High performance nano-scale measurements by advanced atomic force estimation

Cui, Song

2010

Cui, S. (2010). High performance nano-scale measurements by advanced atomic force estimation. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/46322>

<https://doi.org/10.32657/10356/46322>

High Performance Nano-scale Measurements by Advanced Atomic Force Estimation

Cui Song

School of Electrical & Electronic Engineering

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2010

Statement of Originality

I hereby certify that the content of this thesis is the result of work done by me and has not been submitted for a higher degree to any other university or institution.

.....

Date

.....

Cui Song

Acknowledgements

It is my pleasure to thank many people who had offered generous advices and helps during my four years of Ph. D study.

First of all, I am grateful to my supervisor Prof. Soh Yeng Chai. His vision enlightens me to work towards the proposed research directions. I also appreciate the detailed discussions with him which help me understand the essentials of true research, how research work should be carried out and how the research problems could be tackled.

Secondly, I would like to express my warm and sincere thanks to my co-supervisor Dr. Lim Ser Yong. His support has influenced my chosen research directions.

Throughout my research studies, I have benefitted greatly from close interactions with Dr. Zhang Ying. He introduced me to the topic of nano-scale measurement during my undergraduate study. Since then, I have benefited a lot from his suggestions and guidance.

I would also want to thank Dr. Kok Shaw Wei, Dr. Ng Boon Ping, Mr. Dong Zhao-gang, Mr. Liu Zhuang and Mr. Seck Hon Luen of Singapore Institute of Manufacturing Technology (SIMTech) for their discussions in theoretical derivations and experiments. I am also thankful to Mr. Li Guoqi, Mr. Luo Dan, Ms. Wang Wei and Mr. Zhou Yong

of NTU for their friendship and fruitful discussions.

Last but not least, I am grateful to all my family members for their love, understanding and support during these years. I would like to thank my parents and parents-in-law for their understanding of my devotion of time and efforts to research. My deepest appreciation goes to my wife Ms. Luo Xuewen. Her love motivates me to overcome the most difficult issues in my research studies and I deeply appreciate her sacrifices in personal interests and time in order to make a good home environment for me to undertake my research.

Table of contents

Acknowledgements	ii
Preface	iv
1 Introduction	1
1.1 Background	1
1.2 Motivations and Contributions	3
1.3 Organization of the Thesis	9
2 Principle of AFM	12
2.1 Introduction	12
2.2 The Probe-sample Interaction	12
2.3 Operation Mode	15
2.3.1 Contact Mode	16
2.3.2 Non-contact Mode	16
2.3.3 Tapping Mode	17
2.4 Theoretical Model	18
2.5 Principle of AFM Sensing	20

2.6	Positioning Regulation in AFM	23
3	Linearity Indices and Linearity Improvement of 2-D Tetra-lateral Po-	
	sition Sensitive Detector	24
3.1	Introduction	24
3.2	Mathematical Model	27
3.3	Linearity Indices	29
3.4	New Estimation Formulae	34
3.5	Simulation and Experimental Results	37
3.6	The Effect of Spot Size on Our Proposed Formulae	45
3.6.1	Theoretical Analysis	47
3.6.2	Simulation Results	50
3.7	Conclusions	51
4	Improved Accuracy of 2-D Non-contact Range Sensors Using 2-D Tetra-	
	lateral PSD	54
4.1	Introduction	54
4.2	2-D Non-contact Range Sensor Based on Triangulation Principle	55
4.3	Proposal of New Formula and Simulation Study	59
4.4	Conclusion	64
5	Improved Measurement Accuracy of the Quadrant Detector Through	
	Improvement of Linearity Index	65
5.1	Introduction	65

5.2	Theoretical Modelling	67
5.3	The New Estimation Formulae	69
5.3.1	The Uniform Circular Beam Profile	71
5.3.2	The Gaussian Circular Beam Profile	73
5.4	Simulation Results	75
5.5	Experimental Results	79
5.6	Conclusions	83
6	A Novel Probe-sample Separation Estimation Scheme for Atomic Force	
	Microscopy	84
6.1	Introduction	84
6.2	Theoretical Model	87
6.3	Proposed Estimation Scheme	89
6.4	Simulation Results and Discussions	92
6.5	Conclusions	95
7	An Accurate Separation Estimation Algorithm for the Casimir Oscil-	
	lator	96
7.1	Introduction	96
7.2	Review of the Casimir Force	100
7.3	Mathematical Model of the Casimir Oscillator	104
7.4	Dynamics of The Casimir Oscillator	106
7.4.1	Small Nonlinearity	106
7.4.2	Hysteresis	106

7.4.3	Period Doubling and Chaos	108
7.5	Gap Separation Estimation Scheme	109
7.6	Simulation Results	114
7.6.1	Small Nonlinearity	115
7.6.2	Hysteresis	117
7.6.3	Period Doubling and Chaos	118
7.7	Conclusions	119
8	Improved Separation Gap Estimation for Dynamic Systems Actuated by Casimir Force	121
8.1	Introduction	121
8.2	Problem Formulation	123
8.3	Our Estimation Method	125
8.4	Simulation Results	128
8.4.1	Ideal Casimir Force Model	128
8.4.2	Modified Casimir Force Model	131
8.5	Conclusions	134
9	Conclusions and Future Work	137
9.1	Conclusions	137
9.2	Future Work	140
	Author's Journal Publications	142
	Bibliography	144

A Properties of Cost Function Eq. (7.13)	163
B Properties of Cost Function Eq. (8.9)	166

Preface

Atomic force microscopy (AFM) was invented by G. Binnig and his collaborators in 1986 after the invention of scanning tunnelling microscopy (STM) in the early 1980s. The AFM system deploys a tiny probe to interact with the sample such that topography information of the sample can be obtained. The resolutions of the AFM topography images are in the nanometer scale and they are limited only by the probe diameters. Besides, in AFM imaging, there is no requirement on sample preparation and conductivity. Thus, the AFM provides an easy platform to observe, measure and manipulate atomic interactions. Since its invention, AFM has become an indispensable tool in the research areas of material sciences and biological sciences. The AFM is also widely used in many industry applications in microelectronics and data storage.

Although many research areas and industry applications have benefited from the success of AFM, there is still an urgent need to improve the measurement resolution and accuracy of the AFM in order to further exploit the atomic characteristics at the scales of atoms or even smaller. In addition, AFM has been modified as a Casimir oscillator to study quantum fluctuations where highly accurate and precise measurement is also desired. Driven by this motivation, this thesis will focus on the improvement of accuracy

and resolution of the AFM and its modified system, the Casimir oscillator.

