Open-Loop Transient Atomic Force Microscopy

by

Seyed-Mahdi Olfat

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Master of Applied Science

in

Systems Design Engineering

Waterloo, Ontario, Canada, 2023

© Seyed-Mahdi Olfat 2023

Author's Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

The Atomic Force Microscope (AFM) is an instrument for measuring, in fact “seeing”, phenomena at nanoscale (10^{-9} m) and all the way down to the atomic scale ($< 10^{-10}$ m). It was borne out of a need to observe physical reality below the resolution of optical microscopes. Invented in 1986 by Binnig, it has aided scientists, researchers, and engineers spanning many scientific and industrial domains. The typical sensing apparatus of the AFM is a very sharp tip (a few atoms wide) attached to the free-end of a fixed-free micro-beam. The tip is brought close to the desired specimen to initiate localized force interactions between the top-most atoms of the specimen and the bottom-most atoms of the tip. The tip-sample coupled system introduces a relative shift in the deflection of the microcantilever, this shift is interpreted as the magnitude of the interaction force and used for topographical mapping as well as mechanical/electrical/thermal characterization of a single point on the specimen. By recording the cantilever deflections while laterally scanning the sample, an area is “imaged”.

According to resonant sensing theory, the sensitivity of the deflection of an oscillating cantilever beam driven at its resonant frequency is increased by a factor proportional to its quality factor (Q). As such, it is very common for an AFM cantilever to be driven at (or near) its first resonant mode in order to increase the Signal-to-Noise ratio (SNR). Whilst away from the sample, the steady-state response is perfectly harmonic with a force-response phase difference of 90 degrees: the very definition of resonance. However, near the sample, the response becomes anharmonic, nonlinearly modulated by the tip-sample interactions. This anharmonic response needs to be demodulated to quantify the interaction forces. The deviation from the harmonic response is only instantly described if the instantaneous frequency and amplitude are known. Alas, this is not possible. System design engineers are confronted with the ultimate compromise, namely, the Heisenberg uncertainty principle. More specifically, the energy spread of any time-frequency transformed signal forms a rectangle in the time-frequency plane, this "Heisenberg resolution box" has a minimum surface area of $1/2$ that ultimately limits the attainable time-frequency resolution.

This thesis proposes a framework for a holistic approach to an open-loop bottom-up AFM system design. Design decisions and compromises are discussed and analyzed based on the desired requirements such as the SNR, the minimum acceptable noise floor, the demodulation scheme, and the maximum/minimum response times. An electrothermally driven and piezo-resistively sensed single-chip AFM (sc-AFM) is used for testing and verification. System Identification is carried out in the frequency domain to characterize the noise and establish the best linear approximation (BLA) transfer function of the AFM cantilever. For imaging, the AFM cantilever is continuously driven by an excitation signal while the cantilever deflection signal is sampled and filtered digitally. An analysis window is selected to adequately capture both transient and steady-state responses of the cantilever deflection signal. Moreover, a complete digital processing pipeline is proposed and implemented using the STFT for nonlinear time-varying spectral signal processing. Finally, imaging results based on amplitude-modulation AFM (AM-AFM) are demonstrated.

Acknowledgments

I would like to acknowledge the support and guidance of Professor Eihab Abdel-Rahman and Dr. Alaaeldin Ahmed. I would also like to thank Professor Raafat Mansour, Dr. Neil Sarkar, Geoff Lee, Duncan Strathearn, David Morris, and Zhenle Cao for the fruitful discussions and interactions that have influenced this report.

I am deeply indebted to my family, Mom, Dad, Ehsan, and Amy for their unconditional love and support.

Table of Contents

Author's Declaration ii
Abstract iii
Acknowledgments iv
List of Figures viii
List of Tables ix
List of Abbreviations x
List of Symbols xi
1. Introduction 1
1.1 Motivation 1
1.2 Objective 2
1.3 Outline 2
2 Literature Review 3
2.1 Overview 3
2.1.1 Probe Model 5
2.1.2 Thermomechanical Noise 6
2.2 Tip-Sample Interactions 7
2.2.1 Models of Contact 8
2.3 State Estimations 9
2.3.1 In-Phase-Quadrature-Phase 9
2.3.2 Kalman Estimator 10
2.3.3 Lock-in Amplifiers 10
2.3.4 Lyapunov Estimator 11
2.4 Feedback Control 11
2.5 Advanced Techniques 11
2.5.1 Multifrequency 11
2.5.2 Q Control 12
2.5.3 Information Theory 12
3 Signals 13
3.1 AFM Signals 13
3.1.1 Noise processes 14

3.2 Systems Analysis 14
3.3 Sampling Theory 15
3.4 Oversampling 15
3.5 Fourier Analysis 15
3.5.1 Mathematical Definitions 16
3.5.2 Convolution Theorem 17
3.5.3 Time-Shift (Frequency-Modulation) Theorem 17
3.5.4 Frequency-Shift (Time-Modulation) Theorem 17
3.5.5 Parseval's (Power) Theorem 17
3.5.6 Rayleigh's (Energy) theorem 17
3.6 Digital Signal Processing 17
3.6.1 Time-Limited Signal 18
3.6.2 Band-Limited Signal 21
3.6.3 Fast Fourier Transform (FFT) 22
3.6.4 FFT Convolution 22
3.6.5 Zero-padding 22
3.7 Time-Frequency Distributions 23
3.7.1 Short-Time (Windowed) Fourier Transform 23
3.7.2 Spectrograms 24
3.7.3 Time-Frequency Resolution 24
4 Design 25
4.1 Approach 25
4.1.1 System Block Diagrams 26
4.1.2 Open-loop vs. Closed-loop 28
4.1.3 Frequency Response Function 28
4.2 Model Parameters 29
4.2.1 Harmonic Oscillators 29
4.3 Excitation Signals 30
4.3.1 Real Harmonics 30
4.3.2 Random Multisine 30
4.3.3 Uniform Random-Phase Multisine 31

4.3.4 Periodic Noise Multisine 31
4.4 System Identification 32
4.4.1 Modeling 32
4.4.2 Best Linear Approximation 33
4.4.3 Identification Process 34
4.5 Signal Processing Pipeline 37
5 Results and Discussion 38
5.1 System Under Test (SUT) 38
5.2 System Identification 39
5.3 Imaging 40
5.4 Conclusions 42
5.5 Future Work 42
5.5.1 Multifrequency AFM 42
5.5.2 Input shaping 42
5.5.3 Feedback 43
References 44

List of Figures

Figure 1: Coordinate system of an AFM cantilever oscillating near the sample surface. 3
Figure 2: Rectangular window representations in time and frequency domains 18
Figure 3: Rectangular window magnitude (dB) and phase response 19
Figure 4: Ideal low-pass filter response 21
Figure 5: System and signals block diagram 26
Figure 6: Processing unit system block diagram with synchronous clocking 27
Figure 7: Simplified model of the system under test 27
Figure 8: Uniform random phase multisine 31
Figure 9: Random noise multisine 31
Figure 10: Frequency domain representation of the AFM 32
Figure 11: Digital signal processing pipeline for real-time estimation and of tracking spectral peaks 37
Figure 12: Magnitude and phase response of the BLA estimates for experiment 1 39
Figure 13: Magnitude and phase response of the BLA estimates for experiment 2 40
Figure 14: AFM image 41

List of Tables

Table 1: Mode constants for free and supported cantilever. 7
Table 2: Energy fields and forces based on the Hamaker constant and simple geometries. 8
Table 3: Spectral performance of common windows 20
Table 4: System-Under-Test specifications 38

List of Abbreviations

ADC	Analog-to-Digital Convertor
AM-AFM	Amplitude-Modulation Atomic Force Microscope
AFM	Atomic Force Microscope/Microscopy
BLA	Best Linear Approximation
CPU	Central processing unit
DAC	Digital-to-Analog Convertor
DC	Direct Current
DMA	Direct Memory Access
DFT	Discrete Fourier Transform
DTFT	Discrete Time Fourier Transform
FFT	Fast Fourier Transform
FPGA	Field-Programmable Gate Array
FN-AFM	Frequency-Modulation Atomic Force Microscope
FRF	Frequency Response Function
FS	Fourier Series
FT	Fourier Transform
GPU	Graphical Processing Unit
MCU	Micro Processing
NC-AFM	Non-Contact Atomic Force Microscope
NMR	Nuclear Magnetic Resonance
OLA	Overlap-Add
RMS	Root-Mean-Square
SPM	Scanning Probe Microscopy/Microscope
STFT	Short-Time Fourier transform
STM	Scanning Tunneling Microscopy/Microscope
SUT	System Under Test
TM-AFM	Tapping-Mode Atomic Force Microscope
VDW	Van Der Waals

List of Symbols

Symbol	Description
$\mathbb{Z}$	integers
$\mathbb{N}$	natural numbers: positive integers including 0
$\mathbb{N}^+$	natural numbers: positive integers not including 0
$\mathbb{R}$	real numbers
$\mathbb{C}$	complex numbers
$\mathbb{C}^N$	complex set of finite size N
$\langle f, g \rangle$	inner product
$\|f\|_2$	L2 norm
$\|f\|_1$	L1 norm
$\boldsymbol{\ell^2}(\mathbb{R})$	finite energy (square integrable) signals
$f * g$	convolution of functions f and g
$\tilde{f}$	estimation of f
$\overline{a}$	vector/matrix transpose
$\dot{f}(t)$	time derivative $\frac{df}{dt}$
$< f_N >_M$	time average of the signal f periodic in N observed over M periods
$E\{f\}$	expected value of f. For a discrete-time function $f_N[n]$ that is periodic with asymptotically ergodic steady state response ($M \to \infty$ observations), the expectation may be written as the time average over the observations instead of an ensemble average over all samples: $E\{f_N\} =< f_N >_M = \lim_{M\to\infty} \frac{1}{N(M+1)} \sum_{m=-M}^{M} \sum_{n=m}^{m+N-1} f_N[n]$

1. Introduction

Optical microscopes provide an amazingly rich view of physical phenomena occurring in the range of $10^{-6}m$ (micrometer or micron for short). However, the process of guiding and focusing reflected light (or an electromagnetic wave of any kind) to reconstruct the topography of the reflective surface has an ultimate limit that is proportional to the wavelength of the light used. For optical microscopes, this is known as the Abbe diffraction limit. A modern optical microscope that uses visible light (wavelength of $\sim 500nm$) the best that can be achieved is about $180nm$. Other wavelengths have been exploited to achieve better spatial resolution, but they are all diffraction-limited systems.

Scanning Probe Microscopes (SPM) are a different class of imaging systems that are not diffraction limited. SPMs rely on the local interactions of a sharp tip (a few nanometers) near the surface. The first such instrument, the Scanning Tunneling Microscope (STM), was published in 1981 by two physicists Gerard Binnig and Heinrich Rohrer [1]. They were awarded the Noble Price in Physics five years later in 1986. That same year, they published their second SPM-based microscope, the Atomic Force Microscope (AFM) [2]. Whereas the STM requires the sample to be conductive in order to maintain a tunneling current to the sample, the AFM relies only on the local molecular/atomic surface forces, hence the sample does not need to be conductive. Further, because the AFM is effectively a force sensor, it may be designed to limit the impact it imparts onto the surface, allowing it to scan soft materials, whether biological like skin cells, or artificial like polymers.

Over the next three decades, many new methods of SPM "imaging" were presented and very active branches of engineering and scientific research morphed. Today, the AFM is an indispensable tool for many global industries. Given the immense interest in the AFM both as a discovery tool for scientists and a manufacturing tool for the semiconductor industry, it is hypothesized that the AFM as an imaging system will continue to be engineered from first principles to achieve imaging at the nanoscale as fast and as detailed as physically possible. This report documents the engineering investigation and characterization of an open-loop approach to AFM transient and steady state operations.

1.1 Motivation

The AFM instrument has proven to be an amazingly robust and accurate nanoscale measurement tool. Researchers, scientists, and engineers continue to rely on scanning probe microscopy for being able to run experiments and make fundamental claims about physical processes otherwise completely inaccessible to humans. As the most popular SPM, the AFM system performance will continue to be improved upon incentivized by the backing of researchers and industries. Faster and more accurate AFM systems and methods of operation are explored in academia and in the field. The main motivation and a central argument for this report is that the cantilever deflection signal is the most important signal in the system, it holds the greatest amount of the most sensitive and relevant information; hence optimization of its processing should be prioritized highly.

In a single cantilever drive cycle (approach and retract), the tip-sample interaction is concentrated around the time where the tip and the sample have minimum separation, hence, the transient response may be

extracted from a single cycle near the point of minimum approach. Processing power and strict timing should be maximized for the processing of the deflection signal in a single cycle of interaction, especially in the neighbourhood of the minimum separation distance.

By synchronizing the operations of a digital processing unit with an analog frontend that continuously and periodically drives and samples the AFM dynamic state, real-time digital signal processing algorithms based on the Fast Fourier Transform (FFT) may be used to track and operate the AFM in transient or steady state operations.

1.2 Objective

The objective of this investigation is to study the regularities and structure of tip-sample interactions and the resultant response of the cantilever in each approach-retract cycle. Each cycle will be decomposed and studied for the design of appropriate digital signal processing kernels.

With the knowledge of the (noisy) excitation, the frequency response function (FRF) is assumed to be smooth, system identification in the frequency domain can allow for the detection of weak nonlinearities as well as the characterization of noise in the system.

Finally, due to the smoothness of the FRF and based on the desired system states, the linearized model is used to estimate and track system states in the frequency domain.

1.3 Outline

Chapter 2 covers fundamental aspects of AFM operation and design as discussed in the literature. Its applications and advancements in recent years are surveyed. Common modeling techniques for simulation, system identification, and operation are discussed. System state estimation techniques such as the popular $I - Q$ harmonic analysis as well as statistically based approaches are covered. Finally, advanced imaging techniques and recent research trends are discussed.

Chapter 3 covers relevant signal processing fundamentals and establishes a common design language as it relates to the AFM system. Mathematical definitions are presented, and relevant theorems are proven.

Chapter 4 starts by breaking down the various system components that affect performance. Noise analysis is viewed as essential; noise characterization will allow for quantitative claims regarding image resolution and bandwidth. The model, system identification parameters, and digital signal processing algorithms are presented and discussed.

Experimental results based on the proposed design are presented in Chapter 5. Results are discussed and the performance of the implemented system is evaluated. Noteworthy findings or unexpected anomalies are discussed. The report concludes by suggesting future work related to each design stage is proposed with a brief justification.

2 Literature Review

2.1 Overview

The AFM is loosely comprised of four subsystems:

1. Probe: combination of cantilever and tip
2. 1D vertical scanner: cantilever driver, typically at the base of the fixed-free cantilever that may drive the cantilever into resonance and/or apply step changes.
3. 2D lateral scanner: controls the movement of the sample relative to the probe.
4. Sensor: observing cantilever deflection (position) or cantilever slope (velocity).

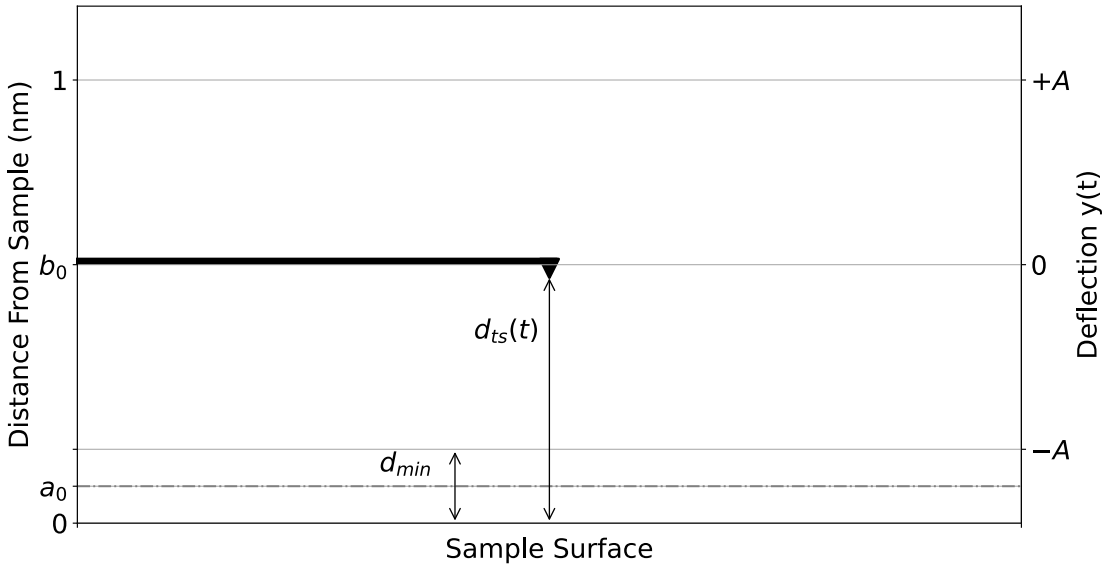


Figure 1: Coordinate system of an AFM cantilever oscillating near the sample surface.

