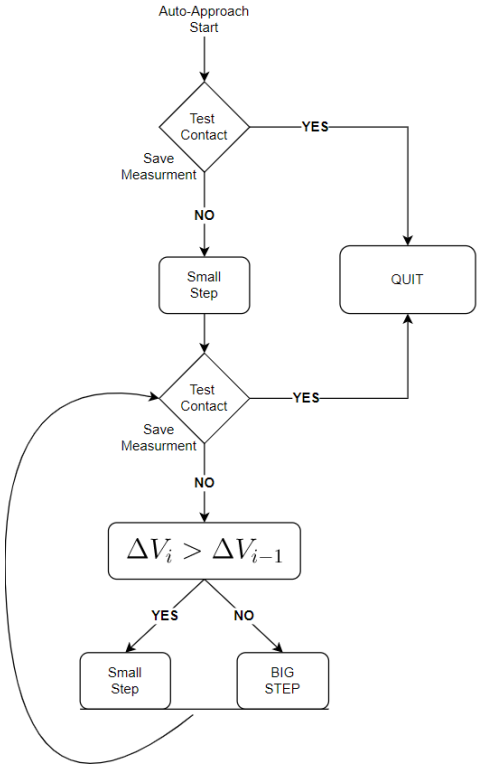

Figure 4.5: Auto Approach algorithm implemented on the AFM prototype. The constant monitoring of the Setpoint's value before and after a step, in conjunction with an appropriate step size, ensures a safe approach.

The trigger to transition from big step to small step was empirically chosen to be about 15% decrease in between steps.

It needs to be stressed that this algorithm is only effective for dynamic mode. In contact mode, the proximity sensing won't work and so the algorithm is broken down to a cumbersome step and test contact iteration that, although quite rudimentary, it really isn't an issue in this mode, since in this mode are used softer (lower spring constants) cantilevers, hence when contact is established the sample won't be under stress.

The big step was chosen to be the 40 ms time interval, corresponding to $1,7 \pm 0.6$ μm, whereas the small step is the 10 ms one (0.5 ± 0.15 μm).

4.2 Z-Axis Control Integration via Arduino

In pursuit of the proposed objectives' completion there was an apparent need to somehow conjugate the ability to perform various tasks, collect and display information resorting to the computer as a user interface.

The specific tasks are : allow the user to tune gains of the PI controller and select a desired setpoint from the computer; implement the algorithm that automatically brings the tip and sample into contact are and give the user tools to interact with the coarse motion and, lastly, visually support the user with the PSD system alignment procedure.

Aside from the gain adjustment mechanism of the PI, the other, when broken down in simple terms, are simple read/write voltages and display them on a computer screen with some program acting as interface. One might argue that there already exists a a data acquisition system that could suits these needs, the *National Instruments chassis* cDAQ-9177 with an ADC and DAC modules. Good engineering practices aim to employ the right tool for the right job and as we shall discuss and there are some critical aspects that scream this is not the proper solution.

As stated previously, each channel on either module can read or write a voltage within the standard -10 to 10V range with 16 bit resolution. This means the modules are capable of specifying a voltage down to $\frac{|-10-10|}{2^{16}} = 0.3mV$ increments. Also, they come with a calibration certificate from NIST.

In the context under which the signals are going to be used, neither the resolution of the equipment nor it's sample rate are appropriate for the task. On one hand, it's specifications are not good enough to be used as a potential digital PID controller (it is worth remembering that the PI is analogue and the arduino will be responsible for the digital gain adjustments, as will be described later on) and on the other hand, it is overqualified for the rest of the tasks. Let us go over some arguments to further elucidate on this matter.

The user first starts the alignment procedure before approaching the tip and sample. When the user is aligning the PSD system he/she will try to to reach a voltage near zero volts. However, in a practical sense, it is only worthwhile to adjust the signals near the zero volt mark. As long as the photodetector's amplifier is in the linear region and has some tolerance for when contact is established, the user does not have to fastidiously set it perfectly to zero, thus the resolution with which this voltages are read do not need to be stupendous.

Although having a good resolution in the setpoint value may be perceived as a useful tool to have, as it opens the possibility to tweak the interaction force with very precision, in practice this corresponds to a negligible variations of the interaction force, hence it is redundant. Typically one chooses as a setpoint 0.7 the value of the interaction free value.

As for the Sample Rate, the manufacturer specifies up to 100 000 samples per second for either read or write operations. In the visual support task, to put it in simple terms, the human reaction time is, on average, 0.1 seconds. While adjusting the position of the laser spot, the users only requires the displayed value to be refreshed, at maximum, 10 samples per second to perform the task within their best capabilities. At the ADC's full sampling rate, there could an excess of $10k$ samples being acquired that will go unused and even if the displayed value is averaged with 100 samples there is still an excess of two orders of magnitude.

As for the setpoint functionality, the sample rate is often not a problem since it rarely is changed mid-imaging. Even if it were, the only appreciable feature would be that, with a typical line scan speed of about 1 second, the new value of the set point would settle in $1/100000 \times 100 = 10^{-3}\%$ of a line, which is quite unnoticeable.

For the approach feature, the picture was already painted in the Auto-Approach section. There are essentially three time intervals that define a step cycle: the time duration of the step, the cantilever's settle down time and the ADC's sampling acquisition time. The longer step takes 40 ms to complete, then it is needed to wait about 5 time constants ($= Q/2\omega_0$) in order to reach $1 - e^{-5} = 0.999$ of its steady state, estimating this time with the longest cantilever, which has the lowest natural frequency, and a typical quality factor in air for about $100 \rightarrow 10 \times 100/(150 \times 10^3) \approx 7$ ms and it may be disregarded the ADC's negligible time to complete 100 samples. This limits the useful bandwidth of the DAC to output a voltage to control the motor to about $1/(46 \times 10^{-3}) \approx 22$ samples per second which is 4500 times less than the full capabilities of the available device.

Lastly, each module is limited to 4 channels. With the current setup, the *DAC* module has 3 out of those 4 being used as a voltage source to the *x*, *y* and *z* piezoelectric Amplifiers, the one applied to the *z* direction is exclusively used for the force spectroscopy mode. With the intended added functionalities, there is the need to have at least 2 more channels, one for the setpoint voltage and another to actuate the motor circuitry, which would imply the acquisition of another module. In the ADC case, all the 4 channels already being used, one for each photodetector output, Vertical and Horizontal signal, another for the Amplitude signal and the forth to read the output of the PI controller and no more signals could be added. More importantly, the way the National Instruments software is implemented, there can only be one instance of the API in usage at all times, this effectively means that only one program can control the *chassis* at a time. From experience, it is undesirable to have a single program clustered with features buried in the software as this make it feel unintuitive and with some information hidden by default. With this in mind, the design of the prototype's software was conceived to be modular, meaning that there is a program for each critical aspect of the instrument and the user could easily have them operating in parallel. This work focuses on the controller part of the software. Additionally, with this typo of modular architecture, the controller can be seamlessly integrated into other instruments, such as the force feedback microscope present at the laboratory [23].

From this outline, it can be concluded that there is no need to use this high performance, high cost device for these tasks. The remaining available channels could be used for much resolution/timing critical operations.

From a technology point of view, most, if not all, requirements for the project, are available in a single chip, a microcontroller. A microcontroller can be thought of as a computer. Within the chip is embedded a processor (CPU), programmable memory, whether in the form of cache, RAM or ROM, Input/Output physical pins, WiFi, USB protocol, ADC, DAC and many more peripherals. The only computer traditional feature that is not included is Graphical Unit Processor. Unlike a computer, microcontrollers are employed to handle a finite number of events that occur with some periodicity or based on some external input. They're programs are generally static, meaning they are only programmed once. It is through these devices that the practical implementation of our project will be made possible and, conveniently, there was already an Arduino available at the laboratory.

4.2.1 Arduino Overview

The Arduino is a platform that goes beyond the simple description of microcontroller. It has a few additional hardware components, it provides learning material and, most importantly, it provides software support. The software comes in the form of a seamless programmer that allows the upload the of the program script via USB connection, an intuitive environment to write code in C/C++ language and also many libraries that abstract away protocols and direct hardware manipulation that the user can resort as simple function calls. All these features follow the open-source philosophy.

They offer different boards with different layouts/ microcontroller units (MCU). The one that this works is based on is the Arduino Uno Rev3 Board, which from now on will simply be referred to as *Arduino*.

At the core of the Arduino is the Atmega 328p 8-bit microcontroller in a PDIP package. The 8-bit term referrers to the length of it's operating registers and the number of bits used to store a variable. This particular microcontroller operates under a 5 V Logic level and it features 32k bytes of programmable flash program memory, minus 2k used by the Arduino's bootloader - this sets the limit size of the programm running inside it- 23 programmable input/output pins which may have alternate functions, 6 of

which are exposed to the Analog to Digital converter circuitry and 6 other that support Pulse Width Modulation (PWM). It has USART(Universal Synchronous Asynchronous Receiver Transmitter), I^2C and SPI (Serial Peripheral Interface) hardware protocol interfaces. These short length communication protocols that may be used to transfer data with other devices. It also contains Timers and Counters and a Watchdog Timer and features Interrupts.

In order to ascertain if the Arduino Uno Rev3 is a device/platform appropriate to implement into the project a more in-depth look at the both the microcontroller's feature specifications and board layout is needed. This analysis was made possible by consulting the Atmega 328p datasheet and the various source files from the Arduino Board that are easily available for consult since they are an open source project.

Stating with the ADC, figure 4.6 is the schematic of the whole circuitry. The ADC itself is a successive approximation analog to digital converter (SAR ADC) with 10-bit resolution as it is highlighted in a black box in figure 4.6.

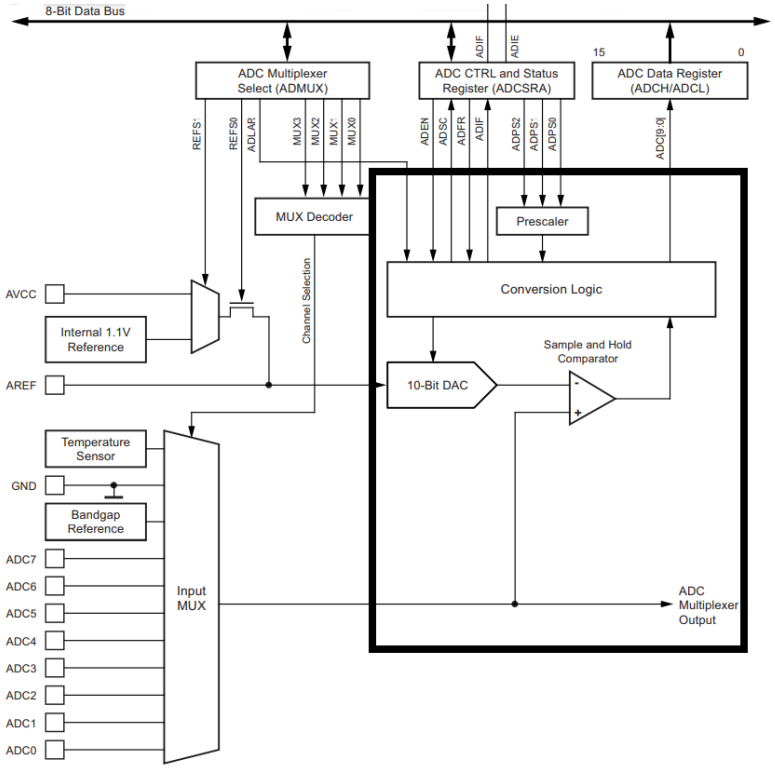

Figure 4.6: Block Diagram of the Analog to Digital Converter of the Atmega 328p microcontroller. Adapted from [24]

In a successive approximation analog to digital converter, as can be seen by the block diagram, the voltage to be converted is fed to the sample and hold circuitry that locks it from varying while the conversion takes place. The conversion is made by comparing the input voltage with a reference one. The reference voltage is outputted by the DAC. At the beginning of a conversion, the conversion logic sets the output voltage of the DAC at half the maximum value of a reference voltage, i.e starts with the Most significant bit turned on. The analog comparator signals if the DAC voltage is greater or less than the input voltage. If it is greater, the conversion logic holds this bit of the binary representation of a voltage and proceeds to increment the following significant bit of the DAC. When the DAC's voltage is greater than the input voltage the conversion logic unit will reset the current bit. This constitutes one cycle of the SAR ADC. This process is repeated from Most Significant bit to Least Significant bit (from

the right most bit to left most bit) until it has traversed all bits.