Our approach to accuracy and resolution enhancement of the AFM system is by tackling the measurement errors that arise from its components such as the position sensitive detector (PSD) and the sensing principle in the AFM sensing system. In our work, we have proposed a new concept of linearity indices to quantitatively measure the measurement inaccuracies of various PSDs. Based on these linearity indices, new design formulae are proposed for 2-D tetra-lateral PSD and quadrant detector. Their measurement accuracy are significantly improved via our new formulae as well as robustness to the incident beam size and parameter variation. Besides the AFM sensing system, inaccuracies also come from the probe-sample separation estimation. To address this issue, a novel estimation algorithm has been proposed and accurate estimates can be obtained regardless of different nonlinear dynamics. Several methods have also been derived to improve the separation gap estimation in Casimir oscillator which is a modified AFM system.

Chapter 1

Introduction

1.1 Background

Atomic force microscopy (AFM) belongs to the family of scanning probe microscopy (SPM) which uses a probe to interact with the sample via scanning line by line. Vital information of the sample such as topography and optical properties can then be obtained. The probe has to be very sharp with a probe radius of less than 10 *nm* and the height of the probe is around 10 μm . The first family member of the SPM is scanning tunneling microscopy (STM) which was invented in the early 1980s. With a conductive probe, the STM can be used to image the topography of conductive samples. The principle of STM is based on quantum tunneling concept. When the conductive probe is brought near to the conductive sample, electrons can tunnel through the separation between them if a low level bias is applied [1]. The tunneling current varies if the separation between the probe and the sample changes. In order to image the topography of non-conductive samples, the AFM was invented and its principle lies in sensing the change of the atomic force between the probe and the sample when there is a change in the separation gap [2]. Other than sample topography, optical properties, magnetic

properties and other material properties can also be exploited by using other SPMs like scanning near field microscopy (SNOM) which are built on top of the AFM. Compared to the resolution of the optical microscopy which is at most 200 *nm* in practice due to the diffraction limit, the vertical resolution of the AFM is below 0.01 *nm* and the lateral resolution is less than 0.1 *nm* [3]. Besides the topography measurement, the AFM can be used to fabricate nanometer-scale structures where molecules are transferred by the probe from other places to the sample, similar to the process of inking on paper by using a pen [4]. Last but not least, there is little requirement on the sample preparation and wet samples can also be imaged by the AFM. This is a significant breakthrough compared to the compulsory coating of the sample and the vacuum environment required by electron microscopy. Currently, the AFM is widely used in the semiconductor and data storage industries where it is used to measure the size and the shape of particles. Researchers in material sciences and life sciences also use AFM to visualize and measure particles or structures at or below the molecule level.

There are two types of AFM in the market, the tuning fork based AFM and the cantilever based AFM. In the tuning fork based AFM, the fork is mounted in a similar manner as in a watch: its base is fixed and both prongs are oscillating in opposition to each other. The dynamical forces of the two prongs are then compensated in the base [5]. The symmetry of the prongs is broken when one of them is subject to a probe-sample interaction dominated by the shear force. The deflection of the prongs will be changed and it can be detected by measuring the current. The cantilever based AFM system consists of a probe grown on a cantilever, a position sensitive detector (PSD) to

measure the cantilever deflection indirectly and a feedback loop to adjust the position of the sample to maintain a constant separation during scanning. The probe-sample interaction in the cantilever based AFM is the force interaction. When the probe is near the sample, the separation dependent atomic force between the probe and the sample will lead to deflection of the cantilever. Usually, the deflection is in the range of nanometers which makes it difficult to measure. To provide an easy measurement, a laser light is incident on the cantilever and reflected back to PSD. The cantilever deflection will lead to a change of the laser optical path which in turn alters the position of the light spot on PSD. Then the movement of the light spot on PSD in the range of millimeter is proportional to the cantilever deflection. Readout circuits can be implemented readily to measure the position of the light spot on PSD and the cantilever deflection can be deduced from the spot position measurement. A piezoelectric actuator is used to adjust the probe-sample separation. The adjustment is based on the information of the force measurement in order to maintain a constant separation during the scanning process. If the constant separation is maintained, the vertical adjustment of piezoelectric actuator during the scanning is exactly the topography of the sample.

1.2 Motivations and Contributions

Although the AFM is becoming more and more important in many scientific areas and industries, it is still desirable to improve the accuracy, repeatability and resolution of the AFM system if one wants to further exploit and observe the physical phenomena in an even smaller scale. Driven by this motivation, our research is focused on the

improvement of the accuracy and the resolution of the cantilever based AFM system. In addition, our research work is also extended to the measurement accuracy improvement of the Casimir oscillator which is a modified AFM system.

The accuracy and the resolution of the whole AFM system is mainly limited by the accuracy and the resolution of each component. For example, the quadrant detector (QD), which is the core of the AFM sensing system, suffers from a serious nonlinearity problem and the measurement error is significant when the laser spot is away from the center of the PSD and the measurement error is dependent on the beam intensity profile. Here, the accuracy of the PSD measurement is defined as measurement error of less than 1%. This problem seriously degrades the measurement accuracy of the whole AFM system as it "sees" the atomic interaction. It is not possible to know the atomic physics if the "eye" is blurred. In addition, the resolution of the AFM is also limited as only the center area of the PSD has minimum measurement errors. On the other hand, the QD has been widely used in machine alignment, robot vision and other applications. All these applications require an accurate PSD sensing with high resolution. In our research, another type of PSD, 2-D tetra-lateral PSD is considered to replace the QD due to its advantage that its measurement errors is independent of the laser beam profile but only dependent on the centroid of the spot. It eliminates the unknown factor of the beam profile. However, it has the same problem as the QD in terms of serious measurement errors when computing the spot centroid by the conventional formula which was obtained by experience without an in-depth analysis. Although many approaches have been suggested to improve its accuracy, their tedious

procedures make them unsuitable in real implementation. After a rigorous theoretical modelling and analysis, the concept of linearity indices is proposed to measure the measurement errors quantitatively. Based on the proposed linearity indices, a new and simple formula is proposed to calculate the spot centroid. The effective area of the PSD with various measurement errors is enlarged significantly. This comparison has been tabulated as some PSD applications may have their own criteria on the accuracy. Thus, readers can judge the significance and contribution of our method based on the detailed quantitative comparison from their own perspectives. Simulations and experiments have shown that it is effective in the real environment. In practice, the light beam projected on the PSD may have a large beam size in PSD applications. And it is difficult to measure and control the beam size. Thus, it is of interest to investigate the effect of beam size on the accuracy improvement achieved by our newly proposed formula. To facilitate the simulation process, a theoretical analysis is done to simplify the numerical calculation of the beam position estimates when the beam size is considered. Our simulation results have shown that the performance of our new formula is not degraded even if an extremely large light spot is projected on the PSD. In addition, measurement errors exist not only in the PSD itself but also in the reflection paths of the laser based on the triangulation principle. Another set of formula is thus proposed to reduce the overall cascaded errors for the AFM sensing system. Last but not least, the concept of linearity indices can be applied to improve the measurement accuracy of other PSDs as well. As mentioned before, the QD also suffers from the nonlinearity problem. We note that the measurement error of the QD is dependent on the beam intensity profile.