Figure 1 illustrates the tip-sample coordinate system for an AFM probe operating in dynamic (resonant) mod For a forced harmonic drive at a frequency near a cantilever resonant mode, the cantilever oscillates periodically with an amplitude of A. The instantaneous deflection of the probe from its mean position is represented by $y(t)$. In one cycle of oscillation, the tip travels a total distance of $2A$, with the strongest interactions occurring at a maximum negative deflection of $y(t) = -A$. Alternatively, the instantaneous tip-sample separation $d_{ts}(t)$ is a positive measure of distance that is not only a function of the oscillation amplitude A and the mean position b_0, but also topographical variation on the sample, the subject of the desired AFM image. If an assumption may be made that the bias position of the cantilever does not change in operation, $d_{ts}(t)$ and $y(t)$ represent the same signal. However, because the mean cantilever position b_0 defines the deflection $y(t)$ coordinate system, $d_{ts}(t)$ is not observable in general. In fact, the exact bias position b_0 is a baseband signal that is not immune to environmental variations. More importantly, variations on b_0 may be caused by the tip-sample interactions, which is part of the desired estimation.

Maximum tip-sample interaction occurs at the point of minimum tip-sample separation. This is annotated as d_{min} in Figure 1. For a constant b_0, $d_{min} = b_0 - A$. Since b_0 is always greater than A by definition, d_{min} is a positive measure of distance and may change once every oscillation. In theory, given enough force, d_{min} may approach zero, but, as the tip-sample separation nears interatomic separation distances around 0.2nm, strong repulsive forces start to repel the tip. A typical threshold is 0.2nm. This is shown as a_0 in Figure 1. The interatomic distance plays an important role. It defines the nominal distance of separation between the tip and the sample where "contact" is made. Below the interatomic distance $d_{min} < a_0$ the forces experienced by the tip are repulsive. Above the interatomic distance, the tip and the sample are in an attractive regime of interaction [3].

The most widely adopted cantilever deflection sensing mechanism is a laser-based system that is calibrated to shine off the cantilever (opposite the tip) and into a sensitive quadrature photodetector. This may be called a velocity sensor as it effectively observes the slope of the cantilever deflection (vertical variations at the tip near the end of the cantilever) as time variations of the photodetector. More recently, strain sensors embedded within the cantilever have been used as deflection sensors and replace laser-based sensing altogether. This deflection sensing mechanism may be called position sensing as it tracks instantaneous vertical deflections along the cantilever.

In typical operation, the lateral scanner moves the tip or the sample in a raster-like pattern so a different point on the sample is sensed by the tip. The deflection sensor tracks the position of the cantilever as it is being scanned. Since the absolute shift in a cantilever's vertical position is a weak baseband signal, the cantilever is typically operated near (or at) its first flexural mode. This effectively increases the sensitivity of the cantilever to any perturbations by the mode's quality factor Q. This is an example of "resonant" sensing where a dynamic mode (mechanical resonance) is used to increase Signal-to-Noise (SNR) ratio. AFM resonance frequencies range from a few kHz to 100s of kHz. A mechanical shaker at the base of the cantilever is typically used to drive the cantilever into mechanical resonance. This forced oscillation modulates the tip-sample interactions around the resonant frequency. Hence, the sensor deflection signal becomes a superposition of the forced oscillation response and the tip-sample disturbance.

In amplitude-modulation AFM (AM-AFM) [4], the amplitude envelope of the sensor signal carries the tip-sample interactions. In frequency modulation FM-AFM, the instantaneous frequency of the sensor signal carries the tip-sample interactions. In order to keep the SNR constant over the scan area, a controller is used to close the loop between the deflection signal and the drive signal. In AM-AFM, the controller adjusts the sinusoidal drive signal amplitude. In FM-AFM, it is the phase that is adjusted.

Perhaps the most exciting new development in the last few years is that of video-rate or high-speed AFM (HS-AFM) imaging systems. High speed AFM imaging capable of capturing sub-100ms nanoscale events have been reported. These AFMs can observe biomolecular events as they unfold in real-time [5]. This has led to an explosion of publications and new fundamental understandings of cell morphology and cell mechanics [6][7].

Some AFM system limitations and performance bottlenecks include calibration and alignment of the various mechanical and optical components of the system. Long load/unload times of the sample. Tip

wear over time with non-linear effects on image resolution requiring periodic replacement of the tip. Limited spatial throughput (small scan range) and limited temporal throughput (slow image turnaround). There has been a considerable effort over the last three decades to address these shortcomings. Rigorous engineering optimizations by various teams have simplified the operation of the AFM. System calibration as well as optical alignments have been simplified through automation and/or precise mechanical assembly. Tip wear issues have been addressed on two fronts. First, advances in nanofabrication and nanopatterning have greatly improved the sharpness and mechanical characteristics of the tip, making it inherently less susceptible to tip wear. Second, dynamic imaging modes such as tapping-mode AFM (TM-AFM) [8] and non-contact AFM (NC-AFM) [9] minimize the contact between the tip and the sample. More recently, fully integrated AFMs-on-a-chip have been announced [10][11], representing a paradigm shift in the system realization of AFMs. This level of component integration and miniaturization has increased imaging robustness and reduced costs.

2.1.1 Probe Model

The cantilever system is a distributed infinite-dimensional continuous-time system. It may have an infinite number of eigenmodes, each with infinite number of harmonics distributed along the length of the cantilever. The simplest view of such a system is that of a simple harmonic oscillator (SHO) with no damping effects. The tip is assumed to be located at the very end of the cantilever ($x = L$), lumped to a single point centered at half the width. These assumptions lead to a single ordinary differential equation that describes the vibration mode of the tip as a function of time only:

$$m\dot{y}(t) + k_s y(t) = 0$$

Where k and m are both positive constants. k_s is the lumped cantilever (spring) stiffness and m is an equivalent lumped mass of the cantilever. It is typical to normalize the differential equation to $m = 1$, which simplifies the equation without any loss of generality.

The partial differential equation governing the transversal vibrations of the cantilever at all positions along the length of the cantilever is governed by a fourth-order partial differential equation that is a function of time and position:

$$C\frac{\partial^4 y}{\partial x^4} + \frac{\partial^2 y}{\partial t^2} = 0$$

Where C is a geometry and material dependent constant, for a rectangular beam, it equates to $\frac{Eh^2}{12\rho}$; E is the modulus of elasticity (or Young's Modulus), h and w are the thickness and width of the beam, and p the density of the material. The spring constant is given by:

$$k_s = \frac{Ewh^3}{4L^3}$$

2.1.2 Thermomechanical Noise

White noise present in the cantilever arising from random mechanical fluctuations was first investigated for its role in FM-AM [12]. The next major insight was provided through the techniques of eigen-decomposition [13]. Using the equipartition theorem, the rectangular cantilever's longitudinal flexural modes are defined as orthogonal functions of continuous time and position along the length of the cantilever. This provided a unique geometric insight into the forced response of the tip place at the end of the cantilever.

The equipartition theorem states that if a system is in thermal equilibrium every independent quadratic term in its total energy has a mean value equal to $\frac{1}{2}k_BT$, where k_B is the Boltzmann constant. In thermal equilibrium each vibration mode has a mean thermal energy of $\frac{1}{2}k_BT$. The mean square deviation of the deflection is the sum of the mean square deviations of each vibration mode:

$$< y^2 >_p = \sum_{p=1}^{\infty} < y_p^2 >$$

Where p is the indicator of each mode. It is shown that the rectangular cantilever's mean square deviation per mode is:

$$< y_p^2 > = \frac{12k_BT}{k_s\alpha_p^4}$$

With temperature T in Kelvin. Finally,

$$< y^2 >_p = \frac{12k_BT}{k_s}\sum_{p=1}^{\infty}\frac{1}{\alpha_p^4}$$

Where the wavelength of a vibration mode is L/α_p and is related to the vibration mode's angular frequency as:

$$\left(\frac{\alpha_p}{L}\right)^4 = \frac{12\rho\omega_p}{Eh^2}$$

Finally, it can be shown that $\sum_{p=1}^{\infty}\frac{1}{\alpha_p^4}$ converges to $\frac{1}{12}$ and the total mean amplitude of thermal noise due to all modes is:

$$\sqrt{< y^2 >_p} = \sqrt{\frac{k_BT}{k_s}}$$

Note $\alpha_p$is part of the analytic solution to the eigen-decomposition of a fixed-free and fixed-supported (by the sample) rectangular cantilever. For a supported cantilever the external force at the tip is not 0. Table 1

lists the solutions to α_p. The relationships revealed in Table 1 may be use for modal analysis, system modeling, and system identification, especially in multifrequency AFM operations where multiple modes carry sample information.

Mode p	α_p	
	Free	Supported
1	1.88	3.93
2	4.69	7.07
3	7.85	10.21
4	11	13.35
$p \geq 5$	$\left(p-\frac{1}{2}\right)\pi$	$\left(p+\frac{1}{4}\right)\pi$

Table 1: Mode constants for free and supported cantilever.

Some insightful take aways are:

- The amplitude of the thermal noise falls by a factor of 100 when going from the first to the sixth vibration mode.
- The first 10 vibration modes already contribute 99.93% to the total mean deflection.
- Assuming the spring constant $k_s$is given in $\frac{N}{m}$, the amplitude of thermal noise is $0.064\ nm$ per $\sqrt{k_s}$or $\sqrt{< y^2 >_p}= \frac{0.064nm}{\sqrt{k_s}}$. This sets the ultimate noise floor for the measurement of the deflections. In other words, if there was a sensor with perfect knowledge of the deflections, its resolution would be this. AFM cantilever stiffness values typically range between $0.25\frac{N}{m}$ and $10\frac{N}{m}$.

2.2 Tip-Sample Interactions

The characteristic of interactions between two small bodies may be described for a "range" of interaction distance. For interatomic range of interactions on the order of $0.2nm$, chemical/covalent bonds dominate and are the cause of highly non-linear repulsive forces in that range. The change from attractive to repulsive regime occurs round 0.3nm for two atoms. From 0.3 to a few nm, a class of attractive forces are called short-range Van der Waals (VDW) forces. The interaction from 0.3nm to 1nm is highly nonlinear with up to 7^{th}order interactions. As the separation increases from $1nm$ to $5nm$, the order of the interaction force decreases until it reaches a slow varying first order drift. For separations above a few nm, these interactions are typically referred to as the long-range vdw forces. The long range vdw forces persist up to and above $10nm$ before other interaction strength weakens and other potential fields (electrostatic) start to dominate.

The interaction forces discussed so far may be modeled as the sum of two inverse power laws. The most general expression for the potential energy of interaction between two atoms is given by the Lennard-Jones Potential:

$$U_{LJ}(d) = -\frac{A}{d^6} + \frac{B}{d^{12}}$$

Proposed by Lennard-Jones in 1924 after quantum theory had shown that the attractive exponent is 6. Negative sign indicates attractive force, positive sign indicates repulsive. For interactions between two atoms, $A = 10^{-77} Jm^6$ and $B = 10^{-134} Nm^{12}$.

The inverse power law relationship may be generalized by families of exponentials. The Morse model is a generalization of the inverse power laws and has been used for modeling interatomic interactions.

$$U_{Morse}(d) = -Ae^{\frac{d}{A_d}} + Be^{\frac{d}{B_d}}$$

Note that the tip-sample interaction force as a function of separation distance may be defined as the gradient of the interaction field.

Van der Waals simplified the model of interactions by assuming atoms, molecules, and small particles in general (10s of nm) may be considered closely-packed hard-spheres with a constant packing radius and atom-atom potential. These constants are called the Van der Waals packing radius and the van der Waals constant C_{VDW}, respectively.

2.2.1 Models of Contact

The elastic interaction of two colliding bodies was first formalized by Hertz [14]. The model is some time referred to as Hertzian contact[15] [16]. A major simplification to the Hertzian model was proposed by Hamaker [17]. Hamaker simplified the general model of elastic contact by considering the geometries of the interacting bodies as simple spheres or flat surfaces. The Hamaker constant may be viewed as the surface potential of a flat surface over a given area, it is the result of integrating force contributions of packed vdw atoms in a volume of a given density. It is given by:

$$H = \pi^2 C_{VDW} \rho_1 \rho_2$$

ρ_1 and ρ_2 are the number of atoms per unit volume. The various limiting equations of for macroscopic bodies are strictly correct only for values of d_{min} very much smaller than R ($d_{min} < R/100$).

Table 1 captures relevant interaction models based on the Hamaker assumptions.

	Limiting Condition	**Potential Energy U_{ts}**	**Force F_{ts}**
Two atoms / small molecules	$R1, R2 \approx a_0 \approx 0.1nm$	$-\frac{C_{VDW}}{{d_{min}}^6}$	$-\frac{C_{VDW}}{{d_{min}}^7}$
Macro spheres	$R1, R2 \gg d_{min}$	$-\frac{H}{6d_{min}}\left(\frac{R_1 R_2}{R_1 + R_2}\right)$	$-\frac{H}{6d_{min}^2}\left(\frac{R_1 R_2}{R_1 + R_2}\right)$
Macro sphere, flat surface	$R1 = R \gg d_{min}$	$-\frac{HR}{6d_{min}}$	$-\frac{HR}{6d_{min}^2}$

Table 2: Energy fields and forces based on the Hamaker constant and simple geometries.

According to [18], a simple yet effective model of AFM tip-sample interactions is that of a spherical particle of radius R_1near a flat surface of $R_2 \to \infty$.

$$F = -\frac{d(U_{ts})}{d(d_{min})} = -\frac{2HR^3}{3(2R + d_{min})^2 d_{ts}^2}$$

For $d_{min} \ll R$

$$F_{ts} = -\frac{d(U_{ts})}{d(d_{min})} = -\frac{2HR}{6d_{ts}^2}$$

For $d_{min} \gg R$

$$F_{ts} = -\frac{d(U_{ts})}{d(d_{min})} = -\frac{2R^3}{3(d_{min} + R)^4}$$

For molecules that interact via van der Waals forces, once their diameters are greater than about 0.5 nm, they must already be considered as particles or else the strength of their interactions will be underestimated.

2.3 State Estimations

A powerful and realistic model for state estimation is that of a stochastic process consisting of the desired deterministic signal in additive white noise. Estimation can take two forms: minimax, Bayesian. If the joint probability distribution of the signal in noise is known, it may be filtered out to achieve better signal estimation even in good SNR cases. Bayesian estimators guarantee certain average performances given a priori knowledge of the signal and the noise. Relying on a priori knowledge makes Bayesian estimators unreliable and biased in the face of large error vectors (assumed knowledge of the system was not close to the real signal sensed). The minimax approach on the other hands starts from the representation space of all possible outcomes and converges on the estimator that minimizes the maximum possible error over the representation space. In system design, minimax is ideal for guaranteeing absolute performance metrics such as minimum achievable height resolution and minimum response time (absolute time from stimulus to the detection of the stimulus). In other words, minimax operators are well suited for designing guaranteed sensor performance. Bayesian estimators on the other hand are more resilient to unmodeled conditions and are desirable for estimating sample information. Note that a statistical estimator requires some a prior information regarding the system model. This may or may not always be feasible.

In AM-AFM and FM-AFM, the term “demodulation” is the widely used in AFM literature [19] to refer to the estimation of amplitude, phase, or frequency variations. The prevailing solution is the so called $I - Q$ method described in the next section.

2.3.1 In-Phase-Quadrature-Phase

Start by specifying the amplitude and phase time variations of a real harmonic signal as functions of time. If we know a(t) and its phase, we can recover f(t) as

$$f(t) = a(t)\cos\varphi(t)$$

For small variations in instantaneous frequency, $a(t)$ is slow varying, and the phase shift $\varphi(t)$ can be estimated as small variations in baseband.

2.3.2 Kalman Estimator

The general implementation of a Kalman filter is an optimal recursive Bayesian filtering technique. The prior is a measure of belief (conditional and joint probabilities of random variables) in the system parameters (modeled as a deterministic process in white Gaussian noise). It is updated based on new evidence (observations from a sensor in white Gaussian noise). The posterior is the belief in the system models after incorporating the new observation(s). The relative scaling of belief in the model vs. belief in the new evidence is called the Kalman gain, which may itself be time-varying.