The architecture of this ADC determines the minimum elapsed cycles needed in order to complete a single conversion. A conversion is started with the selection of the channel susceptible to the conversion from the multiplexer, then the sample an hold circuitry takes, effectively, 2 ADC clock cycles [24] in order to stabilize the voltage to be read and finally the iteration thought the various 10 bits takes precisely 10 cycles, adding up to a total of 13 ADC clock cycles to complete a conversion. The number of ADC clock cycles in conjunction with it's frequency determine how long it takes to obtain a sample. In order to guarantee the maximum resolution of 10 bits, the successive approximation circuitry requires an input clock frequency between 50kHz and 200kHz REF. The ADC module contains a prescaler, which generates an acceptable ADC clock frequency by dividing the external clock frequency by 2,4,8,16,32,64 or 128. The Arduino Board contains a 16 MHz external oscillator and the only prescaler value that falls within the recommended frequency is the 128 value, hence the ADC clock cycle has a frequency of $\frac{16\times10^6}{128} = 125$ kHz. Now, it is possible to compute the minimum time to obtain a complete sample by multiplying the total number of cycles with the time of a cycle : $13 \times \frac{1}{125\times10^3} \approx 100\mu$s which equates to about 10×10^3 samples per second. It should be noted that this values are a best case scenario that won't ever be achieved unless the ADC is set to free running mode, in which case the ADC makes back-to-back conversions, i.e as soon as one sample is acquired another conversion cycle is automatically started.

On the bottom-left side of the diagram are the physical pins where the signals to be read are connected, represented as small squares. There are a total of 6 pins labeled as inputs for the ADC. Notice the channels are connected to an analog multiplexer, whose single output goes into the sample and holder circuitry of the ADC. It is clear that only a single channel can be read at a time.

Above the ADC input pins, there are the AVCC and AREF pins. Their purpose is to provide the DAC circuitry a voltage reference. The AVCC pin provides a voltage straight from the microcontroller power supply, whereas the AREF pin is intended to be connect to an external voltage reference if the user so desires. The voltage supplied to the DAC as a reference must be bounded to the operating limits of the microcontroller, i.e 0V-0.3 V and 5V+0.3V as these are the maximum rating specified by the manufacturer. The ADC's precision is determined by voltage reference. While the ADC's resolution specified as 10 bits nominal, however, this figure does not include deviations from ideal behaviour. The manufacturer, very briefly, specifies an absolute accuracy of ±2 least significant bits. This is the compound effect of offset, gain error, differential error, non-linearity, and quantization error parameters in the deviation from ideal behaviour.

The signals intended to be read range from ±10V. In order to be compatible with the ADC they will need to be converted to some other range anywhere between 0-5V but at a maximum ceiling and floor of 0-5V, respectfully. For this application it's best to has the ADC operating with it's maximum 5V range. The reason to convert to 0-5V is to preserve the most signal to noise ratio. Let's suppose the noise in the system is about 50mV peak-to-peak the SNR $= \frac{|0-5|}{50\times10^{-3}} = 100$ whereas if the signal were bound to $0-1.1$V, for instance, the SNR would decrease about 5 times to $\frac{|0-1.1|}{50\times10^{-3}} = 22$ and effectively reducing the resolution of the ADC because now $\frac{50\times10^{-3}\times2^8}{1.1} = 12$ out of $2^8 - 1 = 255$ bits are filled by noise. Within the $0-5$V range, using an effective bit resolution of $10-2=8$ bits, the ADC can distinguish voltages within $\frac{5}{2^8} \approx 20$ mV that translates to $\frac{|-10-10|}{5} \times 20 \approx 80$ mV in the ±10V range as this range is four times larger than the one used to convert the voltage.

Another important feature present in this microcontroller are Timers and Counters. Timers and Counters allow for an accurate execution of time critical tasks. Timers and Counters are the same entity since keep in Time is the same thing as counting how time intervals. A time interval occurs at each clock tick. This module allows for a flexible time interval operation by setting a prescaler that divides the number

of clock pulses by some integer value in order to lower the frequency of the signal used as clock and thus manipulating the time resolution. Timers/Counters are very flexible tools. In a simple mode, it is loaded into memory an 8 bit number that will be used as a comparison variable. The counter starts out with a null value. Each tick of the clock, the counter increments the value on its address by 1 and compares it to the value stored in the comparison register until a match is found. When a match is found a flag is raised and an interrupt may be triggered. At this point, three actions may be performed: either the timer value resets back to zero in order to repeat the process or decrements back to zero or the timer/counter may be left to be incremented until it reaches the maximum 8 bit value and in the subsequent cycle it will overflow back to its null position. Besides recording time, they also allow for signal modulation, such as Pulse Width Modulation(PWM) generation.

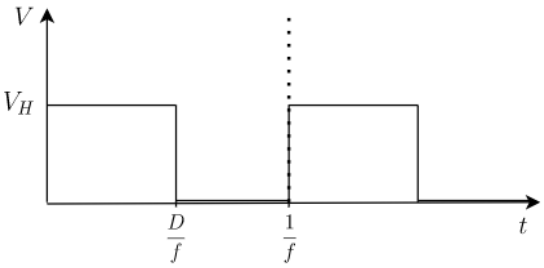

Figure 4.7: Ideal Pulse Width Modulated (PWM) signal

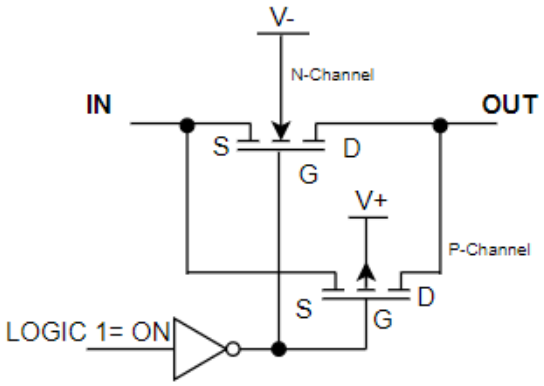

Figure 4.8: Internal circuitry of a switch in an analog multiplexer.

It consists in a signal that is alternated between two values a HIGH, 5V in this case, and a LOW, 0V ideally. The PWM wave form is characterised by two parameters, a portion of time the signal is ”on” (HIGH level), called duty cycle *D*, and the time it takes for the wave to repeat it self, the period of the PWM. Figure 4.7 showcases the generic PWM signal. The signal starts out in a HIGH state and it is toggled when there is a match between the compare value loaded into memory and the current counter number. The counter continues to accumulate ticks until is reaches the the maximum value an 8-bit register can represent. After it overflows, the cycle is renewed. The compare value loaded into memory is the duty cycle and the clock time intervals settings in conjunction with the length of the registers, i.e the maximum ticks it can count to, is the period of the PWM pulse. There are 2 8-bit Timers/Counters that serve 2 pins each and, even tho the Atmega 328p is has an 8 bit microprocessor, there is a third 16-bit Timer/Counter that serve another 2 pins. This totals 6 PWM pins. As will be apparent in a later section, the PWM may be used as a variable voltage source.

Similar to the Timers/Counters is what’s called a Watchdog Timer. This Watchdog Timer has the same functioning principle to the Timers described above however, it has a special feature, it has the ability to reset the MCU. When enabled, if the the Watchdog Timer ends its countdown the system may issue an interrupt and/or may be forcibly reset. The system reset operation restarts the program. It is a very useful feature as it can unlock the processor from either a malfunction or an infinite loop. In order to prevent a system reset, the Counter must not reach a given timeout value and so the processor must programmed to reset the Timer (”pet the dog”). This must be done periodical since as soon as the Counter is reset, it will restart the countdown. This feature is a great security measure that we can fallback on when there is some software malfunction.

An essential feature implemented in the majority of microcontrollers, including the Atmega 328p, are

Interrupts. Interrupts are mechanisms to control the flow of execution of the program whenever a certain event occurs, in other words, whenever a specific event that is programmed to trigger an Interrupt occurs the microcontroller will suspend the main program and execute a piece of code that was programmed to be associated with Interrupt. For instance, in the this specific microcontroller events that may have an Interrupt are: whenever a pin's state level is changed (changed from 0V to 5V), events associated with the timer, i.e whenever a timer overflows, or the counter has found a match, there is an interrupt for when the ADC has completed a conversion, there are also interrupts that signal when a given data protocol has completed a transfer or as received a group of bits, and a few more.

Besides having satisfactory specifications, the Atmega 328p has several hardware implementation of communication/data transfer protocols peripherals. This feature is important for the PI task. The additional circuitry implements one Universal Asynchronous Receiver-Transmitter (UART), one the Serial Peripheral Interface (SPI) and one 2-wire Serial Interface, commonly known as I^2C. These short length communication protocols will be discussed at a later stage.

Now that have been covered some aspects of the microcontroller it is time to briefly look at at the hardware on the board.

One benefit of choosing an open source platform is that their schematics, other design documents and source code are publicly available.

In figure 4.10 it is shown part of the schematic of the Arduino Uno Rev3 Board.

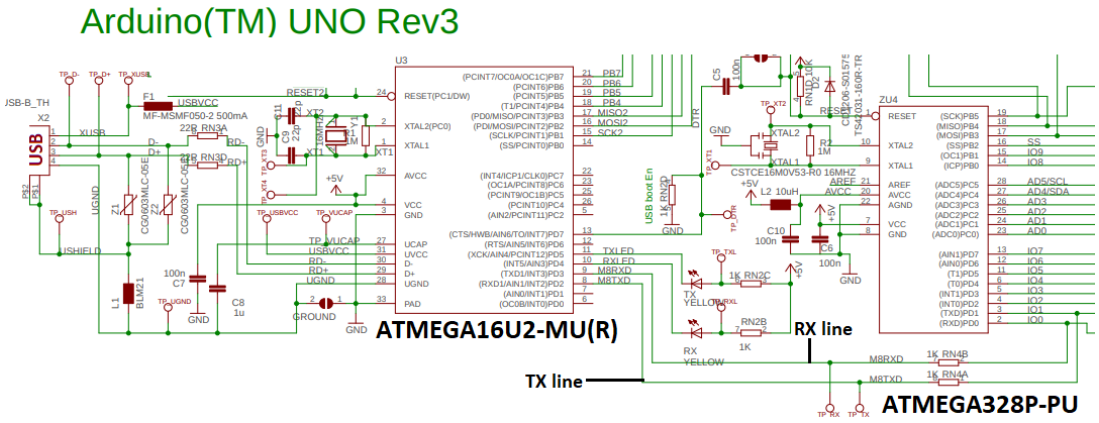

Figure 4.9: Arduino Uno Rev3 Board Schematic. The Tx and Rx lines in the main microcontroller (Atmega328p) are connected to the Rx and Tx lines of the secondary microcontroller that subsequently comunicates with the computer via the USB protocol. Adapted from the Arduino website

There are two ways of powering the board either through a DC barrel jack rated with up to 12V, whose voltage is subsequently regulated down to 5 and 3.3V, or via the USB connector that already provides a 5V regulated voltage.

In the schematic it is shown two microcontrollers, one is the Atmega 328p which is the main microcontroller, the one that will be running our program and executing the tasks, and a secondary one that was implemented in order to provide a USB 2.0 protocol hardware implementation, i.e the use of this microcontroller is to allow the communication with a computer via USB.

Special attention must be payed on how both of the microcontrollers are connected. At the bottom right side of the secondary microcontroller diagram symbol, it is noticeable that the pins Tx and Rx of this chip are connected to the Rx and Tx pins, respectfully, of the Atmega 328p, the main microcontroller. These are the UART Transmit and Receive lines. This protocol allows for a asynchronous data transmission from each device. This implies that the main microcontroller communicates with the

secondary one via UART and, subsequently, the secondary microcontroller converts that data into USB protocol compatible that makes it's way to the computer or vice versa. Two important notes here, firstly this is how the Atmega 328p microcontroller sends and receives data from the Computer via USB, and secondly, if we are to take advantage of the USB protocol to constantly send and receive information from the computer we cannot use this two pins for any other purpose.

Having touched on the unavailability of the UART protocol due to our application needing a constant stream of data from and to the computer, it is worth mentioning a few notes on the other two remaining communication protocols, I^2C and SPI. Without going into much detail about these protocols, it is important to mention some features that distinguish both. I^2C is a two-wire communication protocol, one line is the Data Line and the other is the reference clock to synchronize the recipients of the messages. This protocol allows the communication with several devices. All the devices share a single data line and so only one can talk at a time. As for the data structure scheme, the I^2C starts with a START condition where the Master takes control over the data line, and afterward it sends a packet of data containing 9 bits in which is always present a 7 address bits to signal what device it wants to transmit data and, after recognizing it's address, the peripheral device must respond with an ACKNOWLEDGE bit. After having exchanged information, the Master performs a STOP condition signaling the other devices it has finished. This is to say that the I^2C data protocol has built in handshake and error handling. Since the address line is 7 bits wide, it supports up to $2^7 = 128$ different peripherals with a single shared data line.