Its advantages of fast response and low cost make it a suitable choice as a position detector in many industry applications. With respect to different beam profiles, a series of calculation formulae for the QD are proposed. Simulations and experiments show that the measurement accuracy is greatly improved.

Inaccuracies of the AFM system also arise from the estimated separation gap between the probe and the sample. The probe-sample separation is estimated from the probe-sample interaction. The interaction in the AFM is dominated by the atomic force between the probe and the sample where its strength and characteristics are highly related to the atomic structure of the material and the probe-sample separation. The probe dynamics can be described by a second order ordinary differential equation subject to the nonlinear separation dependent atomic force as input. Rich information about the atomic force and the separation gap is embedded in the system dynamics but unveiling the information is not an easy task because unknown parameters are nonlinearly parameterized in the system. In theory, there is no guarantee that the designed algorithm can accurately estimate the probe-sample separation as well as other atomic force related parameters from the states of the probe dynamics. In practice, it is assumed that the amplitude of the probe vibration has a linear relationship with the probe-sample separation which is not always true even under the normal operation condition [6], and there is no available method to estimate other force related parameters. Inaccurate measurements have sometimes been obtained and it is not possible to know when the measurement error is produced in the imaging process. Our efforts are focused on how to estimate the probe-sample separation and other force related parameters with min-

imized errors under the normal condition. Under the normal operation condition, the probe dynamics are dominated by the resonant sinusoidal with weak harmonics. While most researchers are still finding ways to extract the information from the harmonics rather than the fundamental sinusoidal only, the formula to calculate the unknown probe-sample separation and other parameters is derived from the information of the amplitude and the phase of the fundamental sinusoidal and its higher order harmonics. Simulation results show that the separation estimation is accurate and robust.

The atomic force consists of several forces from different origins. The Casimir force is one of them. The Casimir force exists between two conductive materials. It originates from the density change of the electromagnetic waves imposed by the conductive materials. The understanding and measurement of the Casimir force is highly important in designing modern micro- and nanoelectromechanical systems (MEMS and NEMS) [7] and understanding fundamental physics [8]. The Casimir oscillator is a dynamic oscillating device where the interaction force between the conductive probe and the conductor surface is only the Casimir force [9]. By eliminating the influence of the electrostatic force, the AFM can also function as a Casimir oscillator. It is found that the vibration amplitude of the Casimir oscillator is sensitive with respect to the change of the separation between the two conductors. Thus, the Casimir oscillator is a potential candidate as a position sensor at the nanometer scale. But its nonlinear dynamics make the separation estimation difficult. In our simulations, different nonlinear phenomena are observed under different conditions. To estimate the separation accurately even when the system is experiencing complicated nonlinear dynamics, we propose a novel

nonlinear parameter estimation method. The rational of our method is that we first simplify our problem of estimating multiple unknown parameters to a single unknown parameter estimation problem by constructing a cost function with only one local minimum and no local maximum with respect to the difference between the single unknown parameter and its estimate. Thus, the golden section search method can be readily applied to search for the local minimum of the cost function and the accurate estimate. Theories and simulations showed that our method is accurate and robust. Other than the separation gap estimation in the Casimir oscillator, our scheme is also applicable to many other parameter estimation problems with similar mathematical models like the gap separation estimation problem in electrostatically actuated MEMS [10]-[12]. Based on the same rationale, a modified algorithm is proposed to estimate the separation in all dynamic systems actuated by the Casimir force. This method is based on the modified AFM setup. It is shown in simulations that our modified algorithm is simpler and more robust to noise compared to our algorithm developed for Casimir oscillator in Chapter 7. Our modified algorithm is also effective even when real experimental conditions like the finite conductivity of the interacting materials and the surface roughness are considered.

In conclusion, it is believed that the results achieved will form a systematic framework to improve the accuracy and the resolution of AFM system and the Casimir oscillator which is a modified AFM.

1.3 Organization of the Thesis

The thesis is organized as follow. In Chapter 1, the history and background of the AFM system is presented. The motivation and importance of our proposed research direction is also discussed this chapter followed by a brief description of our research contributions towards the identified research directions. To improve the measurement accuracy and the resolution of the AFM related system, it is important to understand the working principle of the AFM. Thus, a detailed description of its working mechanism and its three core subsystems, the sensing system, the probe-sample interaction, and the regulation system is discussed in Chapter 2. The measurement errors of the AFM system is partially due to the measurement errors of the position sensitive detector (PSD) which is deployed in the sensing system. In Chapter 3, the concept of linearity indices is proposed to measure the degree of measurement errors of various PSDs. Based on the linearity indices, a new and simple calculation formula is proposed to estimate the light spot position projected on a 2-D tetra-lateral PSD which belongs to the PSD family. Simulations and experiments have shown that the accuracy of the position estimates calculated from the new formula are improved significantly. The effective area where the measurement accuracy can be trusted is significantly enlarged. More than that, the effect of spot size on our new accuracy improvement scheme is also studied. Simulations have shown that the effective area with distortion less than 1% is enlarged 5 times when our new formulae are used even if the beam has a beam size which is 5% of the side length of the PSD. Thus, the 2-D tetra-lateral PSD with our new formula can be implemented in the AFM sensing system to improve the overall

measurement accuracy and resolution. Besides the measurement errors of the PSD, the sensing system itself also suffers from a nonlinearity problem. In Chapter 4, the overall measurement errors of the AFM sensing system is considered and a novel estimation algorithm is proposed to reduce the overall measurement nonlinearity. Actually, the concept of linearity indices can be applied to other PSD family members as well. In Chapter 5, we have applied this concept to the linearity improvement of the quadrant detector (QD) as an example. A series of calculation formulae are proposed to estimate the spot position on QD with respect to different beam profiles. From simulations and experiments, it can be seen that the effective area with certain measurement distortions is greatly improved. Besides the measurement errors from the AFM sensing system, the estimation errors from the probe-sample separation estimation also contribute to the measurement errors of the AFM. In Chapter 6, a novel separation estimation algorithm is proposed to provide an accurate estimate of the probe-sample separation. The Casimir oscillator is actuated purely by the Casimir force to study the quantum fluctuations which can be realized by a modified AFM system. But the challenge to provide an accurate estimate of the separation gap when nonlinear dynamics are observed in the Casimir oscillator bottlenecks its development as a nanometric sensor. In Chapter 7, we have simulated various nonlinear dynamics in the Casimir oscillator under different conditions. An accurate estimation scheme is then proposed to estimate the separation gap. In Chapter 8, a modified separation estimation algorithm which is easier to apply and more robust to noise is proposed for the Casimir oscillator. Real conditions like the finite conductivity of the interacting materials and the surface roughness are also

considered in our modified algorithm. Last but not least, a conclusion is drawn in Chapter 9, together with several areas and ideas for future work.