Any discrete recursive Bayesian estimator has two computational steps per recursion, the first step incorporates the new measurement to update the belief about the system error covariance matrix. The second step builds on the first to project the system state estimates forward. The Kalman filter is a surprisingly effective and powerful tool that has been applied to many domains of engineering, the discrete Kalman filter is capable of the following:

- Both the model and the observations are assumed to be noisy (closer to real life).
- The observation vector and the state vector are not restricted to be joint-wide sense stationary, this is an impactful upgrade from optimal causal linear filtering.
- In the case of *linear time-invariant* systems (steady-state estimation), the Kalman gain converges fast and may be calculated offline.
- The Kalman filter is designed to start at some time $n = 0$ without stability concerns. All that is needed to keep it an unbiased estimator at step n is to ensure an unbiased estimation at $n - 1$. Ultimately, if the initial state estimations at $n = 0$ are unbiased, the estimator is unbiased for all time.

For amplitude and phase estimation, the Kalman filter has been used to estimate a complex exponential in Gaussian white noise given noisy observations of the dynamic cantilever response. [20]

2.3.3 Lock-in Amplifiers

The simplest implementation of the I-Q method filters f(t) is a two-phase filter bank centered around the carrier frequency ω that splits the signal into an in-phase component and a quadrature-phase (90deg shifted and projected onto the positive imaginary axis) component.

Multiplying the signal with the in-phase $\cos(\omega t)$ and quadrature $\sin(\omega t)$ filters, we end up with a summation of three terms for each transformation:

$$f_I(t) = 2f(t)\cos(\omega t) = a(t)\left[\cos(2\omega t)\cos\left(\varphi(t)\right) - \sin(2\omega t)\sin\left(\varphi(t)\right) + \cos(\varphi(t))\right]$$

$$f_Q(t) = 2f(t)\sin(\omega t) = a(t)\left[\sin(2\omega t)\cos\left(\varphi(t)\right) + \cos(2\omega t)\sin\left(\varphi(t)\right) + \sin\left(\varphi(t)\right)\right]$$

According to the convolution theorem, multiplication in the time domain corresponds to convolution in the frequency domain. Hence, two out of the three terms are “high-frequency” as they exist at twice the modulating frequency. The last term is a baseband term. By low-pass filtering the I-Q terms (from dc to the carrier frequency), the baseband variations are estimated as the low-passed values:

$$\tilde{f}_I(t) = a(t)\cos(\varphi(t))$$

$$\tilde{f}_Q(t) = a(t)\sin\varphi(t)$$

Finally, the amplitude and phase are recovered through the nonlinear but simple polar-to-rectangular transformation:

$$\varphi(t) = \tan^{-1}\left(\frac{f_Q(t)}{f_I(t)}\right)$$

$$a(t) = \sqrt{f_I(t)^2 + f_Q(t)^2}$$

If the carrier signal frequency is known or can be tracked, using a phase-lock-loop for example, better SNR and passband filtering may be used to pole-zero cancel the high frequency terms first.

2.3.4 Lyapunov Estimator

Given the regularity and periodicity of an AFM amplitude response, many time-varying aspects of the Kalman filter may be simplified, greatly reducing computation per update. The Lyapunov estimator is a simplification of the time-varying adaptive Kalman filter to a constant gain steady-state amplitude / phase estimation [21].

2.4 Feedback Control

A stable feedback system plays an important role in the robust operation of the AFM. First, feedback allows the system to maintain SNR as it is disturbed by unknown surface forces. Second, tighter system integration is achieved by framing the operation of the lateral and vertical actuators as well as the observations of the cantilever states as a Multi-Input-Multi-Output (MIMO) controls problem. The latter case is one of the main reasons for the success ideo-rate AFM systems. Non-raster lateral scanning [22][23] and advanced controllers that are model-based [24][25] or learning-based [26] have successfully proven to increase imaging bandwidth while maintaining image quality.

Conversely, open-loop operation is susceptible to any unmodeled system dynamics that may arise anywhere along the open-loop path, including unpredictable surface topographies. The main advantages of open-loop operation are its simplicity in concept and implementation. In open-loop, the error signal (deviation from steady-state operation) is the desired signal. Furthermore, there is no impact on the forced response of the cantilever other than that of reference (excitation) signal and the tip-sample disturbances. A controller on the other hand tries to regulate the forced operation over time, invariably affecting it.

2.5 Advanced Techniques

2.5.1 Multifrequency

Multifrequency or multimodal AFM is a relatively new mode of imaging in which multiple spectral components are identified and tracked during of the power spectrum are used for imaging [27][28].

Multifrequency operation typically refers to observation, estimation, or operation of the system beyond a single spectral component, which is typically the first eigenmode. Higher harmonics of the first mode are reported to be sensitive to tip sample interaction [29] and may be used in force spectroscopy [30].

In bimodal operation, multiple eigenmodes of the cantilever (typically the first two flexural modes) are excited and each mode is tracked separately. The most robust and promising implementation is that of AM-AFM operation of the first mode and FM-AFM operation of the second [31].

2.5.2 Q Control

The dimensionless quality factor (Q for short) plays a vital role in determining the achievable SNR and the available system bandwidth of resonance-based imaging modes (AM-AFM, FM-AFM). The quality factor may be estimated as the ratio of the centre frequency of the resonant peak f_0 (in Hz) and the width of the resonant peak at half-power. In AM-AFM, a reduction in the Q factor reduces the settling time of the cantilever amplitude according to the following equation:

$$t_{settling} \approx \frac{Q}{f_0}$$

For instance, given a Q of 100 and a resonant frequency of 10 kHz, the settling time is 0.02 seconds, hence the effective imaging bandwidth per point is 50Hz. In other words, the resonating cantilever will have to go through 200 resonant cycles before the cantilever amplitude reaches its final steady state.

Q control can be leveraged to speed up imaging bandwidth. Decreasing Q minimizes the time required for transient tip-sample interactions to decay at any scanned point on the sample. On the other hand, an increase in the quality factor increases the sensitivity of the cantilever to surface forces, allowing for soft-landing [32] of the tip, as well as a higher sensitivity to frequency shifts in FM-AFM. This is especially useful in air or liquid where the medium itself imposes damping onto the system (e.g., squeeze-film-damping in air and water). Note that if the aim is to reduce Q, the optimal response [33] is a critically damped eigenmode which is perfectly flat over the frequency range where the resonant peak used to exist.

The response dominated by the Q factor of the resonant mode is proportional to the velocity (first time-derivative of the position). By feeding back on the velocity response, the overall velocity response of the feedback system can reduce/increase the Q factor [34][35]. The same technique has been used to suppress the vibrational modes of lateral piezoelectric actuators to improve noise rejection and imaging bandwidth [36].

2.5.3 Information Theory

There has been an increased interest in recent years to analyze and estimate the transient dynamics [37] from the deflection signal of the first few cycles of interaction between the tip and a novel location on the sample. In transient imaging [38] the estimations are continuously updated and interpreted for imaging. Wavelet transforms [39] have proven to be very effective for handling the irregularities of the deflection signal which is smooth as it approaches the sample but very abruptly changes (a singularity) when closest to it (piece-wise regular).

3 Signals

The word "signal" is a widely used term engineering and may mean very different things depending on context. A signal typically refers to a mathematical representation of a system or process. For a physical process such as the oscillating AFM cantilever, the signals representation is that of a continuous-time system with one or more inputs and outputs. Disturbances such as environmental variations may be included in the set of inputs. The system (or its signals representation) is said to be deterministic if all outputs are exact functions of the inputs and the system states only. If, on the other hand, randomness or uncertainty is present in the inputs or states, the system is stochastic and behaves probabilistically. Stochastic signals cannot be known exactly but may be modeled using statistical tools to guarantee certain confidence levels.

Signals may also be classified based on their energy and power contents. Generally, any signal f of size (or period) N is a square-integrable complex function in the Hilbert space $\mathbb{C}^N$. It has finite power and a bounded L2 norm. It may be said to belong to the space of function in $\boldsymbol{\ell}^2(\mathbb{R})$:

$$\boldsymbol{\ell}^2(\mathbb{R}) = \sum_{n=-\infty}^{\infty} |f[n]|^2 < +\infty$$

The finite energy signal has a bounded L1 norm and its power averages out to 0. It may be said to belong to the space of function in $\boldsymbol{\ell}^1(\mathbb{R})$:

$$\boldsymbol{\ell}^1(\mathbb{R}) = \sum_{n=-\infty}^{\infty} |f[n]| < +\infty$$

Stochastic processes (thermal noise) or finite energy discontinuous signals (e.g., delta function) are equivalent to (or approximated by) a set of functions in the intersection of $\boldsymbol{\ell}^1(\mathbb{R})$ and $\boldsymbol{\ell}^2(\mathbb{R})$: $\boldsymbol{\ell}^1(\mathbb{R}) \cap \boldsymbol{\ell}^2(\mathbb{R})$. For example, white gaussian noise does not belong to $\boldsymbol{\ell}^1(\mathbb{R})$ because it is ill-defined in time. It does however have a finite non-zero L2 norm, its variance. Furthermore, its value in time may be measured. Hence, the white gaussian noise process is a stochastic signal that may be approximated by functions in $\boldsymbol{\ell}^1(\mathbb{R}) \cap \boldsymbol{\ell}^2(\mathbb{R})$. For the space of functions belonging to $\boldsymbol{\ell}^1(\mathbb{R}) \cap \boldsymbol{\ell}^2(\mathbb{R})$, inner products and norms are conserved by the Fourier transform up to a factor of 2π. This is the class of signals considered in this report.

3.1 AFM Signals

For the AFM, the signal of interest is that of the instantaneous deflection of the tip $y(t)$ as show in Figure 1. This corresponds to the instantaneous position of the tip. The inputs include the excitation signal that actuates the cantilever, the ambient environment (air/vacuum/liquid), cantilever-tip physical properties, and the tip-sample interactions. In the presence of noise in the system, the response also becomes noisy. This may be observed as random variations in the instantaneous amplitude of the response signal. The statistical properties of these variations depend on the statistical characteristics of the noise.

3.1.1 Noise processes

As noted in Section 2.1.2, even an isolated non-driven AFM cantilever is inherently noisy. Since the goal of the AFM is to provide true and unbiased information, the signal $y(t)$ should be treated statistically and include confidence levels in estimation. Furthermore, any other noise process affecting AFM operation should be included as separate inputs and modeled stochastically. These include environmental disturbances and systematic noise. Temperature fluctuations or mechanical vibrations of buildings are examples of environmental disturbances.

Systematic noise refers to any noise processes arising from system components. They typically refer to electronic noise produced by semiconductor devices. Shot noise arises in semiconductor circuits for any generated DC signal. The charge carries are discrete particles with average behaviour. Hence, it is modeled as white noise. Flicker or "$1/f$" noise is a low-frequency noise component with a magnitude that attenuates with frequency. It arises from the timing variations of captured and released charge carriers from "trapped" states in the semiconductor. Flicker tends to dominate spectrum near DC, typically in 10s of Hz (near DC).

In digital signal processing, systematic error may present itself due to the limited size of number representation. For a wide range of signals, it has been shown that the quantization error is nearly a white noise [40].

3.2 Systems Analysis

It is assumed that the system is stable. That is, any bounded finite duration input results in a bounded finite-duration output. The digital representation of $y(t)$ is a discrete-time function. For a periodically driven cantilever, the response (in the absence of the influence of the sample) is also a periodic signal. In the presence of noise, it is stochastic with an unknown instantaneous value.

The deflection signal $y(t)$ of an AFM cantilever driven by a single sinusoid $x(t)$ at the cantilever's modal frequency is a periodic signal in f_d. Under the simple harmonic oscillator model discussed in Section 2.1.1 the steady state response of $y(t)$ may be calculated by taking N samples per period $T_s = 1/f_d$ over M periods. The signal is then simply a delayed and/or scaled version of $x(t)$. For $M \to \infty$, this steady-state response calculation is exact. The steady-state response has a finite non-zero total average power and infinite total energy as it oscillates for all time. Nonetheless, restricting the analysis to an integer number of periods, one may formulate an equivalent finite duration sequence and treat it as a finite energy signal.

The transient response on the other hand refers to the variations in $y(t)$ caused by some non-periodic finite duration event (signal). Both the transient response and the steady-state response under the SHO are deterministic. If the cantilever is modeled as a linear system and all tip-sample interactions are modeled as polynomials of maximum order p

3.3 Sampling Theory

A continuous-time signal $y(t)$ may be uniformly sampled by an Analog-to-Digital Convertor (ADC) with a linear transfer function mapping the strength of the measured signal onto the digital unit space. For a quantization resolution of B-bits, there are 2^B digital representations of the analog signal. Digital samples may vary from 0 to $2^B - 1$. Unit norm may be achieved by a simple multiplication of the ADC samples by $\frac{1}{2^N-1}$. For N samples collected at uniform intervals of T_s seconds, the signal $y(t)$is discretized to $y_d(nT_s)$ where the d-subscript indicates the discrete dual of $y(t)$. The discrete representation is that of $y(t)$ sampled at uniform instances nT_s for all time. As T_s is a constant (uniform time samples), it may be normalized 1 so that the discrete representation of $y(t)$ may simply be written as $y[n]$.

The Fourier transform of the time domain delta function $\delta(t - ns)$ for n and s in $\mathbb{N}$ is the complex exponential $e^{-jns\omega}$. The approximation of a uniformly sampled signal by an ADC is a linear projection that results in N complex sinusoids uniformly spread around the unit circle on the complex plane. A basic result from linear systems theory is that complex sinusoids are in fact eigenfunctions of linear systems. Hence, the response of any linear filtering is a superposition of complex sinusoids weighted by the filter's impulse response:

$$h[n] = \frac{1}{N}\sum_{k=0}^{N-1} e^{j\omega_k n}$$

For a constant uniform impulse response, this is a uniform sampling of the frequency domain via unit spectral lines defined at $\omega_k$and separated by $\frac{2\pi}{N}$.

3.4 Oversampling

When the main components of the signal are bandlimited, oversampling is still useful as it spreads the aliasing of out of band noise or nonlinear effects over a wider bandwidth. Subsequent low-pass filtering can remove them altogether.

Increasing the sampling rate introduces a redundancy between the sample values of the band-limited signal. Redundancy is useful for reducing added noise. For digital systems, it is much easier to increase the sampling rate (oversample) than it is to improve quantization resolution.

3.5 Fourier Analysis

Any real function of time $x(t)$ along with its Fourier dual function of frequency $X(j\omega)$ or simply $X(\omega)$ be categorised based on whether it has a finite/infinite duration in time (or bandwidth in frequency) and whether the representation is continuous/discrete in time (or frequency). Moreover, Fourier theory dictates that if time is discrete, bandwidth is finite, and if frequency is discrete, time duration is finite. Consequently, a real signal of time has four possible Fourier transformations [41], each with a unique inverse transform.

3.5.1 Mathematical Definitions

3.5.1.1 Fourier Transform (FT)

The Fourier Transform (FT) is continuous and infinite in time, continuous and infinite in frequency.

$$FT\{x(t)\} = X(\omega) = \int_{t=-\infty}^{t=\infty} x(t)e^{-j\omega t}\,dt$$

$$\omega \in (-\infty,\ +\infty)$$

3.5.1.2 Discrete Time Fourier Transform (DTFT)

The Discrete Time Fourier Transform (DTFT) is discrete and infinite in time, continuous and finite in frequency.

$$DTFT\{x[n]\} = X(\omega) = \sum_{n=-\infty}^{n=+\infty} x[n]e^{-j\omega n}\,dt$$

$$\omega \in (-\pi,\ +\pi]$$

3.5.1.3 Discrete Fourier Transform (DFT)

The Discrete Fourier Transform (DFT) is discrete and finite in time, discrete and finite in frequency.

$$DFT\{x[n]\} = X(\omega_k) = X[k] = \sum_{n=0}^{n=N-1} x[n]e^{-j\omega_k n}$$

$$\omega_k = \frac{2\pi k}{N},\ k = 0,\ 1,\ \ldots\ N\ -1$$

3.5.1.4 Fourier Series (FS)

The Fourier Series is continuous and finite in time, discrete and infinite in frequency.

$$FS\{x(t)\} = X(\omega_k) = X[k] = \int_{t=0}^{t=P} x(t)e^{-j\omega_k t}\,dt$$

$$\omega_k = \frac{2\pi k}{N},\ k\ \in \mathbb{N}$$

P is the total duration of the finite signal.