On the other hand, the SPI uses at least 4 pins, a Controller Out Peripheral In (COPI) line used by the master to send data to the peripheral devices, a Controller In Peripheral Out (CIPO) line used for the peripheral to send data to the master, a clock line (SCK) and an active low Chip Select ($\overline{CS}$) used by the master to address the peripheral it wants to communicate with. This protocol shares with all connected devices the COPI,CIPO and SCK line but it needs one chip select line per peripheral. Obviously, this protocol is more expensive in term of pins used. Communication wise, it is a straight forward, the master, in our case it will be always the Arduino, manages the devices it either sends or receives data by lowering the $\overline{CS}$ line. However, unlike I^2C, it lacks any built in handshake or error handling procedure. This must be done via software.

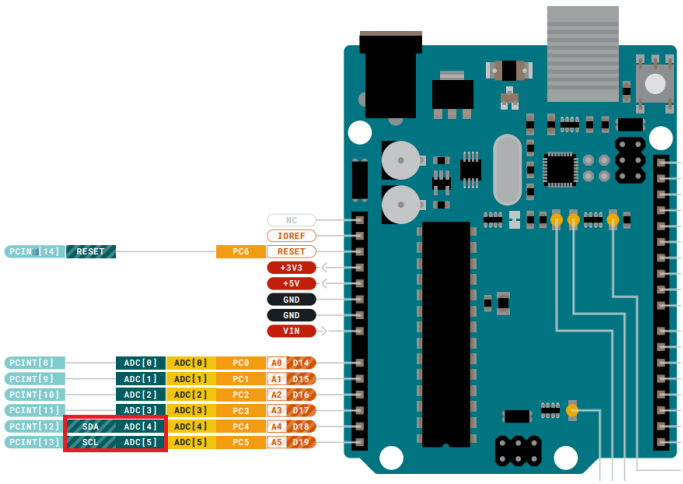

Figure 4.10: Arduino Uno Rev3 Board Pinout Schematic. It can be seen int the botton left side pins that the ADC 4 and 5 pins have as alternate function the SDA and SCL lines of the I^2C protocol. Adapted from the Arduino website

Ideally we could make use both communication protocols since there is a big influence on the availability of components on the market that implement each protocol. However, there is a big constraint for our application. Unfortunately, as it is possible to confirm by figure 4.10, the I^2C pins have the alternate

function of being an ADC input pin. For our application, we prioritize having available as much ADC pins as possible hence, we decided to use SPI to communicate with the components, even thought it greatly uses more pins which is a scarce resource in the microcontroller used.

However, there is a big constraint for our application. Unfortunately, as it is possible to confirm by figure 4.10, the I^2C pins have the alternate function of being an ADC input pin. For our application, we prioritize having as much ADC pins as possible available hence, we decided to use SPI to communicate with the components, even thought it greatly uses more pins which is a scarce resource in the microcontroller used.

In light of this overview, it can be concluded that the Arduino, even thought it is not a high performance device relative to today's standards, has satisfactory features and characteristics that make it adequate for implementation on this project. Also, the platform it self is a massive deciding factor since there is a great amount of support that, on the software side, greatly simplifies the overhead needed in order to have a finished project.

4.2.2 Interface Circuitry

The absolute maximum ratings of the Arduino specify a limit of $V_{CC}+0.5$ V and -0.5 V as the maximum and minimum voltage that are safe to the device. Stresses beyond those may cause permanent damage. The vast majority, if not all, of the lab equipment operates under the standard voltage range of -10 to 10 V both as input and output and so there is a need interface the output of the external equipment and the input of the Arduino's microcontroller. This equates to translating the ± 10 V to a 0-5 V level.

To achieve this, it is implemented a Amplifier with a differential configuration as is represented in figure 4.11.

The ideal OpAmp follows the equation $V_O = A(V_+ - V_-)$. With the configuration shown in figure 4.11, the only variable is the voltage provided by the outside circuitry and so we may rewrite the equation as a simple linear transformation :

$$V_O = aV_i + b \tag{4.1}$$

Where a is the slope of the line and b the offset. With that configuration, the linear coefficients may be obtain following Kirchhoff's rules and the OpAmp's virtual short principle ($V_+ = V_-$), it is obtained the following relationship:

$$V_O = \frac{R_1+R_F}{R_1}\left(V_i\frac{R_3R_4}{R_3R_4+R_2R_3+R_2R_4} + 3.33\times\frac{R_2R_3}{R_3R_4+R_2R_3+R_2R_3}\right) \tag{4.2}$$

To obtain the linear coefficients values, that will translate into resistor values, we need to solve the system of equation 4.1 that represent the state of the system at the maximum and minimum allowed voltage values:

$$\begin{cases} 5V = a\times 10V + b \\ 0V = a\times(-10V) + b \end{cases} \tag{4.3}$$

where it is obtained $a = \frac{10}{40} = 0.25$ and $b = \frac{10}{4} = 2.5$. The desired resistor values were selected based on the linear coefficients and the constraint that it was required to have a feedback gain $(R_1+R_F)/R_1 = 3$. The circuit was implemented with $R_1 = 5$ kΩ; $R_F = 10$ kΩ ;$R_2 = 60$ kΩ; $R_3 = 7.5$ kΩ; $R_4 = 20$ kΩ.

There are, however, some signals, as is the case of the amplitude signal provided by the *Lock-in*

Amplifier whose range is bounded from 0 to 10 V. If the Level-Shift Amplifier, as the previous one, is employed only half the operating range is used and it is lost half the resolution it could otherwise be had. Some optimizations in the amplifier circuit for signals like these were planned.

Here there is no need to offset the input signal, only scale it appropriately with a $1/2$ gain. This could be done with the circuit similar to figure 4.11, where resistors R_1 and R_F are replaced with wires, making the feedback gain unity, removing R_4 and the 3.33 V voltage source and selecting the same resistance value for R_2 and R_3. However, the same outcome is achieved with a inverting configuration as shown in figure 4.12 except the output voltage is inverted relative to the input. This was the chosen configuration as the inverting will not have any effect as it will be seen later.

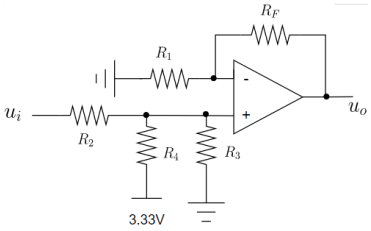

Figure 4.11: Schematic of the ideal interface circuit that translates an input voltage in the $-10-10$ to $0-5$ V, in order to be compliant with the ADC's maximum ratings.

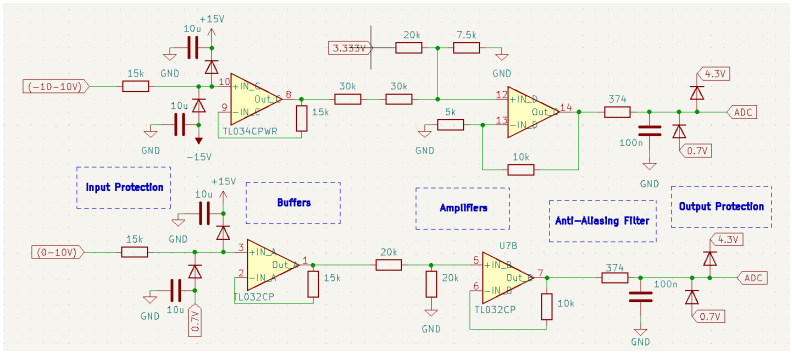

Figure 4.12: Real Interface circuitry implementation between the ADC channels and the external input. In the upper schematic, the circuits accommodate -10 to 10 V as input into $0-5$ V. The bottom circuit shifts voltages from the $0-10$ V ranges into $0-5$ V. The schematic shown was obtained via the Kicad open source PCB design software.

There is always a layer of caution that must go into designing an interface whose input is "open to the outside world", and especially when a human is on the other side. Although specifying an Absolute Maximum Rating Specification for an equipment excuses the designer from any legal responsibility, it would be extremely disappointing if that was the approach. This project is intended to be a versatile and reliable, yet fairly accurate and precise, equipment that would benefit the research group by providing powerful features and improving the overall experience using the AFM prototype. This also means that the device must be robust to some faulty/erroneous use that it will be inherently exposed to.

The full ± 10 V and $0-10$ V Interface circuits are shown in figure 4.12.

The input to the interface circuit is on the right side of the picture. It was decided to shield the output of the voltage source that is connected to the input of this device from any current draws. If the external device's output were connected straight to the the input of the circuit in figure 4.11 it would draw/inject current from/to the external device, even if it the input resistance in the order of the kilo ohm and thus the current has a small magnitude it can be avoided. Furthermore, with the configuration of figure 4.11, the resistor network at the non-inverting input of the OpAmp, effectively decreases the, ideally, infinite input impedance of the OpAmp to the parallel equivalent of the resistor network as seen from the external device. This external device is usually a voltage source with, ideally, zero impedance, however, given the circumstances, the user may present at the input of the interface circuitry a circuit with considerable output impedance and, in conjunction with the input impedance of the interface circuit, a voltage divider network will be formed. The lower this input equivalent impedance is the more loading impact it will have on the input circuitry. As a result, the voltage to be read will deviate from it's actual value. To overcome this problem, it was employed an OpAmp in Buffer Configuration as a first stage of the Interface circuit.

Besides covering the unnecessary loading of the input, it was also addressed over-voltage and over-current conditions. For the over-voltage protection, it was added two clamping diodes at the input of the buffer connected to the power supply. If the input voltage exceeded beyond ± 0.7 V of the supply voltage, the diodes will conduct and short the input to the power rails, not letting this node exceed this limit. A series resistor was added at the OpAmp's input to limit current develop when the diode shorts the the supply and the input node. Also, this resistor prevents large current peaks injected at the input of the buffer.

The input voltage is, effectively, capped at ± 15 V. However, this still poses another problem further down the line. The level shifter Amplifier was designed to maximize the working range of the ADC given an input within the standard Voltage range. Whenever this criteria is exceeded, the OpAmp will still react according to equation 4.1 because it is still withing it's linear range and so it will feed a voltage to the ADC greater or lower that it's absolute maximum ratings and may damage the microcontroller. This is highly undesirable. To protect the input of the ADC, another pair of clamping diodes were added to the output of the level shifter Amplifier as shown in figure 4.12. It does not make sense connecting them to the power supply lines. Since we need to protect the input it makes more sense to bias them at 5 V and ground. With this, whenever the voltage on the input node of the ADC exceeds either 5.7 V or -0.7 V, approximately, the diodes will conduct. Although this is still considered a safe design, this configuration would still exceed the specifications of absolute maximum characteristics, albeit marginally. It was thought that it could still be improved by having a voltage source provide 4.3 and 0.7 V so that it is achieved the clamping at around the 5 V and 0 V. This, however, is never precise since diode forward voltage is current and temperature dependent. Nonetheless, this biasing scheme was employed to be in agreement with the microcontroller's DataSheet and will also be useful in other areas of project. Also, there was implemented a current limiting resistor at the output of the OpAmp to prevent elevated currents when the diodes short this node.

These protection measures imply a performance trade off when we transition from ideal models into the real world and, albeit briefly, should be discussed.

The voltage clamping diodes generate an inherently small dark current that is dependent of the temperature and reverse bias voltage applied to the device. This leakage current in conjunction with the current limiting resistor at the non inverting Input of the OpAmp generate and offset voltage. And this is aggravated by the small current required by the OpAmp to bias itself. In any OpAmp datasheet there is specified this input bias current and other non-ideal deviations. Furthermore, the OpAmp exhibits a small offset voltage between it's inputs. The offset voltage that is built up from these factors is especially degrading when it's feedback gain is high because it will be amplified at the output of the device. At this buffer stage, the total error contributions can be obtained by the following:

$$V_{BufferOS} = V_{OpAmpOS} + R_{limit}(2 \times I_{diode} + I_{bias}) \tag{4.4}$$

At the next stage, the offset voltage contributions from the OpAmp are similar to the ones described previously. In this case, the input bias current generates an offset via the equivalent impedance of the non-inverting input resistor network and the inherent offset voltage from the OpAmp itself. In this stage there is a feedback gain amplifies the cumulative offset voltages by a factor of 3. Also, the output voltage clamping diodes and even the ADC pin leaks some current to the output current limiting resistor which will add another offset. With the values specified from the manufacturers of the all individual components that were chosen, the total offset voltage was computed to be 26 mV, which corresponds to an offset error of about 6 bits in an ADC reading.

Additionally, the output current limiting resistor that was placed will form a voltage divider network with the ADC pin's input impedance. The manufacturer provides a wide interval for this value ($1-100$ kΩ). If it is used the worst figures for the estimation, the output voltage will be reduced by a factor of 0.77, and 0.99 for an impedance in the middle of the range.

This shows a compelling reason to chose low value resistors for circuits where accuracy is highly valued and, although the system will have increased power consumption, there will be less Johnson noise present in the system as it scales with the value of the resistor.