Chapter 2

Principle of AFM

2.1 Introduction

As described earlier, the function of AFM is to image the topography of the sample in nanometer scale. This relies on the measurement of the separation dependent atomic force between the probe and the sample. The information on the topography can be extracted from the atomic force measurement. Although there are different types of AFM from different manufacturers, most cantilever based AFMs contain three major parts, the probe-sample atomic force interaction, the AFM sensing system where a single piece of PSD is deployed to measure the interaction indirectly, and a feedback loop with PZT to regulate the scanning process. The detailed set up is shown in Figure 2.1.

2.2 The Probe-sample Interaction

The resolution limit of the AFM depends on the sharpness of the probe. Indeed, a sharp probe is needed for AFM imaging at the scale of nanometer. To facilitate probe fabrication, the probe is grown on a cantilever. An image of the probe with cantilever is taken under a microscope (Figure 2.2). The cantilever-probe assemble is made of

Figure 2.1: The cantilever based AFM.

Silicon or Si_3N_4 . Currently, the radius of the probe vertex is only 10 nm and the thickness of the cantilever is around several micrometer. Different cantilevers made of different materials are used for a vast variety of samples. The essential guiding principle is that soft cantilevers are more suitable for soft samples like biological cells and stiff cantilevers are made for hard samples like the metal surfaces.

When the separation between the probe and the sample is below 100 nm , an interaction force can be detected [13]. Depending on the environment, the dominating interaction force can be van der Waals' force, capillary force, chemical bonding or electrostatic force. In the dry air environment, the force between the probe and the sample is mainly the van der Waals' force. And the Lennard-Jones potential [14] due to the van der Waals' force between two molecules is given by:

$$w(r) = \frac{c_1}{r^{12}} - \frac{c_2}{r^6} \quad (2.1)$$

where r is the distance between two molecules and c_1 and c_2 are two constants. Generally,

Figure 2.2: The probe grown on the Cantilever.

if the probe is considered as a sphere and the sample surface is considered as a flat surface, the van der Waals' force can be expressed as:

$$F(z) = \frac{B}{z^8} - \frac{A}{z^2} \quad (2.2)$$

where z refers to the distance between the probe center and the sample flat surface. Parameters A and B relate to the material properties. The positive sign indicates that the force is repulsive. Besides, there are some other models to describe the atomic force. For example, in order to avoid the divergence problem when the separation becomes zero, Hoffman [15] proposed that the atomic force is expressed as:

$$F(z) = -\frac{AR^2}{6} \frac{1}{z^3(1 - \exp((b_0 - z)/b_0)) + b_0^3 \exp((b_0 - z)/b_0)} \quad (2.3)$$

where R is the probe radius, A and b_0 are constants. Last but not least, DMT theory is also used to describe the repulsive force [16, 17] when the separation is less than intermolecular distance a_0 :

$$F(z) = \frac{-HR}{6a_0^2} + E\sqrt{R}(a_0 - z)^{3/2} \quad (2.4)$$

where E is the elastic modulus and R is the probe radius. In humid environment, capillarity is a general phenomenon that occurs between two contacting surfaces. When the probe and the sample are in contact, the water in the overlapping region is displaced and accumulated at the edge of the contact area. This amount of water is assumed to form a meniscus when the probe is retracted [18]. In that case, the capillary force has a strong hysteresis where the approaching force and the retraction force are different. In addition, Fa [19] found that if the sample is in aqueous solutions, the electrical double-layer force due to the difference in surface potential exists. And patch charges on the probe and the sample, if any, will interact before the contact [20].

2.3 Operation Mode

Currently, there are several operation modes of AFM when used to measure the interaction force for a large range of research areas. How to choose a suitable operation mode depends on the measurement environment and specific application requirements. A brief summary of the various operation modes is presented here.

2.3.1 Contact Mode

In contact mode, the cantilever is bent due to the atomic force acting on the probe. The cantilever is acting like a spring and the deflection of the cantilever is proportional to the interaction force [21]. In order to obtain a topography image of the sample, the cantilever is bent at a constant deflection by adjusting the separation between the probe and sample at each single spot. To study the relationship between the strength of the net force and the probe-sample separation in different environments, researchers have measured the deflection of the cantilever while changing the separation in steps [22, 23]. Although this method is simple and direct, the influence of noise and drift is very significant and it is difficult to get a repetitive experimental result. The adjustment at each spot is used to construct the sample topography. The atomic resolution can be achieved under ultra-high vacuum and low temperature condition where the noise is minimized [24].

The friction force can also be measured in contact mode. Instead of measuring the cantilever deflection, the frictional force is proportional to the twist of the cantilever. However, the cantilever twist may lead to nonlinear coupling when measuring the deflection and the twist of AFM [25]. Thus, an accurate alignment of the AFM sensing system is needed.

2.3.2 Non-contact Mode

In non-contact mode AFM, the image is obtained without the mechanical contact between probe and sample. A piezoelectric actuator is used to vibrate the cantilever holder

which in turn produces the oscillation of the cantilever [26]. The vibration amplitude and the resonant frequency of the cantilever are very sensitive to the separation between probe and sample. They can be used as reference signals to control the separation gap and construct the topography image. Since there is no contact between probe and sample, the strength of the atomic force is relatively low. However, the sample will not be scratched by the probe in this operation mode. More importantly, the noise and drift are significantly reduced in the detected signals as only signals at the vibration frequency and its harmonics are needed to calculate the material properties and surface topography.

2.3.3 Tapping Mode

In order to have a larger atomic force, the cantilever is closer to the sample and there is a contact between probe and sample during the cantilever oscillation. Compared to contact mode AFM where the probe is always in contact with the sample during the measurement, the contact time in tapping mode AFM is only a small portion of the oscillation cycle. This will reduce the wear and the tear of the probe and scratch of the sample compared to contact mode. Since the contact force is much larger in tapping mode AFM compared to non-contact mode AFM, better resolution can be achieved. Currently, this is the most popular operation mode of AFM. Besides using amplitude and frequency of the cantilever vibration to image the topography, the phase difference between the excitation and vibration signal can be used to characterize various material properties such as adhesion, elasticity and viscoelasticity [27].