Based on the above definitions, some relevant practical implications are outlined in the following remarks:

1. A discrete, real, and time-limited signal can only be Fourier transformed via the DFT. This implies that in digital signal processing, only the DFT is realizable.
2. The spectrum of the discrete time-limited signal obtained from the DFT has a frequency resolution of $\frac{2\pi}{N}$.

3.5.2 Convolution Theorem

Convolution in the time-domain corresponds to pointwise multiplication in the frequency domain. Conversely, pointwise multiplication in the time-domain corresponds to convolution in the frequency-domain.

$$(x * y)[n] \leftrightarrow X[k]Y[k]$$

$$x[n]\mathrm{y}[\mathrm{n}] \leftrightarrow (X * Y)[k]$$

3.5.3 Time-Shift (Frequency-Modulation) Theorem

A constant delay (shift) in time-domain is equivalent to a linear phase shift in frequency. This is useful in removing constant system delays from the frequency spectrum.

$$x(t - \tau) \leftrightarrow e^{-j\omega\tau}X(\omega)$$

3.5.4 Frequency-Shift (Time-Modulation) Theorem

Multiplication by a complex exponential in the time-domain is equivalent to a constant shift in the frequency domain. This may be used to shift the frequency content of a signal in the time domain.

$$x(t)e^{jvt} \leftrightarrow X(\omega - v)$$

3.5.5 Parseval's (Power) Theorem

In digital signal processing, this theorem is a powerful normalization constraint. Enforcing $< x, y > = < X, Y > = 1$ allows for complete and consistent transformations between time and frequency domain representations as needed.

$$< x, y > = \sum_{n=-\infty}^{+\infty} x[n]\overline{y[n]} = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} X[k]\overline{Y[k]} = < X, Y >$$

3.5.6 Rayleigh's (Energy) theorem

The energy theorem relates a finite energy signal to its L2 norm. For any finite energy signal, the L2 norm is the mean-square value in the time-domain and the power spectrum in the frequency domain.

$$< x, x > = \|x\|_2^2 = \sum_{n=-\infty}^{+\infty} |x[n]|^2 = \frac{1}{2\pi}\sum_{k=-\infty}^{+\infty} |X[k]|^2 = \frac{1}{2\pi}\|X\|_2^2 = < X, X >$$

3.6 Digital Signal Processing

The uniform samples produced by the continuous sampling of the ADC need to be grouped into finite chunks of samples so that they may be processed in real-time. Processing a finite number of samples time-limits the infinitely long signal.

3.6.1 Time-Limited Signal

The rectangular window is the ideal time-limited view of a finite collection of discrete samples. For a window of size M, it contains M contiguous unit amplitude components and zeros everywhere else. Consequently, it is discontinuous. An example of a rectangular window is shown in Figure 2.

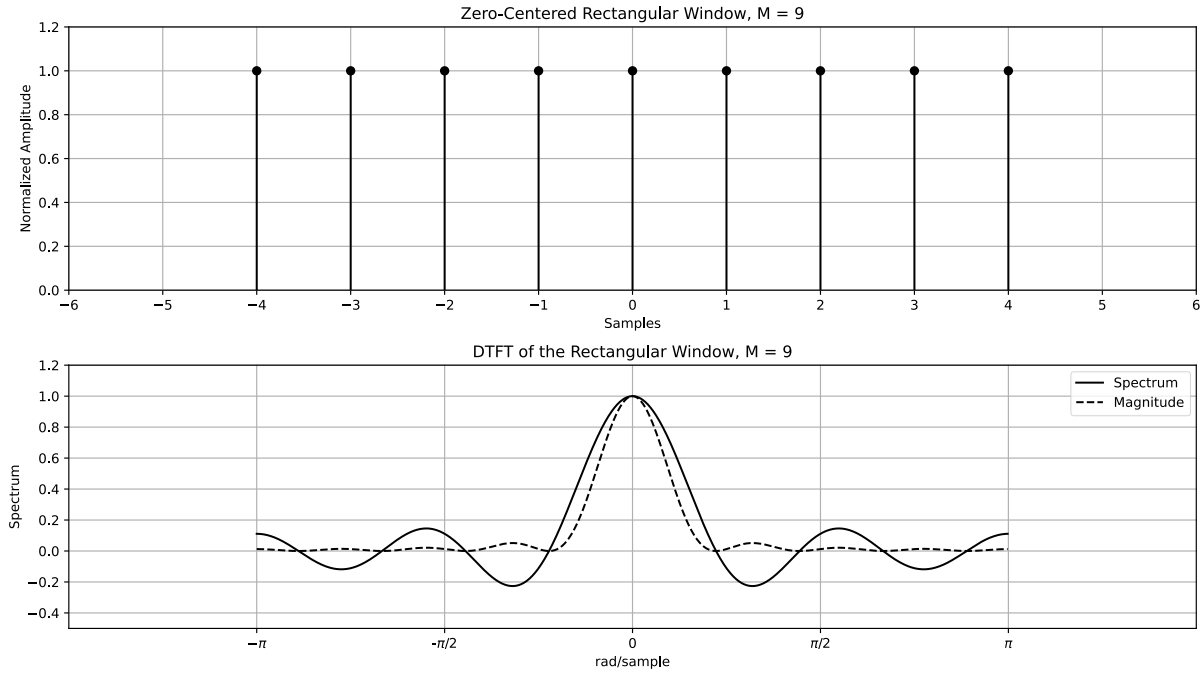


Figure 2: Rectangular window representations in time and frequency domains

The rectangular window may be defined by this piece-wise definition:

$$w[n] = \begin{cases} 1, & |n| \leq \dfrac{M-1}{2} \\ 0, & |n| > \dfrac{M-1}{2} \end{cases}$$

The zero-phase definition has non-zero components for negative time, hence it is not-causal. The window can be made causal by a simple shift to 0 such that $w[n] = 0$ for $n < 0$. The zero-phase form of the discrete time-window has an exact continuous and finite frequency spectrum given by the $DTFT$ transform:

$$W(\omega) = DTFT_{\omega}\{w[n]\} = \sum_{n=-\infty}^{\infty} w[n]e^{-j\omega n}, \;\; \omega \in [-\pi, \pi)$$

$$W(\omega) = \sum_{n=-\frac{M-1}{2}}^{\frac{M-1}{2}} e^{-j\omega n} = \frac{e^{-j\omega\frac{M-1}{2}} - e^{-j\omega\frac{M+1}{2}}}{1 - e^{-j\omega}}$$

Using the closed form of the geometric series $\sum_{n=A}^{B} r^n = \frac{r^A - r^{B+1}}{1-r}$

$$W(j\omega) = \frac{e^{j\omega\frac{M}{2}} - e^{-j\omega\frac{M}{2}}}{e^{j\omega\frac{1}{2}} - e^{-j\omega\frac{1}{2}}} = \frac{\sin\left(\frac{M\omega}{2}\right)}{\sin\left(\frac{\omega}{2}\right)} = M\text{asinc}_M(\omega)$$

Where,

$$\text{asinc}_M(\omega) = \frac{\sin\left(\frac{M\omega}{2}\right)}{M\sin\left(\frac{\omega}{2}\right)}$$

This is referred to as the "aliased sinc", "Dirichlet", or "periodic sinc" function.

Figure 3 illustrates the magnitude and phase response of the rectangular window.

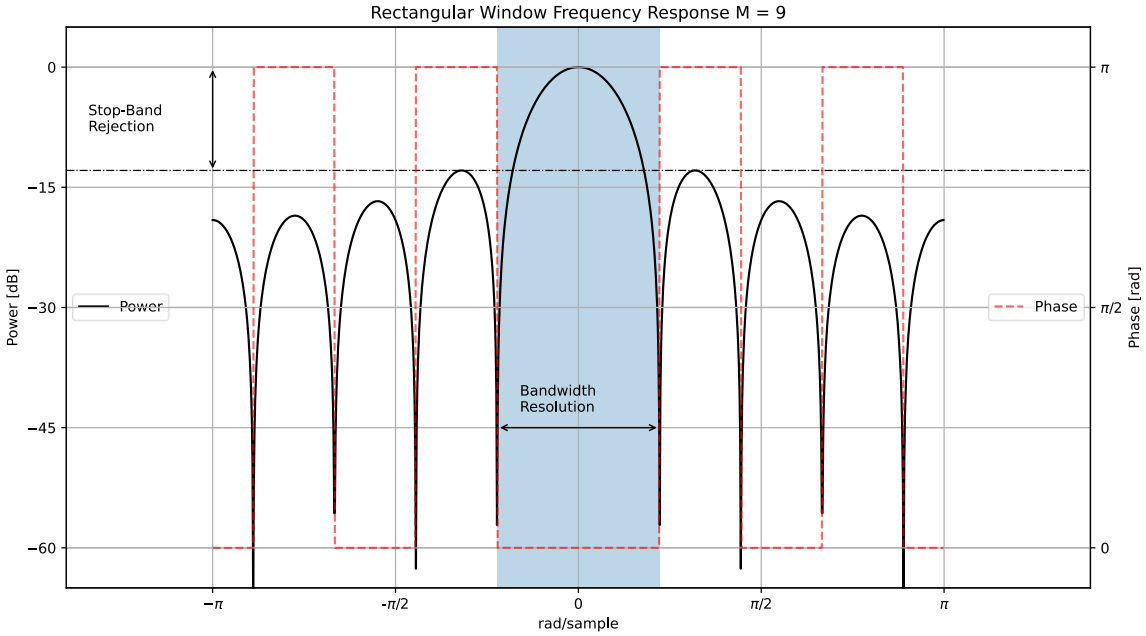


Figure 3: Rectangular window magnitude (dB) and phase response

The window bandwidth (or frequency resolution) is defined as the width of the main lobe between the zeros. This is the minimum bandwidth needed to avoid time domain aliasing.

$$\Omega_\omega = K\frac{2\pi}{M}$$

$K$is a positive integer and depends on the specific window w. This is the dual of the Shannon sampling theorem which similarly defines the minimum window length needed to avoid frequency domain aliasing. This may be used to calculate the size of the desired window required to resolve a sinusoid with frequency f_0. For sampling rate of f_s, the sinusoid period in number of samples is:

$$P = \frac{f_s}{f_0}$$

The minimum window width may be determined as follows:

$$M \geq KP$$

The window stop-band attenuation is defined is the difference in magnitude between the main lobe and the first side lobe. It indicates the window's spectral performance in rejecting noise outside of the main band. Finally, the stop-band roll-off is defined as the decay in magnitude with increasing frequency. The abruptness of the transition of the time-window from its peak to its lowest value (usually the last sample) determines its stop-band rejection (roll-off) performance.

Table 3 provides the frequency sampling capabilities, stop-band rejection, and roll-off rate of some typical analysis windows.

	Passband Resolution (1/K)	**Stop-band Attenuation**	**Stop-band Roll-off**
Rectangular	1/2	-13dB	-6dB/octave
Hanning	1/4	-31.5dB	-18dB/octave
Hamming	1/4	-41dB	-6dB/octave
Blackman	1/6	-71dB	-6dB/octave

Table 3: Spectral performance of common windows

3.6.2 Band-Limited Signal

Figure 4 illustrates the ideal representation of a low-pass filter in the frequency domain.

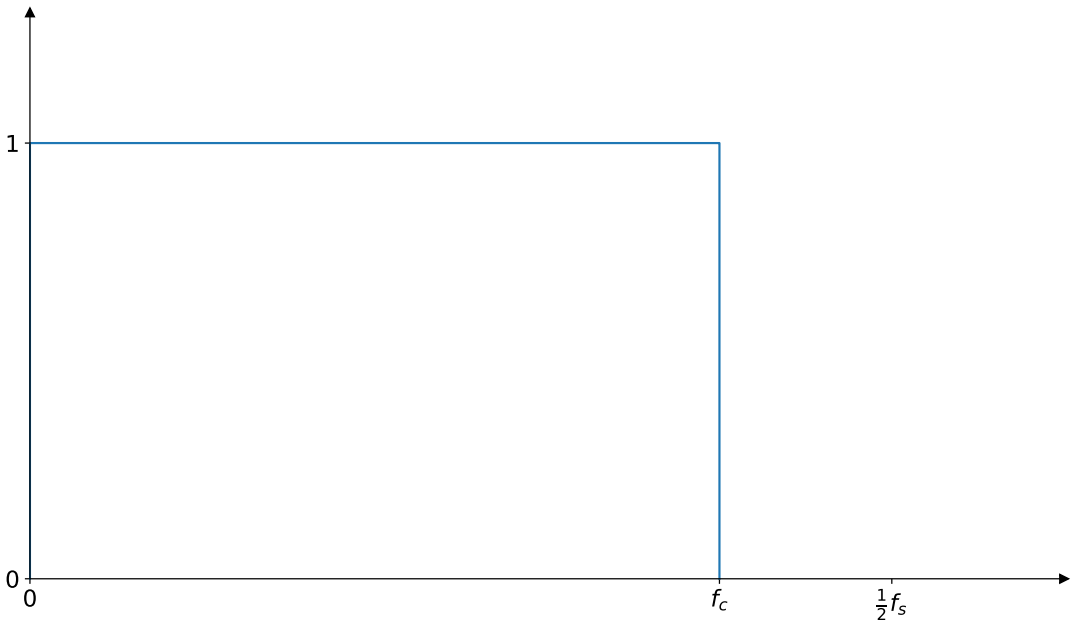


Figure 4: Ideal low-pass filter response

The response may be defined directly in the frequency domain given its cut-off frequency ω_c. as the following piece-wise equation:

$$H(\omega) = \begin{cases} 1, & |\omega| \leq \omega_c \\ 0, & |\omega| > \omega_c \end{cases}$$

The filter's impulse response in time domain may be calculated from the inverse $DTFT^{-1}$transform.

$$h[n] = DTFT^{-1}\{H(\omega)\} = \frac{1}{2\pi}\int_{-\omega_c}^{\omega_c} e^{j\omega n} d\omega$$

$$= \frac{1}{2\pi jn}\left[e^{j\omega_c n} - e^{-j\omega_c n}\right] = \frac{\text{sinc}(\omega_c n)}{\pi n} = 2f_c \text{sinc}(2f_c n)$$

The sinc function or the Cardinal sine function is infinitely long. Therefore, it is not realizable, and it cannot be shifted to be made causal. A crrude approximation may be made by time-limiting the sinc function in order to extract the a causal version of the ideal bandpass filter.

Given the difficulties with the time-domain representation of the ideal bandpass, the best digital filter design that may be achieved is an optimal approximation to the desired frequency response by minimizing a cost function or error criterion.

3.6.3 Fast Fourier Transform (FFT)

The DFT as defined in section 3.5.1.3 is referred to as the "direct method" of computation. Executed on a standard single-thread CPU, the computation is on the order of $O(N^2)$. The Fast Fourier Transform [42] is an alternative algorithm that reduces the computation to the order of $O(NlogN)$, a massive gain for large inputs. It is estimated that for inputs greater than 100 samples the FFT outperforms the direct method of computation. Due to its performance, the FFT has had a huge impact on the proliferation of computing and real-time signal processing. Audio and video signals are routinely transformed using the FFT algorithm. GPUs and custom hardware implementations can beat the FFT given enough adders and multipliers. One of the constraints of the FFT algorithm is that the size of its input must be a power of 2 for maximum computational efficiency. For signals of arbitrary size, this shortcoming may be overcome as discussed next.

3.6.4 FFT Convolution

Based on the convolution theorem discusses in section 3.5.2, an algorithm for the calculation of the convolution of two signals $x[n]$ and $y[n]$ can be described in three general steps:

1. Perform a Fourier transform on each signal to obtain $DFT\{x[n]\} = X[k]$ and $DFT\{y[n]\} = Y[k]$
2. Pointwise multiply their spectra (equivalent to convolution in the time domain), $C[k] = X[k]Y[k]$
3. Inverse Fourier transform to obtain the desired result $DFT^{-1}\{C[k]\} = (x * y)[n]$

In practice however, care must be taken to avoid time aliasing in step 3. For a signal with N nonzero values and a signal with M nonzero values, the convolved signal contains at most $M + N - 1$ nonzero values. The algorithm must guarantee faithful implementation of the convolution operation. The FFT algorithm treats each collection of inputs as periodic in its size, so it cannot handle signals of varying size.