There are a total of 6 interface circuits, one for each ADC pin. Four of those are meant to handle voltages that range from $-10-10$ V and two that handle voltages from $0-10$ V. One out of the four $-10-10$ V interface circuits is not exposed to the user, since it is used to read the output voltage from the PI controller.

There was implemented a 5V external voltage reference for the ADC with initial accuracy of $\pm 0.1\%$. It was leveraged this to make the 4.3 V and 0.7 V voltage sources from a simple voltage divider and buffer amplifiers. These clamping voltage levels will be shared by all the interface circuits and also by the output stage of the PI controller. In anticipation of a worst case scenario, one might expect all the circuits having their over voltage mechanism activated at the same time. The current needed to sink or source from real OpAmps is limited to a few dozens of milliamps, thus it was needed to implement a current sink/source as is shown in figures 4.14 and 4.13. The OpAmp biases the bjt transistor in order to meet the virtual short, i.e $V_+ = V_-$. It provides a current at the base of the transistor. V_{CC} is the positive rail of the power supply in case of a current source or the negative rail in case of a sink. The current sink is the circuit of the 4.7 V clamp, as the voltage at the anode increases from 5 V the diode starts conducting current into the the pnp transistor that sinks it to the power supply rail. The current source has a similar behavior.

It is important to mention that the ± 15 V power supply was not subject to design due to being already available at the lab. This particular power supply is rated up to 400 or -250 mA which is enough for our project.

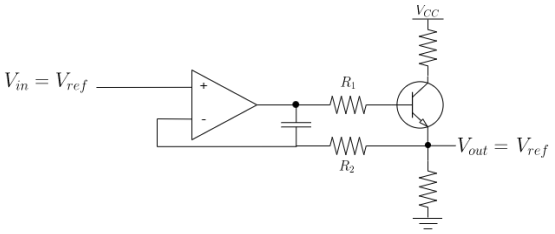

Figure 4.13: Schematic of a current source used for the voltage clamping mechanism

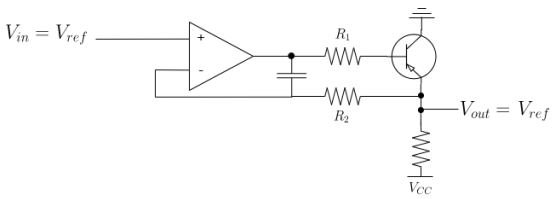

Figure 4.14: Schematic of a current sink used for the voltage clamping mechanism. In this situation, the voltage reference is 0.7 V

4.2.3 PI implementation

This section describes the implementation of the Control System.

The circuit is shown in figure 4.15. The same safety guidelines were also implemented in this section. Here it were chosen different resistor value in order to decrease the trade-off between power consumption and signal degradation.

As was pointed out throughout chapter 2, there are a variety of signals that can be used as input to the controller, the most common ones are the amplitude (AM mode) and the vertical deflection (contact mode) signals. The user is encouraged to swap between them regularly. One design choice of this device is to avert the user from having to constantly plug and unplug the signal cables from the input

terminals every time the mode/configuration is changed. To this end, as similar the Arduino's ADC structure, it was decided to multiplex the input signals to the control system, as shown in figure 4.15. The analog multiplexer allows the user to easily choose the variable for the control system from the computer interface. And, the selection of this device must meet the requirements of having at least 6 inputs, be compatible with $\pm$10V operation and, ideally, the channel switching would be performed via SPI.

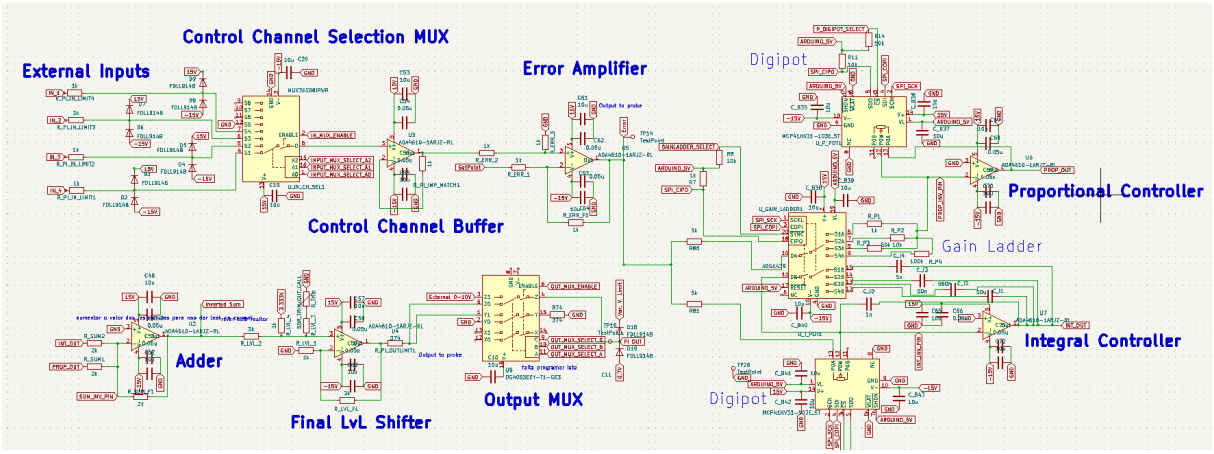

Figure 4.15: PI control system implementation schematic

Each channel in a Multiplexer is, in essence, a switch that is either on or off. The structure of the switch is a N-channel MOSFET in parallel with a P-channel MOSFET as is shown in figure 4.8, to accommodate for the positive and negative signal at the source/drain. This switch is turned on and off via the voltage applied at their gates. This type of switch has no preferred input or output direction.

The output of the multiplexer is connected to the non-inverting input of the OpAmp that is used as a buffer. With this configuration, the input is insensitive to the on-resistance of the multiplexer channel and no signal losses are induced.

Next there is the Error Amplifier. It's circuit configuration is displayed in figure 4.16.

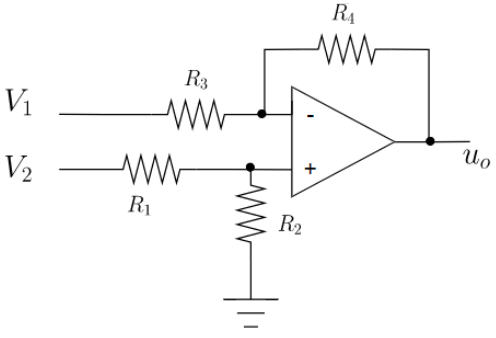

Figure 4.16: Difference Amplifier that produces the error signal of the controller block

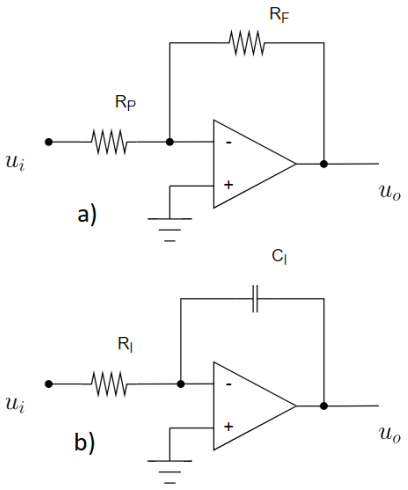

Figure 4.17: Schematic of the simple analog a) Proportional and b) Integrator circuits.

This stage is in charge of computing the difference between the control variable, V_i, and the setpoint

value,V_{SP}. The generic equation for the output of this stage is:

$$V_O = \frac{R_4}{R_3+R_4}\left(\frac{R_2+R_1}{R_1}\right)V_2 - \frac{R_2}{R_1}V_1 \tag{4.5}$$

by selecting resistor values such that $R_4 = R_2$ and $R_3 = R_1$ the equation simplifies to :

$$V_O = \frac{R_2}{R_1}(V_2 - V_1) \tag{4.6}$$

It is of extreme importance to consider the equation 4.6 in relation with how to AFM operates. There is a remark that needs to be made to ensure the control works properly. The order of the input signals V_i and V_{SP} presented at this stage is paramount. It is linked to how the control variable changes with the output of the PI. To further elaborate the idea, let's examine how the control output voltage should behave, for instance, with a variation of the Amplitude signal. In equilibrium, the setpoint value, V_{SP}, and the Amplitude signal, here the control variable V_i, have the same magnitude and nullify the error signal. Let's assume, for simplicity, that the piezoelectric element is displaced at half of the amount it can fully expand. Imagine now, that a disturbance, in the form of a "mountain" in the topology of the sample, is introduced in the system. With increasing tip and sample proximity, following the interaction graph 2.2, the force gradient has a larger magnitude which also dictates a larger elastic response of the sample thus, it adequately decreases the vibration amplitude of the cantilever, per equation 2.25. To properly return the system to the equilibrium distance, the piezoelectric element must have its excess length reduced in order to separate the tip and sample. This means the control system must apply a decreased voltage. Similarly, when the topography of the sample is a valley, the distance between surfaces widens and the vibration amplitude of the cantilever increases. In this situation, the applied voltage to the piezoelectric must increase accordingly. With this in mind, looking back at equation 4.6, the error signal must follow the same functional relationship as the control variable. Since the setpoint voltage is a constant in this equation, the configuration that satisfies this condition is having $V_1 = V_{SP}$ and $V_2 = V_i$:

$$V_{Error} = \frac{R_2}{R_1}(V_i - V_{SP}) \tag{4.7}$$

This train of thought is also applied to the vertical deflection signal. As the force applied at the tip increases the cantilever has it's vertical position increase. The laser spot at the PSD is raised and the upper half of the quadrant has more lit area. According to equation 2.15, the vertical deflection signal is incremented. This contrasts with the behavior of the Amplitude signal with increasing interaction. This is how the system was described to operate, for ease of understanding however, in the actual implementation, the PSD difference amplifier performs the following operation: $Vertical = (C+D) - (A+B)$. As the laser spot is vertically displaced, the upper quadrants generate a larger signal and, following the summing operation, the output is decreased. And, thus the vertical deflection signal is in agreement with the expected error signal behaviour.

The gain R_2/R_1 should be unity so it doesn't pre-amplify the signal.

Up Next is the Proportional and Integral blocks. In figure 4.17 is shown the schematic of the circuits that make up the Proportional and Integral blocks. These are, topology wise, the simplest circuits one can build for the intended effect.

The formulas for the proportional and integral gains are, respectfully:

$$\frac{V_O(s)}{V_i(s)} = P(s) = -\frac{R_F}{R_P} = -K_P \tag{4.8}$$

$$\frac{V_O(s)}{V_i(s)} = I(s) = -\frac{1}{R_I C_I s} = -\frac{K_I}{s} \tag{4.9}$$

Both of these stages invert the signal. This detail was taken into account in both the preceding stage, the error circuit, and the following stage, the summing amplifier, so that the control circuit correctly manages the instrument.

The simple circuits will be the baseline model which will be built upon in order to achieve the intended design goals. With this architecture, the Proportional and Integral block work independently from one another. This is crucial for the tuning of the control system by the user as the gains may be set individually and any interaction between the two modules is minimized.

To achieve one of the main design goals - the digital gain selection of an analog circuit - it was devised the use of a digital potentiometer whose purpose is to increments/ decrements the gains with a small resolution. Note that both the proportional and integral gain scale with the resistor values. Additionally, it was found to be appropriate to have a mechanics through which the gain could be changed by a greater amount, say a order of magnitude for example. Note that by using the potentiometer alone the gain could only be changed within some limit given by the fixed value resistor, in the case of the proportional block, or capacitor, in the case of the integral one. With this setup, the fixed resistor/capacitor is now susceptible to change and increase/ decrease the gain by some order of magnitude, thus covering a broader range without a heavy penalty in the step resolution. To this end, an analog multiplexer was added to the circuit.

As seen in figures 4.18 and 4.19, the multiplexer stage serves the purpose of a ”gain ladder”, were only a line closes the circuit, and thus the component on that branch sets the value of the gain.

Thus gain stages have 2 selective scales, a coarse and a finer one. The course gain consisting of a multiplexed ladder of resistors, in the proportional case, and a ladder of capacitors in the proportional case whose values have increasing orders of magnitude. The finer gainer will be provided by the digital potentiometer.

The digital potentiometer has a finite number of steps in between resistance values , for instance, if a digital potentiometer is 8-bits, it will have $2^8 - 1 = 255$ increments over the some nominal value.

In the Proportional block, the configuration is as follows: there is a fixed value resistor R that in conjunction with the R_i, which symbolizes a particular resistor value in the ladder, form the denominator in the K_P formula. In the feedback loop is where it was placed the digital variable resistor. The formula for the gain is equation 4.10.

$$K_p = \frac{R_{digipot}}{R + R_i} \tag{4.10}$$

With this configuration, two important feature are obtained: firstly, the gain value is easily set up to be less than unity. This is important as in the experimental tuning, the control loop would become unstable with larger than unity gains. Thus the resistor value choices reflect this condition. Secondly, it is possible to turn off this stage completely, either by setting the potentiometer value to zero or opening the circuit in the resistor ladder, since this stage is secondary when it comes to the control system and it in some situations may be useful to be turned off.