2.4 Theoretical Model

In this thesis, we focus on the improvement on non-contact and tapping mode AFM. The cantilever is excited to oscillate in both modes. In order to study the dynamics of the interaction, the detailed modelling of the cantilever dynamics should be considered. The cantilever can be considered as fixed at the holder and free at the other end in AFM shown in Figure 2.3. The governing equation to describe the free vibration of the cantilever is given in [28]:

$$EI \frac{\partial^4 x(y, t)}{\partial y^4} + \rho A \frac{\partial^2 x(y, t)}{\partial t^2} = 0 \quad (2.5)$$

where $x(y, t)$ is the beam displacement, E denotes the modulus of elasticity, I denotes

Figure 2.3: The schematic of a cantilever.

the moment of inertia, ρ is the density of the cantilever, A is the cross sectional area,

the length of the cantilever is assumed to be L and the whole beam is assumed to be uniform and symmetric. Since the cantilever is fixed at one end and free at the other end, the boundary conditions must satisfy the following equations:

$$x(0, t) = 0, \frac{\partial x(0, t)}{\partial y} = 0, \frac{\partial^2 x(L, t)}{\partial y^2} = 0, \frac{\partial^3 x(L, t)}{\partial y^3} = 0 \quad (2.6)$$

The solution of the equation can be written as:

$$x(y, t) = \sum_{j=1}^{\infty} C_j \sin(\omega_j t + \delta_j) \phi_j(y) \quad (2.7)$$

where $\phi_j(y)$ is the mode shape of the cantilever and it is determined by the boundary conditions. They are orthogonal to each other. When the external atomic force and the air damping force are taken into consideration, the system equation becomes [29]:

$$EI \frac{\partial^4}{\partial y^4} [x(y, t) + a_1 \frac{\partial x(y, t)}{\partial t}] + m(y) \frac{\partial^2 x(y, t)}{\partial t^2} + a_0 m(y) \frac{\partial x(y, t)}{\partial t} = f(y, t) \quad (2.8)$$

where a_0 and a_1 are damping coefficients. It is noted that the non-uniformity of the beam is considered in this model where $m(y)$ describes the mass density distribution of the beam. All external forces are described by $f(y, t)$. Further transformations lead to decoupled differential equations:

$$\frac{d^2 Y_i(t)}{dt^2} + 2\gamma_i \omega_i \frac{dY_i(t)}{dt} + \omega_i^2 Y_i(t) = \frac{F_i(t)}{M_i} \quad (2.9)$$

where $Y_i(t) = \int_0^L x(y, t)\phi_i(y)dy$ and $F_i(t) = \int_0^L f(y, t)\phi_i(y)dy$, and M_i , ω_i , K_i and γ_i are generalized mass, frequency, stiffness and damping. Their calculations can be found in [29]. To enhance the sensitivity of the measurement, the vibration frequency of the piezoelectric actuator is selected in practice to be the first resonant frequency of the cantilever. The external force will be dominated by the probe-sample interaction force acting on the probe when the probe-sample separation is small. When the frequency of the excitation signal is set as the first resonant frequency of the cantilever, the response of the cantilever vibration will be largely dominated by its first mode. Since each vibration mode is governed by a second order ordinary differential equation (Eq. (2.9)), the dynamics of the probe-sample interaction can be simplified as [30, 31]:

$$m\ddot{x} + c\dot{x} + kx = f\cos(\omega t) + F_{interaction\ force} \quad (2.10)$$

where m , c and k are the mass, damping coefficient and the spring constant of the model, and $F_{interaction\ force}$ denotes the probe-sample interaction force.

2.5 Principle of AFM Sensing

Since the interaction force can only be detected when the probe-sample separation is less than 100 nm , the vibration amplitude of the probe is controlled to be less than 100 nm [26]. It is difficult to measure the small deflection of the cantilever directly. In practice, a laser is deployed to shine on the back of the cantilever and a PSD is mounted to detect the reflected light. The laser beam is usually generated by a laser diode with output

power of around 3 mW. Since it is focused onto the micro-cantilever, the optical energy is partially absorbed [32] by the cantilever. This heating process leads to deflection of the cantilever due to thermal expansion. This is equivalent to a nano-Newton force acting on the cantilever. This photo-thermal effect will bring measurement errors in contact mode AFM thus a calibration procedure is needed before measurements. However, the photo-thermal effect will not affect dynamic AFM measurements as the introduced optical force is static. On the other hand, photo-thermal effect can be utilized to excite cantilever vibration in liquid AFM. Modulated laser beam at the cantilever resonances is incident on the rear surface of the cantilever [33]. This method is extremely useful in liquid as it decouples the spurious resonances from fluid.

The schematic of the sensing system is shown in Figure 2.4. Vertical movement of the cantilever will change the reflected light path which in turn changes the spot position on PSD. The movement of the spot on PSD is in the range of millimeter which is an amplified version of the cantilever vibration.

The QD PSD is used in most AFM products. The QD is capable of detecting the centroid position of the light spot. It consists of four p-n junction diodes on the same substrate positioned symmetrically around the center and separated by small gaps [34]. The layout of QD PSD is shown in Figure 2.5. When the light spot strikes on diodes, an electric charge proportional to the light intensity on the whole diode is generated and collected as photocurrent. The photocurrents can be measured from the external amplification circuit. Thus, the position information of the light spot can be deduced from the magnitude of photocurrents. The position of the light spot can be calculated

Figure 2.4: The AFM sensing system.

as [35]:

$$\hat{X} = k \frac{(i_a + i_b) - (i_c + i_d)}{(i_a + i_b + i_c + i_d)} \quad (2.11)$$

$$\hat{Y} = k \frac{(i_a + i_d) - (i_b + i_c)}{(i_a + i_b + i_c + i_d)} \quad (2.12)$$

where i_a , i_b , i_c and i_d are photocurrents from each quadrant, $\hat{X}$ and $\hat{Y}$ are the estimated spot positions in the x and y directions, and k is the scale factor. The QD PSD has low inherent noise level. It is widely used in indoor applications. However, it suffers from a nonlinearity problem which is highly dependent on the shape and intensity distribution of the light spot.

Figure 2.5: The schematic of QD PSD.

2.6 Positioning Regulation in AFM

Many researchers are interested in measuring the topography of the whole sample rather than focusing on one single spot of the sample. This requires a scanning system to move the probe or sample to interact the probe over the whole sample. In addition, a feedback mechanism is needed to adjust the probe position to maintain a constant probe-sample separation during the scanning process since there are height differences in different spots of the sample. On the other hand, without the feedback mechanism, the probe may just clash with the sample somewhere or the separation is too far away. Meanwhile, the adjustment of the probe position under the constant separation condition is just the height difference among different spots which is the topographical information. Based on experience, large vibration amplitude of the probe corresponds to large separation. The probe position is adjusted by a PID controller in the feedback system whose input is the difference between the vibration amplitude and a reference value. Lead-based ferroelectric materials, especially lead zirconate titanate (PZT), have excellent piezoelectric properties where it can elongate or shrink depending on the supply voltage [36].

Chapter 3

Linearity Indices and Linearity

Improvement of 2-D Tetra-lateral

Position Sensitive Detector

3.1 Introduction

A 2-D tetra-lateral PSD is capable of providing continuous position measurement of the incident light spot with high sensitivity and resolution in two dimensions. It consists of a single square p-n junction with a resistive layer. Four electrodes are placed along each side of the square near its boundary on the same surface. When the sensor is illuminated nonuniformly, position dependent potential is generated [37] due to lateral photo effect [38] which results in current flow on the resistive layer under the fully reverse-biased condition [39]. Photocurrents can be collected and measured at each electrode to estimate the position of the incident light in two dimensions via an intuitive formula. Besides the 2-D tetra-lateral PSD, 2-D duo-lateral PSD where two electrodes opposite each other are placed on the p -surface and the other two are on the n -surface of a square

p-n junction [40]. These two 2-D sensors, together with QD PSD, have wide applications in position and angle sensing, vibration measurement, optical range finder, camera auto focusing, autonomous navigation and atomic force microscopy (AFM) [41]-[43].