3.6.5 Zero-padding

Zero padding refers to extending the size of the FFT's input signal by 0s. There are a few advantages to doing so:

- Zero-padding in the time-domain is ideal interpolation in the frequency domain (aliasing must be handled separately). Peak estimation is easier when the spectrum is densely sampled.
- The signal's size may be rounded up to the next number that is a power of 2 to take advantage of the FFT speedups available.
- As explained in the previous section, acyclic convolution is achieved for two signals of arbitrary size using zero-padding. Zero adding allows for multiplicative modification in the frequency domain without time aliasing in the inverse FFT [43].

Only signals truly periodic in M should not be zero-padded, they should also use the rectangular window. All digital filtering of a uniformly sampled continuous time signal of has a constant delay (linear phase term) of $N/2$, where N is the size of the filter traps.

3.7 Time-Frequency Distributions

Functions of time and frequency belong to a class of time-frequency distributions. Some examples include the Short-Time-Fourier Transform (STFT) and the wavelet transform.

3.7.1 Short-Time (Windowed) Fourier Transform

Let $x[n]$ be a discrete-time signal defined over all time. Its spectrum $X(\omega)$ is given exactly by the DTFT:

$$X(\omega) = \sum_{n=-\infty}^{n=+\infty} x[n]e^{-j\omega n}$$

$$X_m(\omega) = \sum_{n=-\infty}^{n=+\infty} x[n]w[n - mR]e^{-j\omega n}$$

$$\sum_{m=-\infty}^{m=+\infty} X_m(\omega) = \sum_{n=-\infty}^{n=+\infty} x[n]e^{-j\omega n} \sum_{m=-\infty}^{m=+\infty} w[n - mR]$$

$$\sum_{m=-\infty}^{m=+\infty} X_m(\omega) = X(\omega)$$

If and only if

$$\sum_{m=-\infty}^{m=+\infty} w[n - mR] = 1$$

Let $w_M[n]$ be a of M non-zero samples. Define $X_m(\omega)$ as the spectrum of the windowed signal $x(t)$

centered at some time sample mR where R is a constant integer called the "hop" size:

$$X_m(\omega) = \sum_{n=-\infty}^{n=+\infty} x[n]w[n - mR]e^{-j\omega n}$$

If $R < M$, consecutive windows overlap.

Finally, summing all windowed spectra over the length of the signal, we obtain a collection of *time-ordered* spectra:

$$\sum_{m=-\infty}^{m=+\infty} X_m(\omega) = \sum_{m=-\infty}^{m=+\infty} \sum_{n=-\infty}^{n=+\infty} x[n]w[n - mR]e^{-j\omega n}$$

This may be re-written, without any loss of generality:

$$\sum_{m=-\infty}^{m=+\infty} X_m(\omega) = \sum_{n=-\infty}^{n=+\infty} x[n]e^{-j\omega n} \sum_{m=-\infty}^{m=+\infty} w[n-mR]$$

Hence, $\sum_{m=-\infty}^{m=+\infty} X_m(\omega) = X(\omega)$ if and only if the superposition sum of all shifted window $\sum_{m=-\infty}^{m=+\infty} w(n-mR) = 1$ or some known constant which can be divided out.

This is called the overlap-add (OLA) processor based on the STFT. It may be used to filter an infinitely long signal or process available data in real-time systems.

3.7.2 Spectrograms

The spectrogram is the power spectrum of the windowed Fourier transform. It contains no phase information but may be used to visualize the time-frequency response.

3.7.3 Time-Frequency Resolution

For a statistical signal, the signal bandwidth and resolution may be represented as second order moments of the signal distribution.

Resolutions can be defined as the normalized root-mean-squares (RMS) of the squared products of time-amplitude and frequency-spectrum signals

$$\Delta t \triangleq \frac{1}{\|x\|_2}\sqrt{\int_{-\infty}^{\infty} t^2|x(t)|^2 dt}$$

$$\Delta\omega \triangleq \frac{1}{\|X\|_2}\sqrt{\int_{-\infty}^{\infty} \omega^2|X(\omega)|^2 \frac{d\omega}{2\pi}}$$

Where,

$$\|x\|_2^2 = \int_{-\infty}^{\infty} |x(t)|^2 dt$$

$$\|X\|_2^2 = \int_{-\infty}^{\infty} |X(\omega)|^2 \frac{d\omega}{2\pi}$$

For the STFT, the time resolution is determined by the window length, and the frequency resolution by the width of the main-lobe of the window-transform.

4 Design

In this chapter, the design and open-loop operation of an AFM system are explored. The analysis begins with a systems view of all connected parts. Next, a model for a fully integrated AFM is presented. An analog frontend with digital-to-analog (DAC) and analog-to-digital (ADC) convertors is used to continuously drive and sample the AFM signal. The digital representation of the AFM signals is prepared and processed concurrently by a central processing unit (CPU). This configuration is used to design a digital signal processing pipeline suited for system identification and operation. Design challenges and compromises are explored along the way.

4.1 Approach

One of the essential design ideas is to leverage the processing unit's high frequency, low-jitter digital clock (100s of MHz) as the "master" clock from which all other digital clocks are generated. With careful design, this enables the CPU, the analog frontend, and the imaging events to be phase-locked with the master clock. In other words, all generation and sampling processes are "synchronous" to the rising/falling edge of the master clock. On the other hand, transient events arising from the environment or tip-sample interactions are asynchronous. However, these asynchronous events are slow compared to the speed of the master clock, hence, the timing of asynchronous events may be approximated by the closest synchronous clock edge. In general, all events are timestamped relative to one of the master clock edge transitions. Microcontroller (MCU) based, or Field-Programmable-Gate-Array (FPGA) based digital systems are ideal platform for realization of the system.

The other key idea is to phase-lock (synchronize) the operations of the cantilever DAC drive and the ADC sampling of the noisy input and the cantilever deflection. To establish and keep a consistent timing relationship between the DAC and the ADC, their sampling rates is set as an integer multiple of the other. Furthermore, they are configured to sample continuously throughout the operation. All that is needed to start the operation is a single command that triggers the DAC and ADC sampling processes concurrently. Once the processes are started, collections of sampled input-output data are processed in real-time for state estimation, and ultimately imaging.

For system identification and imaging operations, a two-step calibration procedure is proposed. In step one, a system identification experiment reveals the detailed specifications of the cantilever with no tip-sample interaction. This informs the complexity of the reduced order model suitable for imaging. The second calibration occurs near the sample before imaging has starts. With the known excitation signal, the selected reduced order model from step 1 is tuned to be used for state estimation during imaging.

For imaging, the lateral actuators scan the sample in a raster-like pattern. The first major assumption here is that the start time and the rate of lateral scanning is known. This knowledge is enough to be able to associate cantilever oscillations to a point on the sample. The second major assumption is that the process of lateral scanning is completely decoupled from the operation of the cantilever and has no influence on it. Hence, the cantilever system may be modeled as a single-input-single-output (SISO) system.

4.1.1 System Block Diagrams

Figure 5 gives an overview of the proposed system and the signals interfacing the analog front end with the AFM.

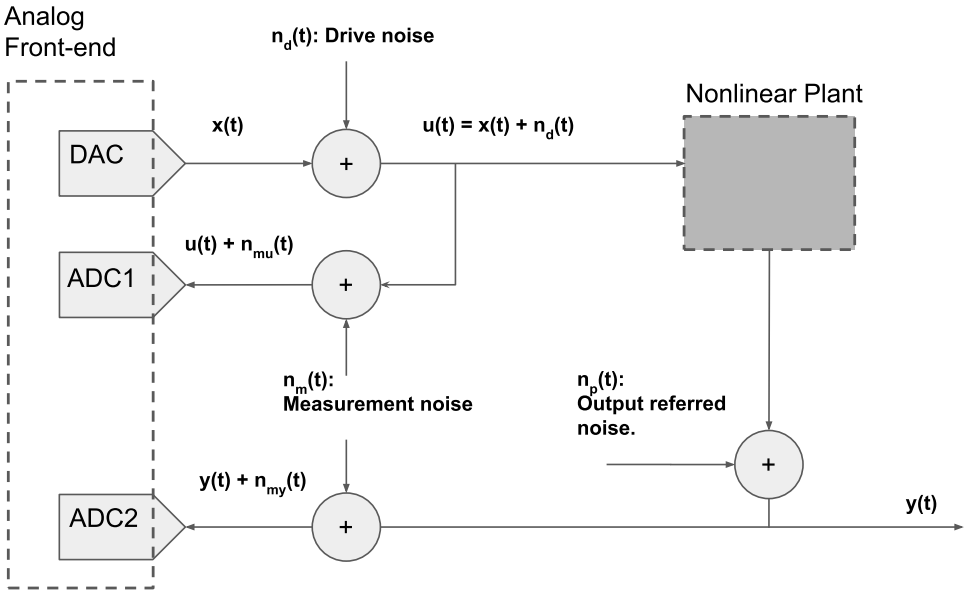


Figure 5: System and signals block diagram

The analog frontend consists of one DAC and two ADC modules. The DAC takes its samples from a look-up-table of values corresponding to the desired excitation signal. The generated signal is the bandlimited analog signal $x(t)$. By the time the signal arrives at the plant's input, it has been disturbed by a an additive noise process, referred to as $n_d(t)$. This noisy version is labelled $u(t)$ and considered the true analog signal as received by the plant. Even though the desired excitation signal is perfectly known, it is very insightful to observe its noisy version in order to observe and reject the noise components. Hence, ADC1 is allocated for sampling $u(t)$. A second ADC module, ADC2 is operated in parallel, and allocated to sample the noisy AFM deflection signal. Sampled analog signals may further be disturbed by measurement noise inherent in analog frontend. This is represented as the additive white noise process $m(t)$. In essence, this system realizes fast, phase-synchronized, and tightly coupled generation and sampling of the input-output signals.

Figure 6 illustrates one possible configuration of the processing unit.

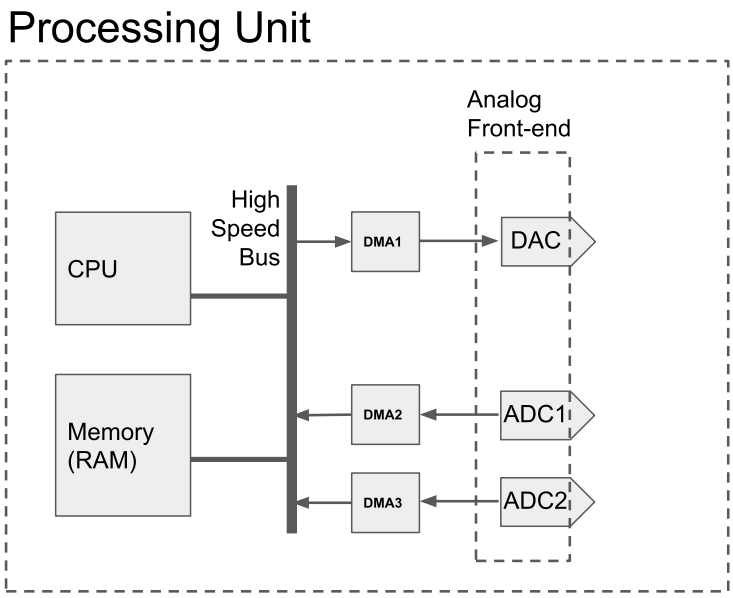


Figure 6: Processing unit system block diagram with synchronous clocking

Direct-memory-access (DMAs) units are used to copy DAC/ADC data from/to main memory without any CPU intervention. This configuration decouples the CPU process from the analog operations. With a CPU clock in 100s of MHz and analog sampling rates in 10s of KHz, the CPU has many free cycles to allocate to the digital signal processing of the sampled input-output data. An FPGA system may also realize such parallel digital/analog signals.

The dominant components of the electrothermally actuated and piezoresistively sensed AFM cantilever are presented in Figure 7.

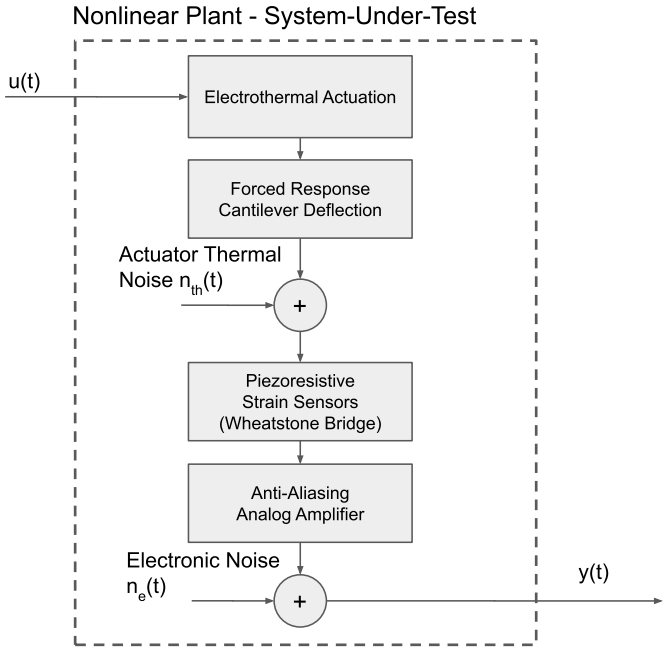


Figure 7: Simplified model of the system under test

The thermal noise $n_{th}(t)$ refers to the subject of the thermomechanical noise processes described in 2.1.2. Electronic noise $n_e(t)$ includes flicker noise, shot noise, and system delays due to the circuit component. Shot noise and actuator thermal noise processes are present over all system frequency bands and contributes to the minimum sensor resolution (system noise floor). Flicker noise distortions dominate near DC and may have significant impact if not considered. Flickr noise maybe identified and actively removed as part of the filtering process.

The excitation signal is a voltage/current variation that excites the electrothermal actuator. A bimorph cantilever is a coupling of two thin films of rectangular shape that have a slightly different coefficient of thermal expansion. The layers are bonded together through a variety of processing fabrication steps. A longitudinal heating element introduces a local variation in the bimorph temperature and forces the structure to expand / contract longitudinally. Direction of expansion and contraction is based on the direction and periodicity of heat fluctuation. Given the variation in the bimorph structure's coefficient of thermal expansion, the layers start to expand / contract at different rates. The bonding force constraints the rectangular layered cantilever to bend in the transverse direction.

4.1.2 Open-loop vs. Closed-loop

Open loop is typically used as the starting point for system identification, controller design, controller verification, and in fact provides enough information so that deep analysis of robustness and stability of the closed-loop variant can be studied. Hence, it is a natural starting point and may be used to influence the design of a controller.

To avoid problematic system deviations, some limiting operations are recommended to ensure the tip nor the sample is damaged. Tip wear/damage is generally undesirable as it degrades the tip and may damage or destroy sensitive samples. One of the downsides of open-loop operation is that if the true system response deviates too widely from the modeled process, the system is unable to operate as designed.

4.1.3 Frequency Response Function

Frequency response functions (FRF) are estimations of the system frequency response using sampled data. Give the frequency band of interest, data driven, robust and stable nonparametric FRF approximations are easily achievable. FRF measurements provide insights into the complexity of the system response. This in turn informs the minimum order for a suitable parametric model.

It is important to note that the FRF is not only a function of the nonlinear plant, but also of the input $u(t)$. The quality of the excitation signal directly impacts thee FRF estimates. The proposed excitation signals are discussed in Section 4.3. For system identification, the system is excited over a wide frequency band to so that all relevant system dynamics may be captured. Wide band refers to a frequency range spanning multiple eigenmodes or harmonics of eigenmodes. Finally, the sampled FRFs and their associated excitation signals are used to calculate the best linear approximation (BLA).

4.2 Model Parameters

The parameters are a collection of static and dynamic system information. Static system information is system specific information, such as constant open-loop system delays, normalization operators. The dynamic system information consists of a collection of desired harmonic/modal parameters. Well separated spectral peaks (eigenmodes/harmonics) are ideal for imaging and may be identified from the BLA. The dynamic system parameters are used to track the desired spectral components and generate the image signal.

4.2.1 Harmonic Oscillators

The simple harmonic oscillator is an ideal resonator with no damping. This is not a realistic model for operation as the signal is always in noise and energy dissipation (damping) processes are unavoidable. Instead, a collection of damped harmonic oscillators are typically used.

$$\sum_{p=0}^{P-1} \ddot{y}_p + \frac{1}{2Q_p}\dot{y}_p + \omega_p^2 y_p = F_{ts}(y, \dot{y};\ t) + u(t)$$

$$\omega_p = +\sqrt{\omega_p^2} > 0$$

Model for a superposition of P damped simple real harmonic oscillators is given by:

$$y(t) = 2\sum_{p=0}^{P-1} a_p e^{-d_p(t+\tau_{ts})} \cos\big(\omega_p(t+\tau) + \theta_p\big)$$

Where d_p is the decay, $a_p$the amplitude and $\omega_p$the resonant angular frequency, and θ_p the phase (center frequency) of the decaying resonant component. τ is the known open-loop delay.