In the Integral block, it is also present a fixed resistor at the circuit’s entrance, which is followed by the digital potentiometer. In the Integrator, the coarse gain is set by the capacitors via the multiplexing ladder in the feedback loop, and there is also a fixed capacitor. The formula for the Integral Gains is given by equation 4.11:

$$K_i = \frac{1}{(R + R_{DigiPot})(C + C_i)} \tag{4.11}$$

The C_i is the value of the capacitor selected in the multiplexer ladder.

Here, the fixed value resistor biases the gain resistance equivalent preventing the gain to shoot to infinity when it has reached the its lowest position possible. (This could be prevented in software also). In the real world, the gain won't rise to infinity, it will be bounded by the open loop gain of the OpAmp.

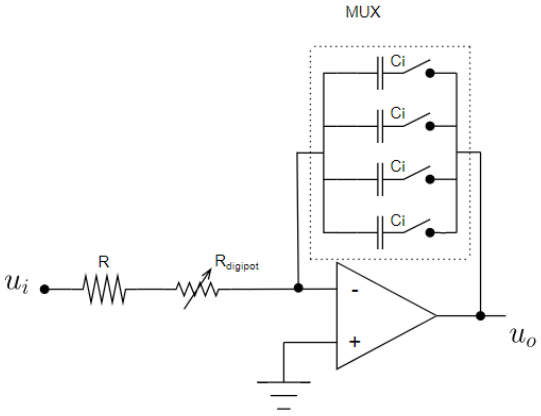

Figure 4.18: Schematic Integrator circuit's implementation that allows for a digital control of the gain.

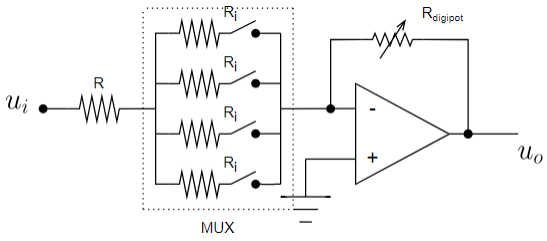

Figure 4.19: Schematic Proportional circuit's implementation that allows for a digital control of the gain.

Both the digital potentiometer and the analog multiplexer are the key elements that enable the digital gain setting without compromising the features of the analog control. A basic explanation of the analog multiplexer was already covered in a preceding text, and now a brief explanation of how the digital potentiometer works is given:

Similarly to it's analog counterpart, the digital potentiometer is a 3 terminal device (The chip itself has more pins but the physical resistance has a 3 terminal configuration), two ends, we may name them A and B and an adjustable wiper, W. In the analog device, the wiper is a contact in the A-B resistance that acts as a voltage divider, and can be physically moved in order to increment/ decrement the resistance ratio. In the digital world there are no moving parts. The bulk resistance AB is made up of a network of discrete resistors that may be seen as "resistor steps". At the end of each resistance piece there is a switch that is connected to the wiper terminal, as shown in figure 4.20.

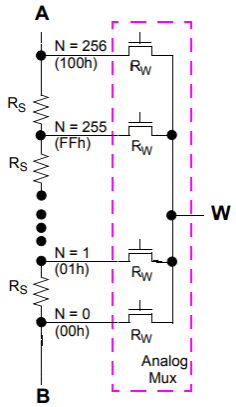

Figure 4.20: Illustration of the working principle of the an 8 bit digital potentiometer. Adapted from [26]

All the switches form themselves an analog multiplexer that selects a single contact point to the wiper. The location of this closed switch is what determines the resistance from terminal A to wiper and, consequently the resistance between terminal B to wiper. The total amount of "resistor steps" from terminal A, whose switch is off, to the final resistance step with its terminal turned on, i.e connected to the wiper, in relation with the total resistor steps along the AB network is what determines the voltage divider ratio.

For an 8bit multiplexer, there are $2^8 = 256$ "resistor steps", i.e. the total AB resistance is divided into 256 smaller ones. The total AB resistance network value divided by the number of switches define the resolution of the digital potentiometer, $(10 \times 10^3)/256 = 39\ \Omega$ for a 10 kΩ potentiometer, for instance. The switches that connect to the wiper have also a non-negligible resistance value, which is called the Wiper resistance. In our specific case, the digital potentiometer is used as a variable resistance, only two terminal are of use, either A-W or B-W. It should be noted that the gain resolution provided by the digital potentiometer decreases as the value of the fixed component in the gain ladder is increased.

Additionally, we are aware that adding this much complementary integrated circuits it is being introduced unwanted capacitance in the overall circuits, which is inevitably ruining the effective bandwidth. To minimize this effect one could distribute the digital potentiometer and the analog multiplexer into two cascading circuits, one that would provide the fine adjustment and the other the coarse. Although it was taken into consideration, we did not implement the circuit this way.

For our application, it is necessary to select a digital potentiometer that is rated for working voltages of $\pm$10V and that is programmable via Serial Peripheral Interface.

With the practical tuning of the PI, it was established the optimal gain values it should be implemented and so the design of the gain adjustment stage took those values into account and added some tolerance around the tuned gains, about one order of magnitude lower and one order of magnitude higher. For the Proportional stage it was selected a 7 bit 5 kΩ digital potentiometer and a 1 kΩ fixed resistor. In the gain ladder, there are a total of 4 resistors with 1 kΩ, 10 kΩ, 50 kΩ and 100 kΩ correspondingly.

For the Integral stage, it was also chosen a 1 kΩ fixed resistor that is followed by a 10 kΩ 7 bit digital potentiometer. As for the capacitor values, the fixed capacitor has a 1 nF nominal value and in the gain ladder the values are 5 nF, 50 nF, 560 nF and a 10 μF.

After the error signal's manipulation from both blocks it is needed to add them, hence, the next OpAmp is performing a summation. This summing stage is also where the signals from the PI are inverted back. The Adder stage as also a peculiar behavior. There is a precaution needed for this stage. When the OpAmp sums the signal from the Proportional Amplifier and the Integrator it is within possibility that the output value exceeds the intended $\pm$10V range that the next stage's design relies on. If the amplifier is left "naked", the voltage would only be capped by the maximum output voltage of the OpAmp (saturation), which is dependent on the specific OpAmp device used. There are OpAmps that don't have symmetrical saturation output voltages (positive and negative maximum output voltages are not equal). It is necessary to reliably limit the maximum output voltage of this stage so that the final Level Shifter behaves as intended. To overcome this situation, it was implemented a Symmetrical Voltage Clamp as can be seen from figure 4.21.

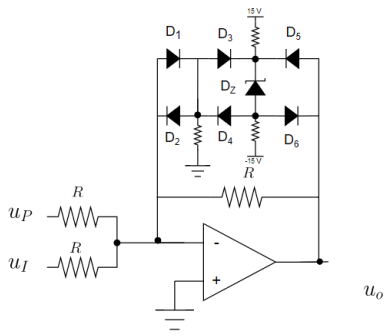

Figure 4.21: Summing amplifier with bounded outputs.

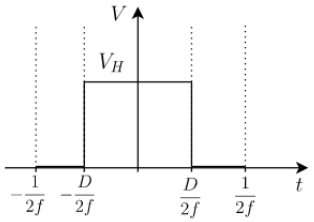

Figure 4.22: Representation of a PWM cycle centered around a symmetrical interval

This circuit takes advantage of the 0.7 voltage drop across a diode when forward biased and the operation of the Zenner diode.

With this configuration, the Zenner diode is reverse biased by the power supply voltage so that the voltage drop from its cathode to anode is it's break down voltage, for our purposes it was chosen a Zenner diode with 8.2 V. Considering the virtual ground at the inverting input, when the OpAmp outputs a negative voltage it will forward bias diode D_6 while diodes D_5, D_4 and D_2 are reverse biased. The circuit can be considered opened at this nodes. The voltage at which the diodes are forward biased is referenced to the virtual ground of the inverting pin. To calculate this voltage we need to add the total voltage drop from all of the diodes that are conducting. If we start from right to left, diode D_1 drops approximately 0.7V, so does diode D_3, now the cathode of the Zenner diode is 8.2V higher in potential relative to it's cathode so the voltage dropped here is the Zenner voltage, 8.2V, and finally diode D_6 will also drop 0.7V. The sum of all of this contributions equals $3 \times (-0.7) - 8.2 = -10.3$ V and so the output of the OpAmp will be bounded to this value. This circuit will have a similar behaviour for positive voltages. The OpAmp will perform the regular summing operation up until the voltage exceeds the voltage drop from the diodes.

The final signal transformation stage is the output Level Shifter. This stage is needed because the majority of piezoelectric elements used in AFM's do not operate with negative voltages and have a limited voltage range. This stage has the same operating principle as the circuit of figure 4.11 and only the linear coefficients are different. The output of this stage meets the required specification of being compatible with the standard voltage range of $\pm$10V. The system of equations that allows to calculate the linear coefficients is the following:

$$\begin{cases} 10 = a \times 10 + b \\ 0 = a \times (-10) + b \end{cases} \tag{4.12}$$

where $a = \frac{10}{20}$ and $b = \frac{10}{2}$.

This system of equation and 4.3 are linearly dependent, and so their coefficients are proportional to each other. This allows for the use of the same Amplifier configuration. Having the same configuration means sharing the same offset voltage (3.33V) and there is only the need to change resistor values.The recycling of the configuration provides a consolidation of other parts of the system. Furthermore, by adding a potentiometer it is possible to not only adjust the maximum output voltage out of the control system but also calibrate it. With an adjustable maximum output voltage this device can accommodate the usage of piezoelectric element stacks with different working voltages ranges, i.e $0-75$V, $0-100$V or $0-150$V.

Finally, the control system has one last multiplexing stage. The purpose of this is to provide an easy commutation between the output of the PI stage and an external signal. The feature allows to quickly switch between an scanning imaging mode to force spectroscopy, note that the force microscopy operation is made with the control system turned off, and the piezoelectric element is actuated via a ramp signal or a sinusoidal one, whereby it is fully stretched and then brought back to it's voltage free length. By incorporating this stage in the device, there is only a single physical output terminal that needs to be connected to the piezoelectric element Amplifier Stage.

In light of the safety precautions, it was also added voltage clamping diodes, one connected to the 0.7V source and another connected an adjustable voltage source intended to be altered, correspondingly, when a new maximum output value is configured in the Level Shift Stage.

Now it will be described the process to design a viable setpoint voltage generation. As quickly mentioned on the Arduino overview, it has a few pins with PWM capability.

To best understand the reason why a signal like this can be used as a variable voltage source it is really useful to look at it in it's Fourier Series representation.

The PWM is a periodic signal and so computing its Fourier Series is straight forward but there are a few consideration needed to be made. For this analysis it is considered an ideal PWM, meaning that the transition from HIGH state to LOW state, and vice versa, is assumed to be instantaneous. This is only an approximation since the there will be some rise time and fall time in the actual system that will contribute with other harmonics. Also, it will be assumed that the HIGHT and LOW voltages are exactly 5V and 0V. In practice this will far from the ideal case as noise will be present in the signals and these levels are depended on the stability of the power supply and how overloaded the microcontroller is with tasks, as it may need more power than what the power supply is able to replenish. By definition, the Fourier Series of a periodic signal with period of $1/f$ is :

$$V_{PWM}(t) = \sum_{n=0}^{n=\infty} a_n \cos(2\pi f n t) + b_n \sin(2\pi f n t) \tag{4.13}$$

where

$$\begin{cases} a_0 = f \int_0^{\frac{1}{f}} V_{PWM}(t) dt \\ a_n = f \int_0^{\frac{1}{f}} V_{PWM}(t) \cos(2\pi f n t) dt, \ \ n > 0 \\ b_n = f \int_0^{\frac{1}{f}} V_{PWM}(t) \sin(2\pi f n t) dt, \ \ n > 1 \end{cases} \tag{4.14}$$

Mathematically, the PWM signal may be represented within a single period over a symmetrical interval, as can be seen from figure 4.22:

$$V_{PWM}(t) = \begin{cases} 5, & -\frac{D}{2f} < t < \frac{D}{2f} \\ 0, & t < -\frac{D}{2f} \wedge t > \frac{D}{2f} \end{cases} \tag{4.15}$$

Where f is the frequency of the PWM of the wave and D is the duty cycle, which is a number between 1 to 0. This was done to represent the PWM as an even function: $f(-x) = f(x)$. The cosine function is also an even function and a sine function is an odd function: $f(-x) = -f(x)$. The product of an even function with an odd one yields an odd function and, conversely, the product of two even functions is even. It can be shown that the integral of an odd function over a symmetrical interval equals zero, while the integral of an even function equals twice the result when evaluated from 0 to the limit of integration. Therefore, coefficients of the sin terms, b_n, in 4.14 are all null.