Compared to the 2-D duo-lateral PSD, the 2-D tetra-lateral PSD has the advantages of fast response, much lower dark current, easy bias application and lower fabrication cost [44]. In addition, its measurement accuracy and resolution is independent of the spot shape and size unlike the QD PSD which could be easily changed by air turbulence [35]. However, it suffers from a similar nonlinearity problem as the QD PSD. Theoretically, the relationship between the position estimate and the real position of the light spot on the 2-D tetra-lateral PSD is implicitly nonlinear. While the position estimate is approximately linear with respect to the real position when the spot is in the center area of the PSD, the relationship becomes nonlinear when the light spot is away from the center and this results in severe estimation inaccuracy. This seriously degrades the measurement accuracy and resolution of the AFM. As the 2-D PSD sensor is deployed in the AFM system to measure the enlarged version of modulated atomic interaction between the probe and the sample, the measurement will only be accurate when the light spot is in the center area of the QD PSD or 2-D tetra-lateral PSD. Thus this nonlinearity problem is a bottleneck to further resolution enhancement of AFM [45] and a PSD with a larger linear range is desired to further enlarge the atomic dynamics. The unknown spot shape and light distribution in the QD PSD make linearity improvement extremely difficult and hence the 2-D tetra-lateral PSD is a potential candidate to achieve a breakthrough in linearity improvement. Attempts have been

made to modify the contact configuration and to redesign the boundary shape of the 2-D tetra-lateral PSD for linearity improvement. Pin-cushion PSD with curvature in boundary [46] and clover PSD with more sophisticated design in boundary and contacts [47] are two representative technologies. Although the linearity of modified PSDs has been improved, their complicated structures make the manufacturing process complicated and result in a much higher cost. And it must be noted that the problem of surface inhomogeneities will further introduce unknown nonlinearities [48]. Other than changing the PSD structure, one may apply advanced signal processing techniques to the original 2-D tetra-lateral PSD when estimating the spot position. This approach will remove the concern of additional cost in manufacturing and other problems associated with a complicated PSD structure. In the literature, estimation methods based on neural network and interpolation algorithms have been developed [49, 50]. However, in order to get an accurate estimate, they need a huge number of training data from the calibration process before the deployment of the PSD which is time consuming. Further, their tedious procedures and slow processing speed make them unrealistic in real implementations that require fast response like AFM. Thus, in order to achieve a better linearity performance without trade-offs in cost and other factors, a new, simple and implementable position estimation formula is needed for the 2-D tetra-lateral PSD.

In this chapter, after a brief review of 2-D tetra-lateral PSD mathematical model, two linearity indices are proposed as measures to evaluate the degree of linearity of 2-D position measurement. The measures are found to be effective when comparing the linearities of the 2-D tetra-lateral PSD and a modified tetra-lateral PSD, which is

a square PSD with L-shape electrodes [44]. Then, we present a systematic procedure for designing new estimation formulae with simple combinations of electrode currents and the coefficients can be optimized to meet any specified linearity characterized by the linearity indices. Following the procedure, a simple estimation formula is proposed for the 2-D tetra-lateral PSD with very significant linearity improvement. More importantly, the design procedure and the derived formula can be applied directly to linearity improvement of other 2-D PSDs like position sensitive silicon detector (PSSD) for heavy ion measurements [51, 52] where only the coefficients in the formula need to be re-customized before application. It is verified by simulations and an experiment that the linearity of the 2-D tetra-lateral PSD is greatly improved when the incident light beam has a negligible beam size. Last but not least, it is shown in simulations that our formulae are effective even when the light beam has a considerable beam size.

3.2 Mathematical Model

The basic structure of the 2-D tetra-lateral PSD is shown in Figure 3.1 (a) and the origin of the coordinate system is the center of the sensor. The cross section of the PSD and its physical principle are shown in Figure 3.1 (b). Lucovsky [53] has shown that the electric potential distribution in steady state is governed by the following equation when it is fully reverse-biased:

$$\nabla^2 U(x, y) = -\frac{\rho_d}{w_d} I_s(x, y) \quad (3.1)$$

Figure 3.1: (a) Basic structure of the PSD. (b) Cross section of the PSD and its physical principle.

where $U(x, y)$ is the electric potential, ∇^2 denotes the Laplace operator, $I_s(x, y)$ is the generated photocurrent which is proportional to the light illumination, ρ_d is the resistivity and w_d is the thickness of the junction. When the light beam has a negligible beam size, the generated photocurrent can be expressed as a Dirac function:

$$I_s(x, y) = I_0 \delta(x_0 - x) \delta(y_0 - y) \quad (3.2)$$

where (x_0, y_0) denotes the position of the incident light on PSD. The length of the electrode is assumed to be 1 unit in this paper for simplicity. In that case, the electric potential can be expressed explicitly when the boundary condition for 2-D tetra-lateral PSD applies [40]:

$$U(x, y) = I_0 \frac{\rho_d}{\omega_d} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \frac{\sin(mx\pi) \sin(ny\pi) \sin(mx_0\pi) \sin(ny_0\pi)}{(n^2 + m^2)\pi^2} \quad (3.3)$$

The electrical currents collected at electrodes can be calculated explicitly from the following integrals:

$$I_i = \frac{\omega_d}{\rho_d} \int_{l_i} \frac{\partial U}{\partial p} |_{p=p_i} dl_i, \quad i = 1, 2, 3, 4 \quad (3.4)$$

where p_i is the normal to each electrode and l_1 to l_4 refer to the lengths of the four electrodes [40]. The conventional position estimation formula can be found in the literature [40]. The position estimate $\hat{X}$ in the x direction and $\hat{Y}$ in the y direction will be dependent on the position of the incident light on the sensor, and they are given by:

$$\begin{aligned} \hat{X}(x_0, y_0) &= k \frac{I_4 - I_3}{I_4 + I_3} \\ \hat{Y}(x_0, y_0) &= k \frac{I_2 - I_1}{I_2 + I_1} \end{aligned} \quad (3.5)$$

where k is the linearity constant in the formula. There is no estimation error when the spot is on the PSD center, i.e.

$$\begin{aligned} \hat{X}(0, 0) &= 0 \\ \hat{Y}(0, 0) &= 0 \end{aligned} \quad (3.6)$$