The damped resonant frequency is given by $\omega_d(p) = \sqrt{d_p^2 + \omega_p^2}$.

The Q factor is related to the model by:

$$Q_p = \frac{\omega_p}{2d_p}$$

Note that this model consists of real (cosine) signals only and may track an exponentially decaying linear response perfectly. This is ideal for modeling the step response of a system to a pulse introduced τ second the past. Nonetheless, it is not capable of tracking the direction of decay. In dynamic AFM imaging, this may be interpreted as the ability to detect a step change in the resonant peak, but not uniquely identify in which frequency direction.

Model for a superposition of P damped simple complex harmonic oscillators is given by:

$$y(t) = \sum_{p=0}^{P-1} a_p \, e^{j\theta_p} e^{(j\omega_p - d_p)(t+\tau)}$$

Model for a superposition of P discrete damped simple harmonic oscillators is. Used in nuclear magnetic resonance (NMR) imaging [44], the model is capable of tracking frequency shifts of resonant components in increasing and decreasing directions of frequency. This is ideal for tracking, without ambiguity, shifts in spectral peaks applicable to

4.3 Excitation Signals

There are two classes of excitation signals considered. 1. wide-band random multisine, 2. pure-harmonic(s). In both cases, DC is assumed to be biased and always set to 0. For multisine signal generation, a collection of complex exponentials may be defined in the frequency-domain and inverse transformed to recover the causal real-time excitation signal. This is substantially more efficient than summing the same collection of harmonics in the time domain, especially as the number of multisines grow large. All signals are bandlimited and time limited.

4.3.1 Real Harmonics

A pure harmonic to drive near or at the desired modal or harmonic spectral component.

$$x[n] = \frac{1}{2}\left(1 + \cos\left(2\pi f_g n/N\right)\right)$$

DC should be set to 0 since it carries no information. Most excitation signals for dynamic operation of AFMs uses a few harmonics to excite the desired eigenmode(s) or harmonics.

4.3.2 Random Multisine

The random multisine may be defined by the following generating function.

$$x[n] = \frac{k}{N} \sum_{k=-\frac{N}{2}}^{\frac{N}{2}} \{X_k\} e^{j\{\varphi_k\}n}$$

k is a randomly generated integer between 1 and $N-1$. The phase $\{\varphi_k\}$ is parametrized by k such that $E\{\varphi_k\} = 0$

$$\{\varphi_k\} = \frac{2\pi f_g}{N} k$$

The amplitudes $\{X_k\}$ may depend on k. f_g is the DAC (generator) sampling rate. Two random phase multisines based on the choice for the amplitudes $\{X_k\}$ are discussed next.

4.3.3 Uniform Random-Phase Multisine

The uniform random phase multisine signal has unit norm (flat) power spectrum, $\{X_k\} = 1$ over the band of excitation. This simulates a bandlimited white noise process. A realization of this excitation signal with 257 samples is shown in Figure 8.

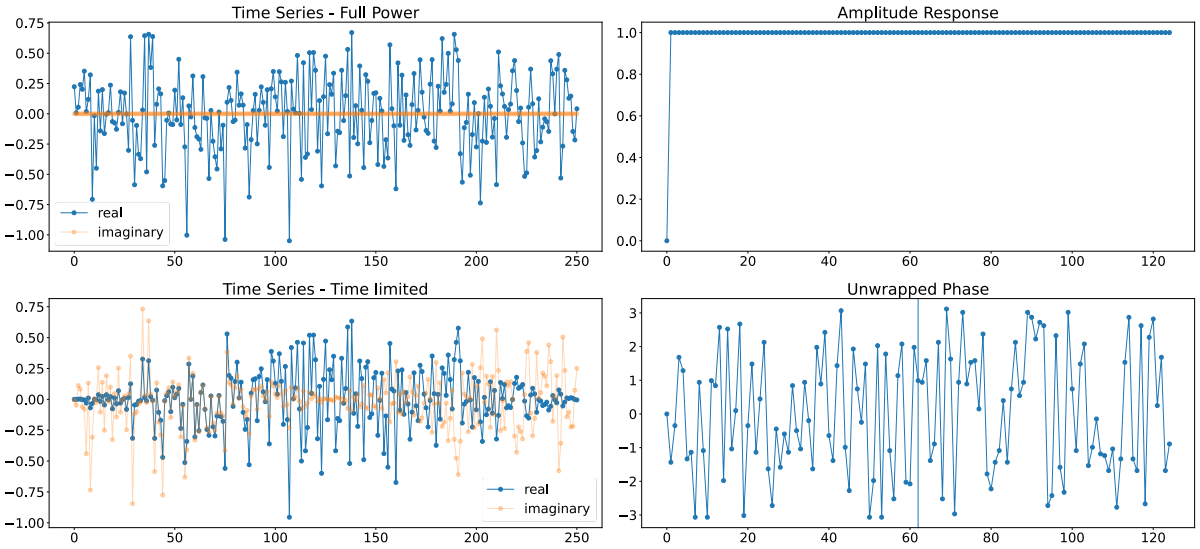


Figure 8: Uniform random phase multisine

4.3.4 Periodic Noise Multisine

A periodic noise multisine is similar to the uniform random-phase multisine but in addition to the random phase generated by k, random positive amplitudes are also generated for all spectral components $\{X_k\}$. This is also referred to as periodic noise. See Figure

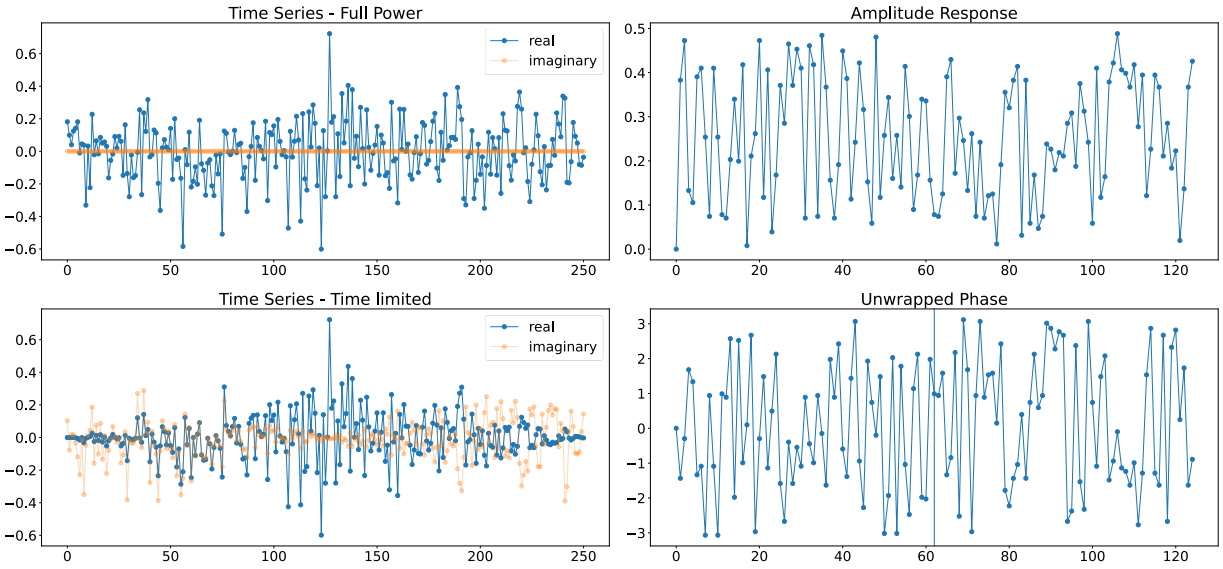


Figure 9: Random noise multisine

4.4 System Identification

For open-loop operations, system identification is crucial. Any unmodeled system dynamics or potential noise disturbances may contribute directly to the error, reducing imaging bandwidth, quality, and resolution.

4.4.1 Modeling

The realistic time-domain plant model shown in Figure 7 is mapped to the following frequency-domain components as shown in Figure 10. What follows is the analysis that has led to this frequency-domain model.

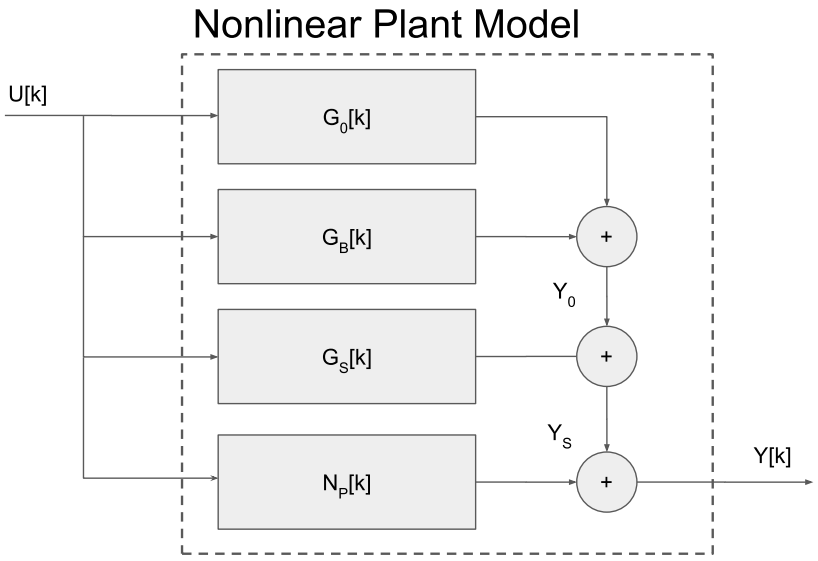


Figure 10: Frequency domain representation of the AFM

$Y[k]$ is defined to be the discrete-frequency function representing the cantilever's instantaneous deflection output $y(t)$ (sensor signal) sampled at spectral lines $\omega_k = \frac{2\pi k}{N}$. We may write this relation as:

$$Y[k] = Y(j\omega_k) = DTFT\{y(t)\}$$

It is important to note that for N ADC samples, there are only $N/2 - 1$ true samples of the system FRF. This is because the samples at DC and the Nyquist frequency carry no useful information. Hence, the samples start at $\frac{2\pi}{N}$, continue at $\frac{2\pi}{N}$ spacings, and end at $\frac{2\pi(N-1)}{N}$. Similarly, $U[k]$ is the FRF of the input signal as received by the plant.

To establish a general relationship (see [45] for the complete mathematical treatment), let $Y[k]$ and $U[k]$ be the sampled and Fourier transformed signals $y(t)$ and $x(t)$. Next, express the input-output FRFs as a mixture of a deterministic signal (signals Y_0 and U_0) and stochastic processes (signals N_Y and N_U) of potentially any order.

$$Y[k] = Y_0[k] + N_Y[k]$$

$$U[k] = U_0[k] + N_U[k]$$

The general non-linear discrete frequency response function $G[k]$ is:

$$G[k] = \frac{Y[k]}{U[k]} = \frac{Y_0[k] + N_Y[k]}{U_0[k] + N_U[k]}$$

The signals and systems in above equation may be represented in the frequency domain as:

$$N_Y[k] = DFT\{(n_e \star g_0)(t) + m_y(t) + n_e(t)\} = N_e[k]G_0[k] + N_{MY}[k] + N_P[k]$$

$$N_U[k] = DFT\{m_u(t) + n_e(t)\} = N_e[k] + N_{MU}[k]$$

Note that the drive noise process $N_d[k]$ propagates through the plant and is a distortion present on the output. This is important for open-loop operation and system identification in general: the quality of the measured FRF is directly proportional to the quality of the input signal.

4.4.2 Best Linear Approximation

Let $G_0 = Y_0/U_0$ be the first order approximation, this represent the deterministic linear-time invariant component. The FRF is re-written as

$$\frac{Y[k]}{U[k]} = G_0[k]\frac{1 + N_Y/Y_0}{1 + N_U/U_0}$$

A Taylor series expansion up to fourth order interactions around G_0 may be carried out directly in the frequency domain: (recall that a Fourier dual of the time derivative $\frac{df^n(t)}{(dt)^n}$ is $(j\omega)^n F(j\omega)$):

$$\tilde{G}[k] = G_0[k]\left(1 + \frac{N_Y}{Y_0}\right)\left(1 - \frac{N_U}{U_0} + \left(\frac{N_U}{U_0}\right)^2\right) = G_0[k]\left(1 + \left(\frac{N_U}{U_0}\right)^2 + \left(\frac{N_Y}{Y_0} - \frac{N_U}{U_0}\right) - \frac{N_Y}{Y_0}\frac{N_U}{U_0} + \frac{N_Y}{Y_0}\left(\frac{N_U}{U_0}\right)^2\right)$$

The BLA is an approximation that may be defined as:

$$G_{BLA}[k] = E\{G(j\omega_k)\}$$

The estimation $\tilde{G}[k]$ is decomposed into four contributing components:

$$\tilde{G}[k] = G_0[k] + G_B[k] + G_s[k] + N_G[k]$$

The BLA is simply the sum of the first two components:

$$G_{BLA}[k] = G_0[k] + G_B[k]$$

G_0 is the linear system response defined above. G_B is assumed to correlate with the odd nonlinear distortions and the power spectrum of the input. It models stochastic biases arising from the noise in the input. It carries no phase information and is given by

$$G_B[k] = 1 + \left(\frac{N_U}{U_0}\right)^2$$

G_s models stochastic components are uncorrelated with but NOT independent from the best linear approximation. Under the right conditions, this class of signals will behave deterministically given the appropriate system state (excitation signal with higher order phase correlations). Since they require the correct conditions and specific excitation signals, they may be wrongly characterized as high frequency transient noise that is independent from the system under test.

$$G_S = \frac{N_Y}{Y_0}\left(\frac{N_U}{U_0}\right)^2 - \frac{N_Y}{Y_0}\frac{N_U}{U_0} = \frac{N_Y}{Y_0}\frac{N_U}{U_0}\left(1 - \frac{N_U}{U_0}\right)$$

$N_G[k]$ is the model variance (uncertainty) due to drive and measurement noise

$$N_G[k] = \frac{N_Y}{Y_0} - \frac{N_U}{U_0} = \frac{N_{MY} + N_{MU} + N_p}{U_0}$$

4.4.3 Identification Process

The identification process is summarized in the following five steps:

1. Create M realization of a random multisine and save in DAC's look-up table
2. For each realization, start the generation (DAC) collection (AD1, ADC2) and
    a. Collect P periods of the input $u(t)$ (ADC1) and output $y(t)$ (ADC2)
    b. Fourier transform u and y and estimate the FRF for each period as $G_p = Y_p/U_p$
    c. Average results over the P periods per realization
3. Phase adjust realization averages based on the "known" phase of the generated excitation
4. Average over all realization to obtain final the final result.

To optimize for the detection of nonlinearities, the goal should be to increase the averaging over the set of responses that include those nonlinearities. Hence, a minimum number of periods $P \geq 2$ and as many realizations M as possible (limited by the total test time) should be selected. On the other hand, if the goal is to minimize the variance of the estimated BLA, a minimum number of realizations $M \geq 2$ and as many periods as possible (limited by the total test time) should be selected.

4.4.3.1 Calculations

Let $\tilde{G}[k][m,p]$ denote the estimated transfer between the input frame $U[k]$and output frame $Y[k]$for the p^{th}period of the m^{th} excitation realization.

$$\tilde{G}[k][m,p] = \frac{\tilde{Y}[k][m,p]}{\tilde{U}[k][m,p]}$$

With no down-sampling/up-sampling, $k = 0 \ldots N-1$

According to the energy theorem, this is the same as averaging over the time series data. Since the excited spectral lines do not vary for the same input, hence, phase information is still preserved at this step.

$$\tilde{G}[k][m] = \frac{1}{P}\sum_{p=1}^{P}\tilde{G}[k][m,p] = \frac{1}{P}\sum_{p=1}^{P}\frac{\tilde{Y}[k][m,p]}{\tilde{U}[k][m,p]}$$

Since this averaging operation is over periods of the same random phase input, phase information is still preserved here. For the final estimation, it would ideal to be able to simply "average" over the realizations.

$$\tilde{G}_{BLA}[k] = \frac{1}{M}\sum_{m=1}^{M}\tilde{G}[k][m]$$

However, because the phase distribution of each realization is random, averaging would surely remove all phase information. To avoid this, the known phase information of each realization is used to phase correct the each input-output signal first. Let $\widetilde{U}_o$ and $\tilde{Y}_o$ denote the phase adjusted variants of the input-output estimates. Multiplication of the frequency signal by $e^{-j\varphi_k}$ is equivalent to a delay in the time-domain

$$\widetilde{U}_o[k][m] = \tilde{U}[k][m]e^{-j\varphi_{m,k}}$$

$$\tilde{Y}_o[k][m] = \tilde{Y}[k][m]e^{-j\varphi_{m,k}}$$

where $\varphi_{m,k} \in [0,\pi]$ is the phase of the k^{th} spectral line of the m^{th} realization.