Given the instances where the V_{PWM} is not null :

$$a_0 = 2f \int_0^{\frac{D}{2f}} 5dt = 5D \tag{4.16}$$

$$a_n = 2f \int_0^{\frac{D}{2f}} 5\cos(\pi fnt) = 5\frac{2}{\pi n}\sin\left(\frac{\pi Dn}{2}\right) \tag{4.17}$$

and thus:

$$V_{PWM}(t) = a_0 + \sum_{n=1}^{n=\infty} a_n \cos(2\pi fnt) = 5D + \sum_{n=1}^{n=\infty} 5\frac{2}{\pi n}\sin\left(\frac{\pi Dn}{2}\right)\cos(2\pi fnt) \tag{4.18}$$

Equation 4.18 provides another interpretation of the PWM signal. It expresses the voltage as DC value, $5D$, super-positioned with some time varying component. Let us call this time dependent term ripple voltage, it will be clear in a moment this designation. The ripple voltage is made up of signals with well defined frequencies, as per the argument of the cosine term. It it interesting to note that each frequency is proportional to the PWM frequency with equal spacing between them, and, more importantly, the magnitude of each harmonic is decreasingly smaller according to the coefficients a_n, which are time independent. The ripple voltage can be considered as noise in the system. It is vital to provide the error amplifier a stable voltage. The PWM signal needs to be filtered, in order to attenuate the amplitude of its harmonics.

Before proceeding to design the filter stage it is useful to understand how the characteristics of the microcontroller will affect the filter specifications.

The factors that the microcontroller influences are the resolution of the DC voltage and the frequency of the harmonics. These are, effectively, the error sources for the accuracy of the voltage source. To understand how the microcontroller influences the resolution we need to review how the PWM is generated: the microcontroller uses a timer/counter to increment a counter value. It then compares it to some user defined number. When there is a match it toggles the output state to 0V and continues to increment the counter variable until it reaches the maximum possible number. Hence, with an 8-bit wide register, for instance, the resolution of the duty cycle is $\frac{1}{256}$ that translates to $\frac{5}{256} \approx 20$mV. The duty cycle will determine the value of DC voltage, note that in the time dependent term, the duty cycle is present inside the sine, which is a bounded function from -1 to 1.

The frequency of the PWM, according to equation 4.18 sets first frequency of the harmonic spectrum. This is important due to our intention of applying this signal to a low pass filter to attenuate the ripple voltage. The frequency of the PWM is tied to the clock of the Timer/Counter and the bit depth used for the overflow value. As previously described, the clock of this module can be adjusted via a prescaler that divides the external clock frequency by either 1, 8, 256 or 1024. In conjunction with this, the PWM will repeat it self only after it has reached its maximum count value which, if the number is described in 8 bits, is 256. Thus the frequency of the PWM signal is given by:

$$f_{PWM} = \frac{f_{clock}}{N \times 2^{bits}} \tag{4.19}$$

where N represents the prescaler value, and f_{clock} is the frequency of the external oscillator, which in this case is 16 MHz.

The higher the bit depth and chosen prescaler value, the slower the PWM will be.

In the 16-bit Timer/Counter module the user may customize how many bits are intended to use to represent the maximum count value.

There will be a trade-off between the voltage resolution and the frequency of the PWM. The higher the frequency the more distant the harmonics are from the desired 0 Hz and, consequently, a less demanding filter stage may be designed - the filter's cut off frequency does not need to be close to DC, which in turn prevents a poorer dynamic response and allows for the usage of lower order filter that decreases the filter complexity and cost - but at the expanse of voltage resolution.

Before projecting the filter stage, let's start by discussing what are the acceptable figures for the relevant parameters of the instrument, namely the ripple voltage, the rise time of the signal and the resolution.

Ripple voltage of the setpoint carries on to the error stage and subsequently to the output of the controller. This is an important parameter as it can significantly damage the resolution of the AFM. We aimed to achieve a 5 mV or less ripple voltage.

As for the rise time of the filter's response, it is only important when an image acquisition is underway and there is a need to adjust the setpoint's value. It must stabilize within an acceptable time interval that depends on the line acquisition speed. The longer it takes relative to the line acquisition velocity the more noticeable artifacts will be present in the image. A good estimation starting point would be a line acquisition speed of 1 second per line, this is an average experimental value. Note that one line formation is composed of a Trace and Retrace motion. There is no quantitative value one could estimate, however, with some experimental experience a good figure to start would be about 25% of a trace time, i.e $0.25/2 = 125$ ms.

As per the voltage resolution, the PWM voltage is bounded between 0 to 5V which does not cover all the standard voltage range, $\pm$10V, in order to cover all the possible values of the control variable, and so the 5V must be translated into this range. When converting from ± 10 V to $0-5$ V there is a factor 4 that must be taken into account because the resolution decreases by that amount. Here there is no quantitative value to take into account, although it should be aimed for, at least, a resolution in the second decimal place of the volt scale.

Now let's consider the filter stage. In order to project this circuit, it should be started from the most simple low pass filter, ascertain if it meets our requirements, and move on to more complex architectures if needed.

A first order low pass filter is schematically a simple circuit with a resistor and a parallel capacitor. The Laplace Transform of the Transfer Function of this system is:

$$LP_{1st}(s) = \frac{\frac{1}{R_1 C_1}}{s + \frac{1}{R_1 C_1}} \tag{4.20}$$

Recall that this transfer function is Transformation of the output signal with respect to the input signal: $V_0(s) = LP_{1st} V_i(s)$ in the frequency domain.

By replacing s with $i\omega$, where $\omega = 2\pi f$ in equation 4.20 it is obtained information about the attenuation and phase delay at a given frequency, introduced to a signal after being applied to this particular circuit. One can understand this transformation in the frequency domain like this: We have seen that our particular PWM signal can be described by a combination of sinusoidal signals with unique frequencies. Each unique frequency has a weight. In our particular case, the weights get progressively smaller with increasing frequency. In the frequency domain, a particular frequency is a point whose abscissa is that particular frequency and it's ordinate is the weight, as so the PWM signal is characterized by a discrete set of points that start at the origin with 5D magnitude and then get increasingly smaller. The representation of the low pass filter in the frequency domain is exactly like described previously, except it has a point in every frequency as opposed to our signal that has a discrete spectrum and also, more importantly,

each weight is a complex number. When the PWM signal passes throughout the filter they mix. In the frequency domain the mixing is a simple point by point weight multiplication - each point in the PWM spectrum is multiplied by the its corresponding weight value in the spectrum of the low pass filter. The output signal is the discrete set of values that result from the multiplication. The set is now made up of complex numbers. In the complex plane, these numbers are vectors with an angle, θ and a magnitude, A given by:

$$\Theta = \tan\left(\frac{Im}{Re}\right) \quad , \quad A = \sqrt{(Re)^2 + (Im)^2} \tag{4.21}$$

where Im is the imaginary and Re is the real part of the number.

This quantities are the the Phase and Amplitude of the signal at each frequency.

The ripple voltage was estimated by using equation 4.20 and the first 10 terms of equation 4.18. Firstly, it is determined coefficients of the first 10 harmonics($a_0, ..., a_{10}$). Then, it was computed the low pass coefficients at each of this 10 frequencies. Followed by the multiplication of each coefficient by their correspondent low pass coefficient counterpart affected by their phase difference. Then it is determined the resultant Amplitude value. Finally, ripple voltage is then obtained by adding all the 10 amplitudes into a single value. The value obtained is the peak-to-peak noise imposed on some DC signal. This procedure may be repeated for different RC values as well as PWM frequencies, until it is found satisfactory ripple voltage that agrees with the requirements, while keeping an eye must be kept over the rise time of the filtered signal.

For a simple system like the low pass filter, an expression for the rise time may be easily obtained by exciting it with a step function. The response of the system, in the time domain, can be shown to be:

$$V_O(t) = V_i\left(1 - e^{-\frac{t}{R_1C_1}}\right) \tag{4.22}$$

We find the rise time by solving equation 4.22 with respect to time by setting the output voltage at 90% of the value of the input:

$$e^{-\frac{t}{R_1C_1}} = 0.9 \rightarrow t = -R_1C_1\ln(0.9) \tag{4.23}$$

Thus, the RC product is found to be directly proportional to system's rise time. Figures 4.23 and 4.24 provides a graphical simulation of the filtered PWM signal with different RC values. With all this in mind, it can be calculated that, in order to have a ripple voltage of about 1.5 mV with a 9 bit PWM, we would end up having a system with rise time of about 40 ms.

The Low pass filter system can be upgraded to a Second Order Low pass filter in order to improve the ripple attenuation while maintaining a selected rise time. This can be visualized by the amplitude plot of two low pass filters. Due to the logarithmic nature of the bode diagrams, the resulting second order low pass filter is given by the sum of the individual first order low pass filters, thus, for equal first order stages, the attenuation can be increased to -40 dB per decade without a substantial decrease in the cut-off frequency.

Progressing to a Second Order Low Pass Filter, there are essentially two options, either add a second cascading RC stage to the previous circuit or design a Sallen-Key Low Pass Filter. The Sallen-Key architecture is used to implement second order active filter meaning that it requires an OpAmp. This is not at all an inconvenience since the system already requires an Amplifying Stage to shift the voltage from $0-5$ V to ± 10 V. The Sallen-Key topology may be employed to design the filter with certain specifications. From it it is possible to design Butterworth, Bessel and Chebyshev filters with well tuned

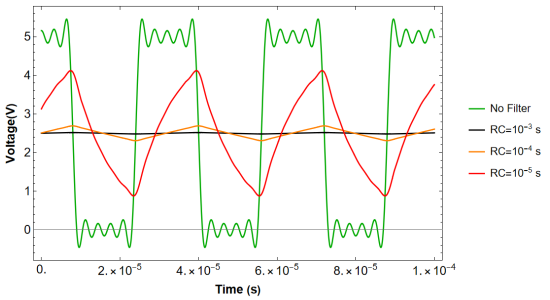

Figure 4.23: Simulation of the ripple voltage of the PWM signal for diferent time constants. It is also shown the PWM signal calculated with equation 4.18 that was truncated at n=10. The PWM wave has a 0.5 duty cycle and 31.25 kHz frequency resulting from a 9bit PWM signal with the prescaler set to 1. Obtained via the Mathematica software.

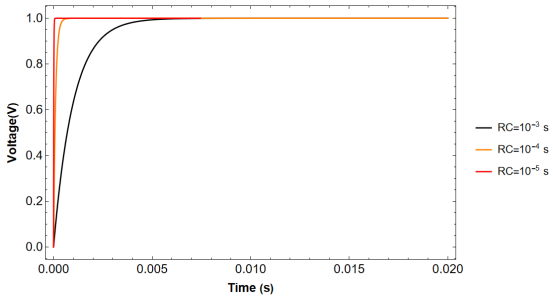

Figure 4.24: Simulation of the rise time of a first order low pass filter with different time constant values. The system is responding to a unit step. nontheless it is indicative of the time it takes to reach steady value. The ripple voltage should be superpositioned with this signal in order to match the real world behaviour. As can be seen, for these time constant, the 125 ms rise time constraint is easily achieved. Obtained via the Mathematica software.

characteristics such as a slow varying gain at smaller than cut-off frequencies, Q value and other filter parameters. For our purpose, we are only interested in the maximum attenuation possible while not compromising in the Rise Time of the Response of the system. Furthermore, with the need to offset and amplify the PWM filtered signal to the suitable range, it would be imposed a constraint in the Sallen-Key Low Pass Filter's cut-off frequency or another amplifying stage would be needed to perform the level shifiting, thus adding complexity and cost. Also, attention would be needed to ensure the stability of the OpAmp is not compromised, since the filter has a feedback loop.

For this reasons, adding another RC filter, if the required performance is achieved, seems like the better option.

The Transfer Function of the second order low pass filter made of two low pass filters is:

$$LP_{2nd}(s) = \frac{\frac{1}{R_1R_2C_1C_2}}{s^2 + s\frac{2}{\sqrt{R_1R_2C_1C_2}}\frac{R_1C_1+R_2C_2+R_1C_2}{2\sqrt{R_1R_2C_1C_2}} + \frac{1}{R_1R_2C_1C_2}} \tag{4.24}$$

If both stages have the same RC values then the rise time is unchanged, and the attenuation is squared compared to the first order low pass filter.