3.3 Linearity Indices

It is found that when the light spot is near the PSD center, $\hat{X}$ is proportional to x_0 and $\hat{Y}$ is proportional to y_0 approximately. This approximation becomes inaccurate when

x_0 and y_0 become larger, i. e. when the spot is further away from the center. These observations can be well explained by Taylor series expansion in two dimensions, e. g. $\hat{X}$ is expanded as:

$$\begin{aligned}\hat{X}(x_0, y_0) = & [\hat{X}(0, 0) + x_0 \frac{\partial \hat{X}}{\partial x_0} \big|_{x_0=0, y_0=0} + y_0 \frac{\partial \hat{X}}{\partial y_0} \big|_{x_0=0, y_0=0}] \\ & + \frac{1}{2!} [x_0^2 \frac{\partial^2 \hat{X}}{\partial x_0^2} \big|_{x_0=0, y_0=0} + 2x_0 y_0 \frac{\partial^2 \hat{X}}{\partial x_0 \partial y_0} \big|_{x_0=0, y_0=0} \\ & + y_0^2 \frac{\partial^2 \hat{X}}{\partial y_0^2} \big|_{x_0=0, y_0=0}] + \dots\end{aligned}\tag{3.7}$$

As calculated previously, $\hat{X}(0, 0) = 0$. $\hat{X}$ depends on both x_0 and y_0 . When the light spot is near the PSD center, x_0 and y_0 are small and it can be shown that $\frac{\partial \hat{X}}{\partial y_0} \big|_{x_0=0, y_0=0} = 0$. Thus, the first term in the expression dominates and that is why it is observed that $\hat{X}$ is linear with respect to x_0 approximately. However, contributions from the nonlinear terms increase when x_0 and y_0 are larger and $\hat{X}$ is no longer linear with respect to x_0 . This will give measurement errors and one is unable to know x_0 accurately from the information of $\hat{X}$ because of the implicit relationship. The estimate will be more accurate only if the degree of linearity is enhanced. In addition, the lower order nonlinear terms in the expression, as compared to higher order terms which are neglected, usually dominate the contributions to the nonlinearity. Thus, the nonlinearity can be reduced if the lower order terms have zero or small values. In this sense, these lower order terms can be used to define linearity indices to measure the degree of linearity for different estimation schemes of 2-D PSD. The linearity indices also serve the purpose of measuring the estimation errors. Before we introduce the linearity indices, it must be

noted that the estimation formulae for all types of 2-D PSD have the property of center symmetry. This is because every 2-D PSD has a center symmetric structure, which means that estimation inaccuracy is center symmetric too. Thus, only the analysis of position estimate $\hat{X}$ will be given in details in the thesis because the same analysis also applies to $\hat{Y}$. The property of center symmetry is expressed as:

$$\begin{aligned}\hat{X}(x_0, y_0) &= \hat{X}(x_0, -y_0) \\ \hat{X}(x_0, y_0) &= -\hat{X}(-x_0, y_0)\end{aligned}\tag{3.8}$$

Or equivalently:

$$\begin{aligned}\frac{\partial^{2m+n+1} \hat{X}}{(\partial y_0)^{2m+1} (\partial x_0)^n} \Big|_{y_0=0} &= 0, \quad m, n = 0, 1, 2, \dots \\ \frac{\partial^{2m+n} \hat{X}}{(\partial x_0)^{2m} (\partial y_0)^n} \Big|_{x_0=0} &= 0, \quad m, n = 0, 1, 2, \dots\end{aligned}\tag{3.9}$$

We propose the following two linearity indices for $\hat{X}$ as measures of linearity in the x and y directions:

$$LH_x^n \triangleq \left| \frac{\frac{\partial^n \hat{X}}{\partial x_0^n} \Big|_{x_0=0, y_0=0}}{\frac{\partial \hat{X}}{\partial x_0} \Big|_{x_0=0, y_0=0}} \right| \tag{3.10}$$

$$LV_x^n \triangleq \left| \frac{\frac{\partial^n \hat{X}}{\partial y_0^{n-1} \partial x_0} \Big|_{x_0=0, y_0=0}}{\frac{\partial \hat{X}}{\partial x_0} \Big|_{x_0=0, y_0=0}} \right| \tag{3.11}$$

where n is an integer such that all $LH_x^i = 0$ for $2 \leq i \leq n-1$ and $LH_x^n \neq 0$. It should be noted that $LH_x^2 = 0$ is always satisfied when it is center symmetric. The position

measurement $\hat{X}$ satisfying this property is referred to as having n th order linearity with respect to the x direction. For example, it can be calculated that $LH_x^2 = 0$ while $LH_x^3 \neq 0$ for the 2-D tetra-lateral PSD with the conventional formula and so the position measurement $\hat{X}$ is referred to as having third order linearity with respect to the x direction. Similarly, the n th order of linearity with respect to the y direction is defined as LV_x^n where n is an integer such that all $LV_x^i = 0$ for $2 \leq i \leq n - 1$ and $LV_x^n \neq 0$. Higher order linearity indices indicate better linearity since more lower order terms have zero values. When linearity indices have the same order, linearity indices with smaller values would indicate better linearity. In the same way, linearity indices for the position measurement $\hat{Y}$ can be defined as well. The linearity indices for the 2-D Tetra-lateral PSD with the conventional formula can be calculated as:

$$LH_x^3 = \left| \frac{\frac{\partial^3 \hat{X}(x_0, y_0)}{\partial x_0^3}}{\frac{\partial \hat{X}(x_0, y_0)}{\partial x_0}} \right|_{x_0=0, y_0=0} = LV_y^3 = \left| \frac{\frac{\partial^3 \hat{Y}(x_0, y_0)}{\partial y_0^3}}{\frac{\partial \hat{Y}(x_0, y_0)}{\partial y_0}} \right|_{x_0=0, y_0=0} = 19.5 \quad (3.12)$$

$$LV_x^3 = \left| \frac{\frac{\partial^3 \hat{X}(x_0, y_0)}{\partial y_0^2 \partial x_0}}{\frac{\partial \hat{X}(x_0, y_0)}{\partial x_0}} \right|_{x_0=0, y_0=0} = LH_y^3 = \left| \frac{\frac{\partial^3 \hat{Y}(x_0, y_0)}{\partial x_0^2 \partial y_0}}{\frac{\partial \hat{Y}(x_0, y_0)}{\partial y_0}} \right|_{x_0=0, y_0=0} = 1.9 \quad (3.13)$$

Since the linearity indices defined in Eq. (3.10) and Eq. (3.11) are universal for all 2-D PSD measurements, the linearity indices for the modified tetra-lateral PSD with