The input-output samples means for the set of M realizations are:

$$\widetilde{U}_M[k] = \frac{1}{M}\sum_{m=1}^{M}\widetilde{U}_o[k][m]$$

$$\tilde{Y}_M[k] = \frac{1}{M}\sum_{m=1}^{M}\tilde{Y}_o[k][m]$$

The BLA is:

$$\tilde{G}_{BLA}[k] = \frac{\tilde{Y}_{M0}[k]}{\widetilde{U}_{M0}[k]}$$

The following calculations quantify all noise components in the system.

The sample input-output covariances are:

$$\tilde{\sigma}^2_{\tilde{Y}Y,n}[k] = \frac{1}{M(M-1)}\sum_{m=1}^{M}\left(\tilde{Y}_o[k][m] - \tilde{Y}_M[k]\right)\overline{\left(\tilde{Y}_o[k][m] - \tilde{Y}_M[k]\right)}$$

$$\tilde{\sigma}^2_{\widetilde{U}U,n}[k] = \frac{1}{M(M-1)}\sum_{m=1}^{M}\left(\widetilde{U}_o[k][m] - \widetilde{U}_M[k]\right)\overline{\left(\widetilde{U}_o[k][m] - \widetilde{U}_M[k]\right)}$$

The (cross-)covariance of the input and/or outputs are:

$$\tilde{\sigma}_{YU}^2[k] = \frac{1}{M(M-1)} \sum_{m=1}^{M} \left(\tilde{Y}_o[k][m] - \tilde{Y}_M[k]\right)\overline{\left(\tilde{U}_o[k][m] - \tilde{U}_M[k]\right)}$$

$$\tilde{\sigma}_{UY}^2[k] = \frac{1}{M(M-1)} \sum_{m=1}^{M} \left(\tilde{U}_o[k][m] - \tilde{U}_M[k]\right)\overline{\left(\tilde{Y}_o[k][m] - \tilde{Y}_M[k]\right)}$$

$$\tilde{\sigma}_{UU}^2[k] = \frac{1}{M(M-1)} \sum_{m=1}^{M} \left(\tilde{U}_o[k][m] - \tilde{U}_M[k]\right)\overline{\left(\tilde{Y}_o[k][m] - \tilde{Y}_M[k]\right)}$$

$$\tilde{\sigma}_{YY}^2[k] = \frac{1}{M(M-1)} \sum_{m=1}^{M} \left(\tilde{Y}_o[k][m] - \tilde{Y}_M[k]\right)\overline{\left(\tilde{Y}_o[k][m] - \tilde{Y}_M[k]\right)}$$

$$\tilde{\sigma}_{G_{BLA}}^2[k] = \left|\tilde{G}_{BLA}[k]\right|^2 \left(\frac{\tilde{\sigma}_{YY}^2}{|\tilde{Y}_0|^2} + \frac{\tilde{\sigma}_{UU}^2}{|\tilde{U}_0|^2} - 2Real\left\{\frac{\tilde{\sigma}_{YU}^2}{\tilde{Y}_0\overline{\tilde{U}_0}}\right\}\right)$$

$$\tilde{\sigma}_{G_{BLA},n}^2[k] = \left|\tilde{G}_{BLA}[k]\right|^2 \left(\frac{\tilde{\sigma}_{YY,n}^2}{|\tilde{Y}_0|^2} + \frac{\tilde{\sigma}_{UU,n}^2}{|\tilde{U}_0|^2} - 2Real\left\{\frac{\tilde{\sigma}_{YU,n}^2[k]}{\tilde{Y}_0\overline{\tilde{U}_0}}\right\}\right)$$

4.5 Signal Processing Pipeline

The proposed real-time signal processing pipeline is shown in Figure 11.

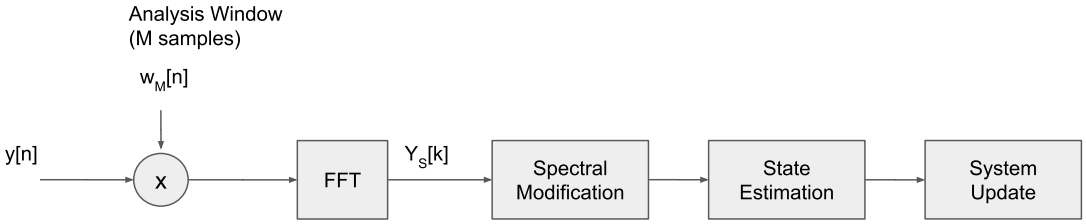


Figure 11: Digital signal processing pipeline for real-time estimation and of tracking spectral peaks

The deflection section is time-limited via the use an analysis window, which also has the benefit of smoothing the power spectrum. Next, the time-windowed dataset is Fourier transformed to produce the spectrum $Y_S[k]$.

If the deflection signal is not periodic in the number of samples applied to the analysis window, leakage effects may start to dominate the spectrum calculations. In this case analysis windows with aggressive stop-band rejection performance, like the Hanning window are preferred. In contrast, the rectangular window has excellent frequency resolution but suffers from leakage artifacts due to its weak rejection of the stop-band components. If the deflection signal is periodic, the rectangular window is the best choice.

The spectrum may simply be estimated from a smoothed version (high order of zero-padding or averaged over many periods) of the spectrum. More aggressive bandpass filtering may be desirable to isolate modes or harmonics that have been identified for imaging.

From the estimated spectrum envelope, frequency peaks are tracked over the duration of the lateral scanner. Finally, estimates are associated with a pixel in the output image space to produce the final image.

5 Results and Discussion

5.1 System Under Test (SUT)

The processing unit selected is Microchip's 32-bit SAME70 microcontroller. Its analog frontend supports the parallel operation of a DAC and two ADCs. System specifications are given in Table 4.

Symbol	Value	Note
f_{Master}	480 MHz	Master Clock Frequency
f_{CPU}	300 MHz	Processing frequency
f_{BUS}	150 MHz	Analog-frontend to CPU bus rate
$f_s = f_{ADC}$	73.2 KHz	Sampling frequency ADC Conversion rates of ADC1 and ADC2
$f_g = f_{DAC}$	73.2 KHz	Generator frequency DAC Conversion rate

Table 4: System-Under-Test specifications

5.2 System Identification

BLA estimations of the cantilever response from DC to ~36KHz using $7ms$ long uniform random phase multisine excitations are shown in Figure 12 and Figure 13 below.

Figure 12 are the G_{BLA} estimates using $M = 10$ realizations of the and $P = 20$ periods. Each color represents the results of a separate experiment for a total of sixteen experiments.

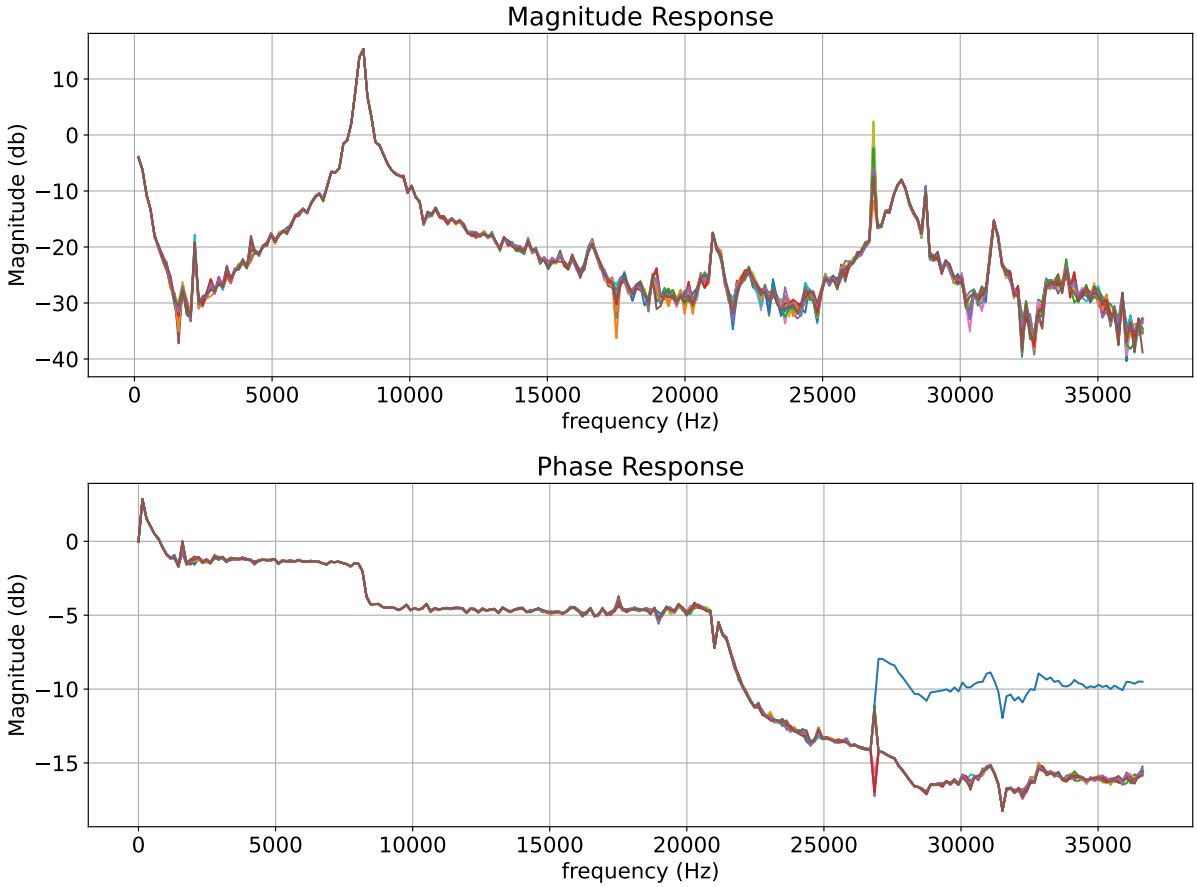


Figure 12: Magnitude and phase response of the BLA estimates for experiment 1

Figure 13 shows the G_{BLA} estimates using $M = 10$ realizations of the uniform random phase multisine and $P = 100$ periods. Each color represents the results of a separate experiment for a total of two experiments.

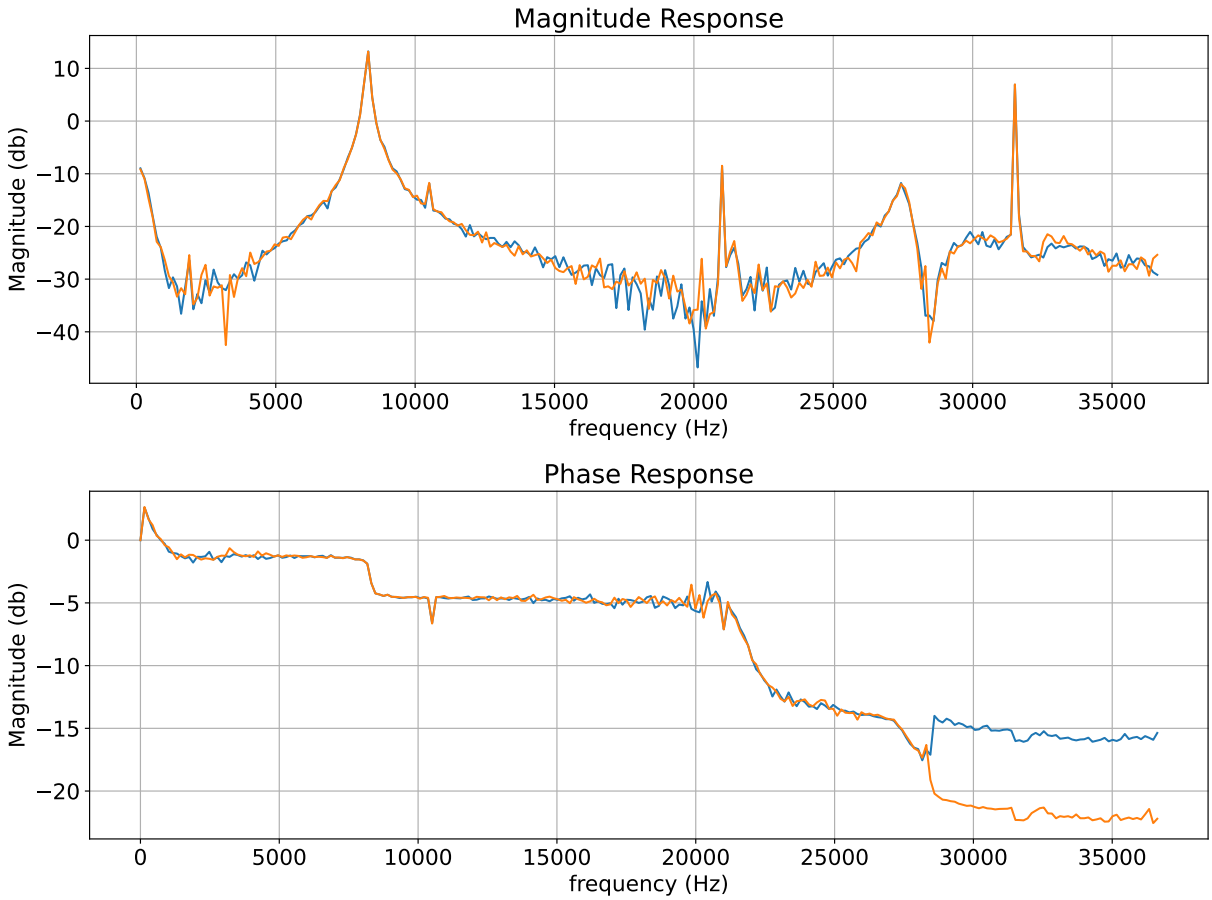


Figure 13: Magnitude and phase response of the BLA estimates for experiment 2

5.3 Imaging

The first resonant mode with a peak at 8.8KHz and Q of 30 is selected for peak tracking. Variations in peak amplitude are tracked and used in image generation. A grayscale image may be constructed with amplitude variations as the grayscale value (pixel intensity). This is shown in Figure 14. Darker pixels indicate increased local topographical height.

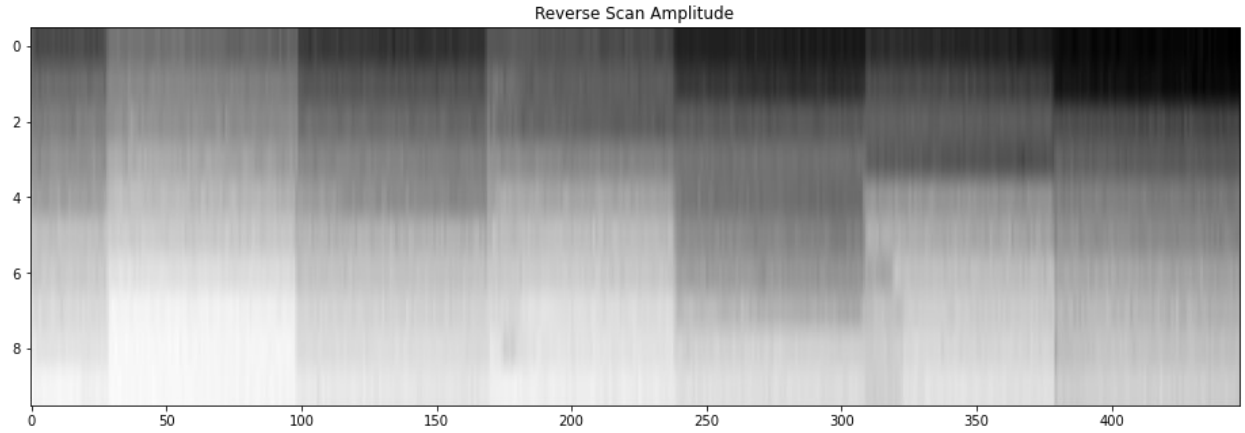


Figure 14: AFM image

The sample use for this AFM image is a standard AFM calibration sample. The size of the scanned area is $10um$ by $3um$. The axes indicate the number of samples used. The material of silicon carbide. It consists of a series of step-discontinuous grating structures with a height 85nm. The inter-grating pitch, measured as the center-to-center lateral separation between any two adjacent gratings is 3um.