After repeating the same numerical simulation routine as earlier described, it was reached a good compromise between all the requirements. It was chosen $R_1 = R_2 = 1$ kΩ and $C_1 = C_2 = 1$ μF and a 9 bit PWM that has a frequency of 31,25 kHz, which equates to a voltage resolution of about 40 mV. It is estimated the system to have a ripple voltage of less than 0.1 mV and a rise time of about 25 ms, meaning the output voltage will reach 99% of the steady state value after about 5 characteristic time which is corresponds to the 25% of a line figure that was deemed satisfactory.

Note the choice of resistor to capacitor values. It was deliberately chosen small resistor values to prevent the aggravation of the input offset Voltage, present in the AmpOp of the next stage, due to leakage currents from both the Arduino pin and and the OpAmp's biasing current.

4.3 Software

There we developed two distinct classes of software for this project. One to program the microcontroller to directly instruct the physical hardware to perform specific tasks, which is known as firmware.

And another to run in a more general type computer to serve as an Interface to the human user, where is provided a GUI (Graphical User Interface) with buttons and other types of objects for the user to send commands to the hardware layer, as well as sections to display information.

4.3.1 Firmware

Perhaps the main reason for the popularity of the Arduino as a learning platform is the seamless experience programming the device with both libraries and user friendly programming environment that facilitates the direct Hardware manipulation. In the background, the Arduino configures the microcontroller to some default state, and the user specifies what parameters it wants to change in the writing process of the sketches. The Arduino is programmed in C. To program the microcontroller it must be kept in mind that the registers control the hardware. Registers are, simply put, special places in the microcontroller's memory, that, unlike general memory to save data, they hold information about a specific hardware process. When it is written a series of bits (most registers in the Atmega 328p are 8-bit wide) the hardware will perform some task, whether turning a pin HIGH or starting an ADC conversion or even hold some information, say about the state of some process. The CPU fetches the information stored in these locations every time the program loops. All memory slots, where data is stored, have a number associated with them, called address, that uniquely defines them. Each register has an address number that maybe consulted in the microcontroller Datasheet.

C is a language that allows direct memory management. A pointer is a variable that stores the address of another variable. This allows us to pragmatically store data in a particular register by referencing it's address value. Accessing or manipulating the value stored at this specific address is what's called dereferencing the pointer variable.

Figure 5.1 contains a few lines of code that serve as an example on how to do direct memory access. Consulting the Datasheet, we find the address value of, for example purposes, the PORTB register is 0x25 in hexadecimal, which corresponds to the number 37 in decimal. This register sets the logic level of D8 to D13 DIGITAL pins. Each pin is set to HIGH, if the bit value is 1 or LOW if the bit value is 0, in each pin when they are configured as output pins. On the first line, we create pointer variable that holds the address of an 8 bit value. We give it the arbitrary name "pointerPORTB", to be more human readable. On the second line, there is a statement, in which we dereference the pointerPORTB and assign it a value, in this case we are explicitly using the binary notation to be clearer what bits of the register or, in the words of the use of this special space of memory, the pins which are turned on or off. Equivalently, it could have been assigned a decimal value between 0 and 256 in order to convey the same message, and hexadecimal even. On the third line, we condense the two above lines into one. The volatile keyword is to signal the compiler to make no optimizations on this variable.

```
volatile uint8_t* pointerPORTB = 37;
*pointerPORB = b11111111;
 *((volatile uint8_t*) 37) = b11111111;
#define PORTB *((volatile uint8_t*) 37);
PORTB = b11111111;
```

Figure 4.25: Snippet of code that show how to access a register in the microcontroller and change the data stored there.

```
80  int analogRead(uint8_t pin){
81
82      uint8_t low, high;
83      // ADSC is cleared when the conversion finishes
84      while (bit_is_set(ADCSRA, ADSC));
85
86      // we have to read ADCL first; doing so locks both ADCL
87      // and ADCH until ADCH is read.  reading ADCL second would
88      // cause the results of each conversion to be discarded,
89      // as ADCL and ADCH would be locked when it completed.
90      low  = ADCL;
91      high = ADCH;
92  #else
93      // we dont have an ADC, return 0
94      low  = 0;
95      high = 0;
96  #endif
97
98      // combine the two bytes
99      return (high << 8) | low;
100 }
```

Figure 4.26: Snippet of code of the analogRead() function in the Arduino's library.

This is, in practice, how to program the microcontroller to our liking, however it is very cryptic for a human to properly read the left side of the statement and so to make it more speedy and human friendly, we may program a macro that substitutes the left side of the statement to PORTB, and so in order to write to the register all is needed is to write the statement shown in line 4. To sum it all up, the PORTBpointer holds the location in memory of the register and then, by dereferencing the PORTBpointer variable, it is written a particular number to this location, that, in this particular case, turns HIGH or LOW the digital pins of the microcontroller.

Each process or feature in the microcontroller has some collection registers associated with them and the manufacturer specifies their use case. We have to keep in mind that although a microcontroller is an incredible versatile device, it is still bounded by it's non-configurable logic blocks, they cannot be configured to do anything other than what they were designed for, they can be programmed via software, but on the hardware level their capabilities are fixed.

A microcontroller may be looked at as a very simple computer. It's CPU possesses a single core, meaning a just few instructions are executed at a time per clock cycle, sequentially. It is common to refer to this pipeline of instructions, a single thread. Microcontroller's executing program is made of a never ending function where the program is looped over and over. This statement is fundamental performance wise, since the Arduino will not be able to attend to some task while it's busy performing another. The only exception to this rule is when Software Interrupts occur. The program will jump from an operation that is processing at the time of the interrupt to complete the task specified in the Interrupt Routine (this is called servicing the interrupt) however, the other task will be left idle until the program returns from the Interrupt Service Routine (ISR). With that in mind, the program must be carefully written in order to not hinder the performance of time critical tasks, such as, for instance, the driving of the motor as it may have catastrophic consequences to the operation of the AFM.

To ensure the best possible performance out of a single core processor, the program must be kept short, in order to be traversed hastily and, most importantly, it should be avoided blocking statements such as while loops or large iterations. This is brought up due to the way Arduino wrote their libraries. Since they are intended for beginners and fast experimentation/ general use, they abstracted a lot of the details on simple function calls, like analogRead(pin number). This is a very simple command that the user calls to start an analog to digital conversion on a specific pin. It's a very simple and intuitive from the user's point of view and it doesn't require an understanding the fundamental process behind an analog to digital conversion let alone what registers are needed to be configured in order to perform the task. Upon further inspection of the source code of the ADC conversion example (a great advantage of building on top of a open source project!) we find the snippet of code shown in figure 5.2. The code reads into human language as such: "While the bit that signals for the end of a conversion is 0 repeat" which effectively means that the function is blocking the execution of the program while a conversion is underway. Although a conversion takes about 13 cycles whose clock is $\frac{1}{128}$ of external clock frequency (0.1 ms), and in a human time perception is quite fast, in the CPU point of view it is a lot of time that the program just sits there and waits. And since the we may collect a few samples it will build up. This get even worse the further complexity is added with different tasks such as reading the serial line/ writing to the serial line or coding delays that, if written as blocking statement, will inevitably make the program not very time efficient.

For this reasons it was chosen to write new firmware that make the whole process more performance oriented. This is not to say that the Arduino libraries were all completely discarded, since it would be unfit to utilize the Arduino platform and not take some advantage of their core library's utility.

The proper solution to the blocking functions is to utilize the very opportune hardware flags that are

implemented on the microcontroller. The hardware flags are bits that the hardware sets/clears in order to signal the state of an operation. The method here is instead of waiting for a given operation to finish, we check for the state of that hardware signal. If it is still ongoing the program continues to traverse. If not, then we execute whatever logic we design to be associated with that task.

For the time critical tasks, the hardware flags are configurable to trigger their corresponding Interrupt Service Routine and make sure there is little delay between the trigger and the action.

Having said that, the optimized code that was developed was targeted at the Analog to Digital Conversion, the Timers/Counters that are needed in order to generate the 16 bit PWM signal and to accurately time the motor steps, as well as the Watchdog Timer whose function is going to be covered up next and finally the UART protocol that allows the Arduino and the Computer Interface to communicate in parallel (Asynchronously). The only Arduino library used was the SPI implementation as it was considered to be acceptable to have a blocking statement, since it is only used when changing the gains and it will be made at the human reaction speed and for that reason won't take up much of the running program's time.

The Watchdog Timer's function in this project is be used to safeguard the Arduino from either being, accidentally, struck on a task or an eventual malfunction from the Interface side. In order to monitor each other's active status, it was coded an Handshake between the two where, periodically, the Arduino and the Interface will send a message to each other that signal's precisely their smooth operation. On the Arduino side, as soon as the handshake message is received, the Watchdog Timer is reset and the program continues normally. In the eventual case the Interface stops sending it's status, the Watchdog timer will run out and a panic action is triggered. The panic action is to retract the tip from the sample by a large distance and then reset the microcontroller. At the beginning of the program, before stating the infinite loop, the microcontroller is deliberately stuck on a loop that only ends when the Interface is turned back on and sending it's status.

On the communication side of things, since there is a need to send messages and data back and forth between the Arduino and the Computer Interface, it was structured a message template that organizes the information and doubles down as a checksum to verify the message integrity (no error occurred in the data transmission):

$$C : C : CC;$$

where C is the equivalent with an 8 bit long character. The command is divided into 3 parts that are separated by the ":" character. The first character is indicative of the task being executed, for instance if the command is to be addressed to the motor, this character is an "M", if the user wants to alter the gains, this bit is a the ASCII representation of "G". The second part of the message is the option of a command, in case of addressing the motor, there are two options, either the motion is upwards "U" or downwards "D". The third part contains numerical data information, for instance if the Arduino is sending the result of an analog to digital conversion that is a 10 bit wide value, then it it split into two 8 bit characters and sent in this part of the command. Finally, the ";" character signals the end of a given message/command.

It was designed to be as short as possible to not prolong the time of transmission but still be reliable.

The microcontroller is responsible for the execution of the Auto-Approach algorithm in order to optimize the it's execution time. When the start command is received, this task is given priority due to it's sensitive nature however, the other tasks are not blocked. The Arduino is always aware of the chosen setpoint and has all the tools to complete the task without the need to consult the computer interface. It reads the voltage of the setpoint, compares it with the chosen value, adequates the step size and performs the step until it either reaches contact or receives a stop condition from the interface issued on behalf of

the user.

As a final note, the Arduino does not process any of the raw data that it collects with the ADC. It is left for the user interface, which has better resources, to conduct this job.

4.3.2 User Interface Software

Figure 4.27 is a screenshot of the Graphical User Interface built for the Control Module. It is a product of both inspiration from other commercial software and user experience insights from the experienced engineers of the AFMaRT laboratory.

Let's start from the graphical presentation: The layout of the interface was designed to be inherently intuitive to the user as it coordinates the steps to be taken in order to start an image acquisition, form the moment the instrument is turned on. This sentence means that the the Interface is organized with the natural order of the tasks the user must complete in order to start acquiring an image.

Western language rules prime our brain to search for information from left to right, top to bottom. This intuition determined the layout of the Interface. Beginning at the top left and making our way right and bottom. This was the principle it was incorporated here. As we can see starting at the top left there is a widget that resembles the physical appearance of the PSD system. The shaded area is representative of the active area of each quadrant, the red circle is the representation of the laser beam spot and the black region is representative of the outer bounds region of the psd system. This widget provides a familiar visual guide to easily align the laser spot with the geometrical center of the photodetector active area. The red circle is moves along the active area as the PSD is in its linear region. Note that this is is not in agreement with the actual physical behaviour. It is an exacerbated motion. The actual laser spot only moves very incrementally in the linear region, recall figure 2.26 . In this widget the red dot reaches the limit of the active area when the signal reads ± 10 V. In reality, the positioning system saturates much earlier, but this choice was made to be a visual aid, and if were an accurate simulation of the real behaviour, the user could not tell, just by looking at the widget if it was close to the center or not. Typically, user will be looking at the monitor screen and it may be separated by some distance, thus it is favorably to have it like this.

Below the photodetector widget, there is a display of the voltage readings of both the Horizontal signal and the Vertical signal in order to affirm the visual guide. The user has the option of choosing which physical channels it plugged the signals. This is added for flexibility. The interface receives all the measurements of the Arduino, converts the conversion value into the appropriate voltage value and then displays them. The update rate of the numerical displays are in agreement with the typical human reaction time. Followed by the expand button that enlarges the photodetector widget when pressed, which is extremely useful when the user has the computer monitor a few meters away from the AFM.

To the right of this section we have the setpoint control at the top and the coarse motion section in the bottom. In the setPoint control the user is prompted to select the control's variable channel to be used. Then he inserts the desired value of the variable that can be specified in volts or milivolts, as the it was implemented a functionality to handle the voltage specification in the form "1.2v" or "100m"(number plus order of magnitude as either the letter v or m, respectfully), for a fast typing experience. The selected value will be used as both the Auto-Approach variable and the signal fed to the controller. The widget where the desired setpoint value is inserted has the extra functionality of

In the coarse motion section, the user is presented with clickable buttons to control the motor. First, there is the pair "Step Approach" and "Step Retract", which commands to approach the sample and tip or separate them apart, respectfully. Both steps have equal lengths that are, by default equal to 10 ms

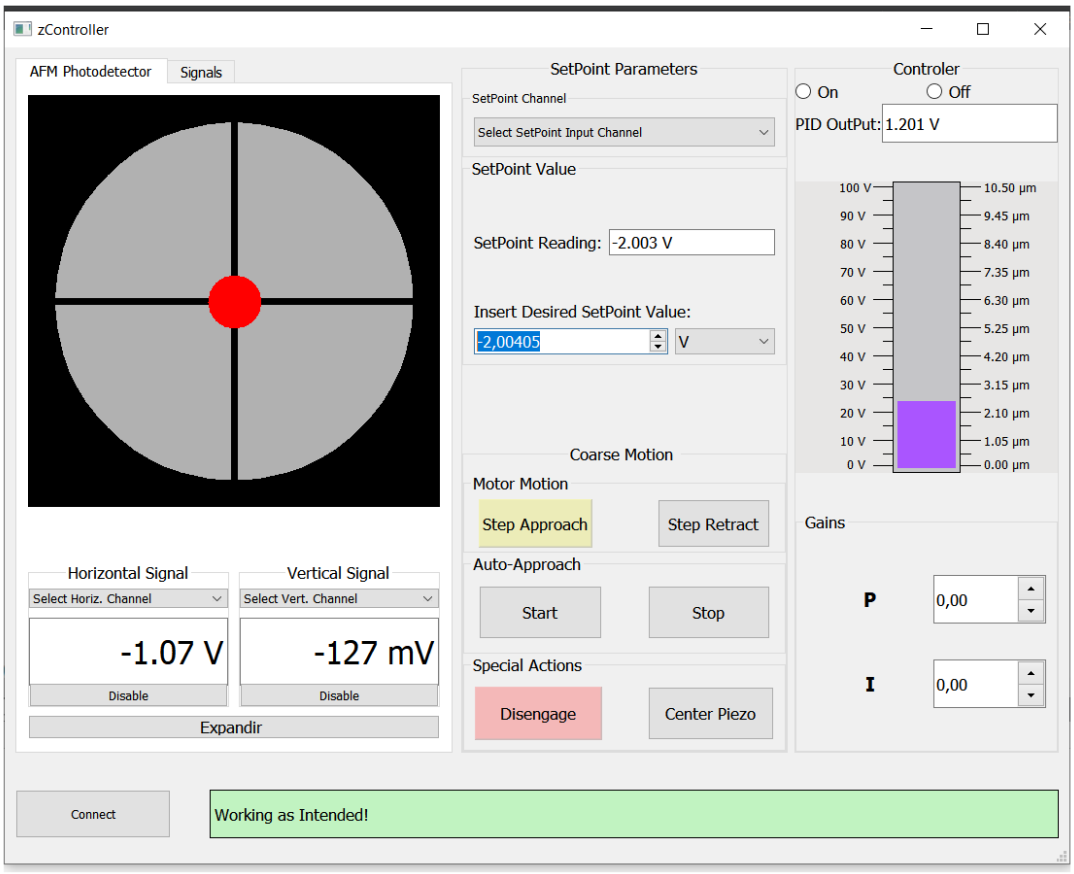

Figure 4.27: Graphical user interface that accompanies the developed hardware. It was used the QT framework to build and design the specialized widgets that display the information to the user.

of current into the motor which roughly results in 0.5μm of length, as stated in section 4.1.1. Also, the Step Approach button has a yellow color to signal precaution to the user. Below are the Auto-Approach buttons. They effectively start and stop the Auto-Approach algorithm sequence. And Finally, there were implemented buttons that perform special actions. The Disengage button pulls the tip and sample apart a macroscopic distance, 500 ms step. This is meant to be either a Panic Button or a Safe Distance for switching samples, cantilever or just end the use of the AFM for the day. The ”Center Piezo” button will place the tip and sample at half a piezoelectric element range of each other, at a certain setpoint, it is meant to be used after establishing contact with the Auto-Approach algorithm. This feature is useful because the image may be started with half a piezoelectric length of margin.

At the right of the Interface, there is the Controller Section. Here the user monitors and controls the Control System. At the top, there are the ”On/Off” buttons that control the multiplexer at the output of the PI stage. Below there is an animated display representative of the Piezoelectric (stretch) displacement by increasing the size of the purple vertical bar along the scale. This widget does not have it’s update rate throttled since it provides a qualitative feedback to the user on whether the PI gains selected are causing the piezo to move too fast/ slow. Last but not least, the Gain Section where the user tunes the PI controller. The selection is made on a scale of 0 to 1000. This was the implemented scheme as it is deemed not relevant from user’s point of view to display the exact value of the gain.

At the bottom of the Interface panel there is a Connect button meant to be used to establish connection to the Arduino either when it resets or is powered off. And also there is a status Label that displays some warning messages to the user and guide him in a troubleshoot process.

Chapter 5

Conclusion

To conclude, the work related to the implementation of the vertical axis control system and the other enhancements carried out and discussed in this dissertation resulted in a box and a computer program. The box contains a PCB that is connected to an Arduino. In the PCB lies the PI controller, the interface circuitry and the other complementary circuitry needed for a successful practical implementation.

As a whole, they enable the user to perform several tasks such as image formation at constant force, the adjustment of the interaction magnitude and the PI gains via software, execute a fast, accurate and automatic approaching of the tip and sample and they let the user manually adjust this distance with a regular and adequate step. The approaching algorithm brings these two surfaces from a distance of a few millimeters to just a handful of micrometers within the reach of the Z piezoelectric element. The manual step adjustment provides a tool to avoid the saturation of the PI signal or to engage/disengage the interaction. Both of these actions are made through the use of the underlying motorised mechanism.

Finally, using the Arduino firmware here developed we are able to read and transmit the data to the computer interface that displays information about important signals in an appropriate manner, thus facilitating the alignment of the laser in the position sensitive photodetector and the monitoring of the output of the controller while the AFM is operating.

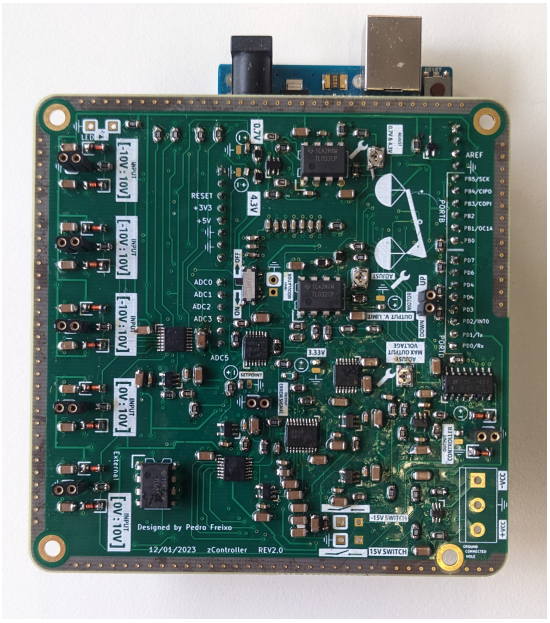

Figure 5.1: PCB that was designed in this project.

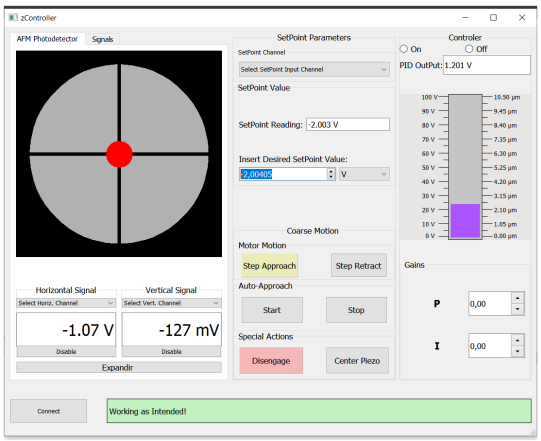

Figure 5.2: Graphical user interface designed in this project.

Bibliography

[1] Daniel Y. Abramovitch et al. "A tutorial on the mechanisms, dynamics, and control of atomic force microscopes". In: *2007 American Control Conference* (2007). DOI: 10.1109/acc.2007.4282300.

[2] Toshio Ando. "Control techniques in high-speed atomic force microscopy". In: *2008 American Control Conference* (2008). DOI: 10.1109/acc.2008.4586984.

[3] G. Binnig, C. F. Quate, and Ch. Gerber. "Atomic Force Microscope". In: *Physical Review Letters* 56.9 (1986), pp. 930–933. DOI: 10.1103/physrevlett.56.930.

[4] Hans-Jürgen Butt, Brunero Cappella, and Michael Kappl. "Force measurements with the atomic force microscope: Technique, interpretation and applications". In: *Surface Science Reports* 59.1-6 (2005), pp. 1–152. DOI: 10.1016/j.surfrep.2005.08.003.

[5] Gerhard Meyer and Nabil M. Amer. "Novel Optical Approach to Atomic Force Microscopy". In: *Applied Physics Letters* 53.12 (1988), pp. 1045–1047. DOI: 10.1063/1.100061.

[6] Fabio L. Leite et al. "Theoretical models for surface forces and adhesion and their measurement using atomic force microscopy". In: *International Journal of Molecular Sciences* 13.12 (2012), pp. 12773–12856. DOI: 10.3390/ijms131012773.

[7] Toshio Ando, Takayuki Uchihashi, and Takeshi Fukuma. "High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes". In: *Progress in Surface Science* 83.7-9 (2008), pp. 337–437. DOI: 10.1016/j.progsurf.2008.09.001.

[8] Maja Dukic et al. "Digitally controlled analog proportional-integral-derivative (PID) controller for high-speed scanning probe microscopy". In: *Review of Scientific Instruments* 88.12 (2017), p. 123712. DOI: 10.1063/1.5010181.

[9] A. Siria et al. "Viscous cavity damping of a microlever in a simple fluid". In: *Physical Review Letters* 102.25 (2009). DOI: 10.1103/physrevlett.102.254503.

[10] Andrew J. Fleming, Yik Ren Teo, and Kam K. Leang. "Low-order damping and tracking control for Scanning Probe Systems". In: *Frontiers in Mechanical Engineering* 1 (2015). DOI: 10.3389/fmech.2015.00014.

[11] Miguel V. Vitorino et al. "Direct measurement of the capillary condensation time of a water nanobridge". In: *Scientific Reports* 8.1 (2018). DOI: 10.1038/s41598-018-32021-0.

[12] T. Sulchek et al. "Characterization and optimization of scan speed for tapping-mode atomic force microscopy". In: *Review of Scientific Instruments* 73.8 (2002), pp. 2928–2936. DOI: 10.1063/1.1488679.

[13] Hamed Sadeghian et al. "High-throughput atomic force microscopes operating in parallel". In: *Review of Scientific Instruments* 88.3 (2017), p. 033703. DOI: 10.1063/1.4978285.

[14] Bharat Bhushan. *Springer Handbook of Nano-technology*. Science Press, 2009.

[15] Chi-Tsong Chen. *Analog and digital control system design: Transfer-function, state-space, and algebraic methods*. Oxford University Press, 2006.

[16] García Ricardo. *Amplitude Modulation Atomic Force Microscopy*. Wiley-VCH, 2010.

[17] John Philip McKelvey. *Solid state and semiconductor physics*. Krieger, 1986.

[18] H. P. Myers. *Introductory solid state physics*. Taylor amp; Francis, 2009.

[19] H. J. Pain. *The physics of vibrations and waves*. John Wiley, 2008.

[20] M. L. Meade and M. L. Meade. *Lock-in amplifiers: Principles and applications*. P. Peregrinus, 1983.

[21] Mario Rodrigues. "Bringing light into the nanoworld: What can you do with an atomic force microscope on top of your synchroton radiation sample holder?" PhD thesis. 2010.

[22] Arthur Vieira. "DESENHO E CONCEÇÃO DE UM MICROSCÓPIO DE FORÇA ATÓMICA". PhD thesis. 2017.

[23] Miguel Vitorino. "Development of a Force Feedback Microscope". PhD thesis. 2020.

[24] *MCP41HVX1 [DATASHEET]*. MCP41HV31-103E/ST. Rev. D. Microchip Technology. 2015.

[25] *ATmega328P [DATASHEET]*. ATMEGA328P-PU. Microchip Technology. 2008.

[26] *Understanding Digital Potentiometer Resistor Variations*. AN1080. Microchip Technology. 2007.

[27] Thor Labs. *Thor Labs Piezoelectric Tutorial*. URL: `https://www.thorlabs.com/newgrouppage9.cfm?objectgroup_id=5030`.