There are a few observations worth investigating. The density of samples in the horizontal direction is much greater in the horizontal directions. This is because the rostering direction is biased in the horizontal direction. of the lateral scanners is in the horizontal scan. Over 450 amplitude estimates are used per each horizontal raster scan. There are only 10 lines scanned.

Given the sample specs, the expected AFM image would expect to see no height variations between any of the gratings anywhere throughout the whole image. In fact, the ideal image would show only two colors. One for the surface at the bottom of the trenches, one for the gratins some $85nm$ away. But the AFM image of Figure 14 clearly shows very dark pixels in the top-right-corner and distortion fades to very light pixel. It is difficult to explain the root cause of this distortion. One possibility could be that the initial alignment at the time of approach between the tip and the sample was off. However, the AFM tip has been laser cut precisely to accommodate any angle bias. It is also unlikely that the precisely manufactured AFM reference grating is physically warped over the imaged area. The other possible explanation is that the lateral scanners are introducing this distortion as they scan the sample. Further investigation is needed to root cause this.

5.4 Conclusions

This investigation set out to establish a framework for the transient operation of dynamic mode AFMs. As part of this effort, various signals and systems models were explored to best match the true characteristic of the deflection signal. Suitable signal processing techniques were investigated for their limitations and capabilities. A fast digital processing unit with an embedded analog frontend was proposed to operate the system. This would be used to oversample the deflection signal and take advantage of fast real-time computation to best estimate the true deflection while imaging. Formal system identification was carried out in the frequency domain. The procedure established the calculations required to characterize the signals and noise characteristics of the open-loop system. It was shown that with wideband random-phase multisine excitation signals are a robust class of signals that can help establish the best linear approximation. With a clear view of the spectrum, the spectral (modal/harmonic) component(s) may be identified and tracked in real-time to produce AFM images.

To validate the performance of the proposed system, a single-chip AFM was used for system identification and imaging. System identification revealed a rich view of the system's power spectrum. Multiple eigen modes and many harmonic components could be seen. Most impressively, the phase response showed clear step changes of π at the eigenmodes and no change in phase for the harmonics.

For open-loop transient imaging, the peak of the first eigenmode was chosen to be tracked. Amplitude variations were turned into an AFM image and plotted. The result showed a faithful recreation of the profile of the reference sample. However, distortion present in the image indicate the presence of some unmodeled process. Note that this indicates one of the major shortcomings of open-loop operation, that is, any unmodeled processes may lead to failure.

5.5 Future Work

This report has shown the richness of information that may be extracted from the AFM deflection signal. Furthermore, a simple implementation of the processing pipeline has produced and AM-AFM and shown very promising results. There are many experiments and investigations that can be done. Some suggestions are discussed below.

5.5.1 Multifrequency AFM

As discussed in Section 2.5.1, multifrequency techniques have been shown to produce excellent AFM images of the sample topography and mechanical properties. The proposed processing pipeline is well suited to track multiple features in the frequency domain. This would help investigate the how the various modes/harmonics behave while imaging and how robust it is to track them.

5.5.2 Input shaping

The ability to generate any arbitrary excitation signal and sample the response simultaneously allows for the possibility of optimizing the excitation signal to excite the cantilever at signals of interest only. This has the effect of whitening or flattening the spectrum everywhere except where the signal is. This

immediately increases the SNR of the sampled response and simplify the identification and tracking of the signal of interest.

5.5.3 Feedback

The pros and cons of feedback have been discussed throughout the report. As was shown in the generate AFM image in Figure 14, open-loop operation is extremely sensitive to random variations or unmodeled processes. Given the direct access to the drive signal, the system is already capable of adjust the drive signal based on system observations. A simple Proportional-Integral (PI) controller could be implemented to work in conjunction with the rest of the system. The combination of a real time digital signal processor and feedback-controller can prove to provide excellent SNR while imaging.

.

References

[1] G. Binnig and H. Rohrer, “Scanning tunneling microscopy,” *Surf Sci*, vol. 126, no. 1–3, pp. 236–244, Mar. 1983, doi: 10.1016/0039-6028(83)90716-1.

[2] G. Binnig, C. F. Quate’ ’, E. L. Gi, and C. Gerber, “Atomic Force Microscope.”

[3] R. Garcí and A. San Paulo, “Attractive and repulsive tip-sample interaction regimes in tapping-mode atomic force microscopy.”

[4] Á. S. Paulo and R. García, “Tip-surface forces, amplitude, and energy dissipation in amplitude-modulation (tapping mode) force microscopy,” *Phys Rev B Condens Matter Mater Phys*, vol. 64, no. 19, 2001, doi: 10.1103/PhysRevB.64.193411.

[5] T. Ando, T. Uchihashi, and S. Scheuring, “Filming biomolecular processes by high-speed atomic force microscopy,” *Chem Rev*, vol. 114, no. 6, pp. 3120–3188, Mar. 2014, doi: 10.1021/CR4003837/ASSET/IMAGES/MEDIUM/CR-2013-003837_0045.GIF.

[6] Y. Sakiyama, A. Mazur, L. E. Kapinos, and R. Y. H. Lim, “Spatiotemporal dynamics of the nuclear pore complex transport barrier resolved by high-speed atomic force microscopy,” *Nature Nanotechnology 2016 11:8*, vol. 11, no. 8, pp. 719–723, May 2016, doi: 10.1038/nnano.2016.62.

[7] G. E. Fantner, R. J. Barbero, D. S. Gray, and A. M. Belcher, “Kinetics of antimicrobial peptide activity measured on individual bacterial cells using high-speed atomic force microscopy,” *Nat Nanotechnol*, vol. 5, no. 4, pp. 280–285, 2010, doi: 10.1038/NNANO.2010.29.

[8] J. Tamayo and R. García, “Relationship between phase shift and energy dissipation in tapping-mode scanning force microscopy,” *Appl Phys Lett*, vol. 73, no. 20, pp. 2926–2928, 1998, doi: 10.1063/1.122632.

[9] N. Sasaki and M. Tsukada, “Theory for the effect of the tip-surface interaction potential on atomic resolution in forced vibration system of noncontact AFM.”

[10] N. Sarkar, R. R. Mansour, O. Patange, and K. Trainor, “CMOS-MEMS atomic force microscope,” *2011 16th International Solid-State Sensors, Actuators and Microsystems Conference, TRANSDUCERS’11*, pp. 2610–2613, 2011, doi: 10.1109/TRANSDUCERS.2011.5969836.

[11] M. G. Ruppert, A. G. Fowler, M. Maroufi, and S. O. R. Moheimani, “On-chip dynamic mode atomic force microscopy: A silicon-on-insulator MEMS approach,” *Journal of Microelectromechanical Systems*, vol. 26, no. 1, pp. 215–225, Feb. 2017, doi: 10.1109/JMEMS.2016.2628890.

[12] T. R. Albrecht, P. Grütter, D. Horne, and D. Rugar, “Frequency modulation detection using high-Q cantilevers for enhanced force microscope sensitivity,” *J Appl Phys*, vol. 69, no. 2, pp. 668–673, 1991, doi: 10.1063/1.347347.

[13] H.-J. Butt and M. Jaschke, “Calculation of thermal noise in atomic I force microscopy,” *Nanotechnology*, vol. 6, pp. 1–7, 1995.

[14] K. Xu *et al.*, “Recent development of PeakForce Tapping mode atomic force microscopy and its applications on nanoscience,” *Nanotechnol Rev*, vol. 7, no. 6, pp. 605–621, Dec. 2018, doi: 10.1515/NTREV-2018-0086.

[15] B. v. Derjaguin, V. M. Muller, and Y. P. Toporov, “Effect of contact deformations on the adhesion of particles,” *J Colloid Interface Sci*, vol. 53, no. 2, pp. 314–326, Nov. 1975, doi: 10.1016/0021-9797(75)90018-1.

[16] “Surface energy and the contact of elastic solids,” vol. 324, pp. 301–313, 1971, Accessed: Dec. 27, 2022. [Online]. Available: https://royalsocietypublishing.org/

[17] “Intermolecular and Surface Forces, 3rd Edition.” https://learning-oreilly-com.proxy.lib.uwaterloo.ca/library/view/intermolecular-and-surface/9780123751829/OEBPS/B978012375182910025 9.htm (accessed Dec. 27, 2022).

[18] C. Argento and R. H. French, “6081 Cite as,” *J Chem Phys*, vol. 80, p. 668, 1996, doi: 10.1063/1.363680.

[19] D. Lagrange, N. Mauran, L. Schwab, and B. Legrand, “Low Latency Demodulation for High-Frequency Atomic Force Microscopy Probes,” *IEEE Transactions on Control Systems Technology*, vol. 29, no. 5, pp. 2264–2270, Sep. 2021, doi: 10.1109/TCST.2020.3028737.

[20] M. G. Ruppert, K. S. Karvinen, S. L. Wiggins, and S. O. R. Moheimani, “A Kalman Filter for Amplitude Estimation in High-Speed Dynamic Mode Atomic Force Microscopy,” *IEEE Transactions on Control Systems Technology*, vol. 24, no. 1, pp. 276–284, Jan. 2016, doi: 10.1109/TCST.2015.2435654.

[21] M. R. P. Ragazzon, M. G. Ruppert, D. M. Harcombe, A. J. Fleming, and J. T. Gravdahl, “Lyapunov Estimator for High-Speed Demodulation in Dynamic Mode Atomic Force Microscopy,” *IEEE Transactions on Control Systems Technology*, vol. 26, no. 2, pp. 765–772, Mar. 2018, doi: 10.1109/TCST.2017.2692721.

[22] M. S. Rana, H. R. Pota, and I. R. Petersen, “Spiral Scanning with improved control for faster imaging of AFM,” *IEEE Trans Nanotechnol*, vol. 13, no. 3, pp. 541–550, 2014, doi: 10.1109/TNANO.2014.2309653.

[23] N. Nikooienejad, A. Alipour, M. Maroufi, and S. O. R. Moheimani, “Video-Rate Non-Raster AFM Imaging with Cycloid Trajectory,” *IEEE Transactions on Control Systems Technology*, vol. 28, no. 2, pp. 436–447, Mar. 2020, doi: 10.1109/TCST.2018.2879939.

[24] A. Bazaei, Y. K. Yong, and S. O. Reza Moheimani, “Combining Spiral Scanning and Internal Model Control for Sequential AFM Imaging at Video Rate,” *IEEE/ASME Transactions on Mechatronics*, vol. 22, no. 1, pp. 371–380, Feb. 2017, doi: 10.1109/TMECH.2016.2574892.

[25] M. S. Rana, H. R. Pota, and I. R. Petersen, “High-speed AFM image scanning using observer-based MPC-notch control,” *IEEE Trans Nanotechnol*, vol. 12, no. 2, pp. 246–254, 2013, doi: 10.1109/TNANO.2013.2243752.

[26] H. Xie, Y. Wen, X. Shen, H. Zhang, and L. Sun, “High-Speed AFM Imaging of Nanopositioning Stages Using H∞ and Iterative Learning Control,” *IEEE Transactions on Industrial Electronics*, vol. 67, no. 3, pp. 2430–2439, Mar. 2020, doi: 10.1109/TIE.2019.2902792.

[27] R. Garcia and E. T. Herruzo, “The emergence of multifrequency force microscopy,” *Nature Nanotechnology*, vol. 7, no. 4. Nature Publishing Group, pp. 217–226, 2012. doi: 10.1038/nnano.2012.38.

[28] J. R. Lozano and R. Garcia, “Theory of multifrequency atomic force microscopy,” *Phys Rev Lett*, vol. 100, no. 7, Feb. 2008, doi: 10.1103/PhysRevLett.100.076102.

[29] R. W. Stark and W. M. Heckl, “Higher harmonics imaging in tapping-mode atomic-force microscopy,” *Review of Scientific Instruments*, vol. 74, no. 12, pp. 5111–5114, Dec. 2003, doi: 10.1063/1.1626008.

[30] R. W. Stark, “Spectroscopy of higher harmonics in dynamic atomic force microscopy,” *Nanotechnology*, vol. 15, no. 3, pp. 347–351, Mar. 2004, doi: 10.1088/0957-4484/15/3/020.

[31] E. T. Herruzo, A. P. Perrino, and R. Garcia, “Fast nanomechanical spectroscopy of soft matter,” *Nat Commun*, vol. 5, Jan. 2014, doi: 10.1038/ncomms4126.

[32] S. Belikov, J. Alexander, M. Surtchev, and S. Magonov, “Digital Q-Control and Automatic Probe Landing in Amplitude Modulation Phase Imaging AFM Mode,” in *IFAC-PapersOnLine*, Jul. 2017, vol. 50, no. 1, pp. 10882–10888. doi: 10.1016/j.ifacol.2017.08.2445.

[33] M. G. Ruppert and S. O. R. Moheimani, “Multimode Q Control in Tapping-Mode AFM: Enabling Imaging on Higher Flexural Eigenmodes,” *IEEE Transactions on Control Systems Technology*, vol. 24, no. 4, pp. 1149–1159, Jul. 2016, doi: 10.1109/TCST.2015.2478077.

[34] J. Mertz, O. Marti, and J. Mlynek, “Regulation of a microcantilever response by force feedback,” *Appl Phys Lett*, vol. 62, no. 19, pp. 2344–2346, 1993, doi: 10.1063/1.109413.

[35] R. W. Stark, “Time delay Q-control of the microcantilever in dynamic atomic force microscopy,” in *2005 5th IEEE Conference on Nanotechnology*, 2005, vol. 1, pp. 435–438. doi: 10.1109/nano.2005.1500745.

[36] N. Nikooienejad and S. O. R. Moheimani, “Frequency Domain-Based Integral Resonant Control Design for a MEMS Nanopositioner,” in *CCTA 2021 - 5th IEEE Conference on Control Technology and Applications*, 2021, pp. 874–879. doi: 10.1109/CCTA48906.2021.9658591.

[37] Z. Wang *et al.*, “Wavelet analysis of higher harmonics in tapping mode atomic force microscopy,” *Micron*, vol. 118, pp. 58–64, Mar. 2019, doi: 10.1016/j.micron.2018.12.007.

[38] A. F. Payam, P. Biglarbeigi, A. Morelli, P. Lemoine, J. Mclaughlin, and D. Finlay, “Data acquisition and imaging using wavelet transform: a new path for high speed transient force microscopy,” *Nanoscale Adv*, vol. 3, no. 2, pp. 383–398, Jan. 2021, doi: 10.1039/d0na00531b.

[39] V. Pukhova, F. Banfi, and G. Ferrini, “Transient eigenmodes analysis of single-impact cantilever dynamics combining Fourier and wavelet transforms,” *Nanotechnology*, vol. 26, no. 17, p. 175701, Apr. 2015, doi: 10.1088/0957-4484/26/17/175701.

[40] R. M. Gray, “Quantization Noise Spectra,” *IEEE Trans Inf Theory*, vol. 36, no. 6, pp. 1220–1244, 1990, doi: 10.1109/18.59924.

[41] S. Mallat, “A Wavelet Tour of Signal Processing: The Sparse Way,” *A Wavelet Tour of Signal Processing: The Sparse Way*, pp. 1–805, Dec. 2008, doi: 10.1016/B978-0-12-374370-1.X0001-8.

[42] J. W. Cooley and J. W. Tukey, “An algorithm for the machine calculation of complex Fourier series,” *Math Comput*, vol. 19, no. 90, pp. 297–301, 1965, doi: 10.1090/S0025-5718-1965-0178586-1.

[43] J. O. Smith, “AUDIO FFT FILTER BANKS.” [Online]. Available: http://ccrma.stanford.edu/

[44] R. Kumaresan, C. S. Ramalingam, and D. van Ormondt, “Estimating the parameters of NMR signals by transforming to the frequency domain,” *Journal of Magnetic Resonance (1969)*, vol. 89, no. 3, pp. 562–567, Oct. 1990, doi: 10.1016/0022-2364(90)90339-B.

[45] “System Identification: A Frequency Domain Approach | IEEE eBooks | IEEE Xplore.” https://ieeexplore.ieee.org/book/6198969 (accessed Dec. 21, 2022).