L-shape electrodes can also be calculated and they are:

$$LH_x^3 = \left| \frac{\frac{\partial^3 \hat{X}(x_0, y_0)}{\partial x_0^3}}{\frac{\partial \hat{X}(x_0, y_0)}{\partial x_0}} \right|_{x_0=0, y_0=0} = LV_y^3 = \left| \frac{\frac{\partial^3 \hat{Y}(x_0, y_0)}{\partial y_0^3}}{\frac{\partial \hat{Y}(x_0, y_0)}{\partial y_0}} \right|_{x_0=0, y_0=0} = 4.3 \quad (3.14)$$

$$LV_x^3 = \left| \frac{\frac{\partial^3 \hat{X}(x_0, y_0)}{\partial y_0^2 \partial x_0}}{\frac{\partial \hat{X}(x_0, y_0)}{\partial x_0}} \right|_{x_0=0, y_0=0} = LH_y^3 = \left| \frac{\frac{\partial^3 \hat{Y}(x_0, y_0)}{\partial x_0^2 \partial y_0}}{\frac{\partial \hat{Y}(x_0, y_0)}{\partial y_0}} \right|_{x_0=0, y_0=0} = 6.7 \quad (3.15)$$

The quantitative calculation shows that the position estimate $\hat{X}$ for the 2-D tetra-lateral PSD has a poorer linearity than its modified version in the x direction but better in the y direction. This conclusion is supported by the simulation result. In the simulation, the light spot is moved from the center in steps of 0.05 in both directions inside the central square of (0.8×0.8) . Then position estimates are plotted for both 2-D tetra-lateral PSD and its modified counterpart in Figure 3.2 and Figure 3.3 respectively. If the estimation is perfect without inaccuracy, they will form a regular grid pattern. In the simulation, when the light spot moves in steps along the x direction from the center, the distance between adjacent position estimates will be nearly equal if $\hat{X}$ has small nonlinearity in the x direction. From Figure 3.2 and Figure 3.3, it can be seen that the distance between adjacent position estimates for the 2-D tetra-lateral PSD varies more significantly than that for the modified 2-D tetra-lateral PSD which means a poorer linearity for $\hat{X}$ in the x direction. When the light spot moves in steps along the y direction for a fixed x coordinate from top to bottom, the position estimates should be well aligned vertically if $\hat{X}$ has small nonlinearity in the y direction. From Figure 3.2 and Figure 3.3, it can be seen that the alignment is poorer in the modified 2-D tetra-lateral PSD which verifies that its position estimate $\hat{X}$ has a larger nonlinearity in the

Figure 3.2: Simulation result of 2-D tetra-lateral PSD using conventional formula.

y direction. These phenomena are well explained by our linearity indices and hence our proposal of these linearity indices as measures of the linearity of 2-D PSD is justified.

3.4 New Estimation Formulae

The difficulty of achieving an accurate estimation lies in the implicit nonlinear relationship between the position estimate and the real incident position. It is impossible to recover exactly the real incident position based on the information of the four current outputs from the electrodes regardless of any tetra-lateral electrodes and boundary design. The only possibility is to improve the estimation accuracy by improving the degree of linearity. While creative designs of the electrodes and boundary [44] have improved

Figure 3.3: Simulation result of the modified 2-D tetra-lateral PSD.

the linearity, further improvement is rather difficult and the technique cannot be applied to other 2-D PSD like PSSD for different applications. In this section, through the derivation of a new formula for 2-D tetra-lateral PSD, a systematic procedure is presented to design estimation formulae to meet any degree of linearity for 2-D PSD. Due to the property of center symmetry, only the derivation for $\hat{X}$ is given and that for $\hat{Y}$ can be obtained in the same manner. The desirable estimation formula which is based on the four output currents, should have a simple arithmetic for easy implementation. We first examine a formula that involves the ratio of two linear combinations of the four currents:

$$\hat{X}(x_0, y_0) = \frac{\sum_{i=1}^4 a_i I_i}{\sum_{i=1}^4 b_i I_i}. \quad (3.16)$$

The coefficients are to be optimized for linearity improvement. First of all, the property of center symmetry must be satisfied, $\frac{\partial \hat{X}}{\partial x_0}$ must be non-zero to ensure linearity and $\hat{X}$ should not be infinity when the light spot is on PSD. To fulfil all these pre-conditions, we have:

$$\hat{X}(x_0, y_0) = \frac{a_4(I_4 - I_3)}{b_1(I_1 + I_2) + b_4(I_3 + I_4)}. \quad (3.17)$$

There is still flexibility for further linearity improvement. Guided by the proposed linearity indices, since $\frac{\partial^3 \hat{X}(x_0, y_0)}{\partial x_0^3}|_{x_0=0, y_0=0} \neq 0$ and $\frac{\partial^4 \hat{X}(x_0, y_0)}{\partial x_0^4}|_{x_0=0, y_0=0} = 0$ in general, we can achieve fifth order of linearity with respect to the x direction by choosing coefficients to satisfy:

$$\frac{\partial^3 \hat{X}(x_0, y_0)}{\partial x_0^3}|_{x_0=0, y_0=0} = 0 \quad (3.18)$$

which requires $b_4 = 1.72b_1$. This relationship is calculated based on the photocurrents and their derivatives which are calculated from Eq. (3.3) and Eq. (3.4). At this stage, there is no more flexibility for further linearity improvement because the coefficients a_4 and b_1 only relate to the linearity slope and not the nonlinearity. With the above $\hat{X}(x_0, y_0)$, the linearity index for $\hat{X}$ with respect to the y direction can be calculated as $|\frac{\frac{\partial^3 \hat{X}(x_0, y_0)}{\partial x_0 \partial y_0^2}|_{x_0=0, y_0=0}}{\frac{\partial \hat{X}(x_0, y_0)}{\partial x_0}|_{x_0=0, y_0=0}}| = 5.5$ which is poorer than the linearity index calculated in Eq. (3.13) for $\hat{X}$ with respect to the y direction when the conventional formula is used. Thus, a new estimation formula with more flexibility terms is needed for linearity improvement in $\hat{X}$ with respect to both directions.

In general, any degree of linearity improvement for 2-D PSD is achievable by optimizing the coefficients of a formula that has enough flexibility terms to set to zero the corresponding nonlinearity terms. But this will give rise to a complicated formula that is difficult to implement. A trade-off between linearity improvement and complexity of the formula is required while satisfying the pre-conditions. Our aim is to achieve fifth order linearity with respect to both x and y directions for position measurement $\hat{X}$. For this purpose, we made use of the following formula:

$$\hat{X}(x_0, y_0) = \frac{\sum_{i=1}^4 a_i I_i \sum_{i=1}^4 c_i I_i}{\sum_{i=1}^4 b_i I_i \sum_{i=1}^4 d_i I_i}. \quad (3.19)$$

With optimized coefficients, the following estimation formulae have been found to satisfy the requirement:

$$\begin{aligned} \hat{X}(x_0, y_0) &= \frac{k(I_4 - I_3)}{I_{sum} - 1.02(I_2 - I_1)} \times \frac{0.7(I_1 + I_2) + I_{sum}}{I_{sum} + 1.02(I_2 - I_1)} \\ \hat{Y}(x_0, y_0) &= \frac{k(I_2 - I_1)}{I_{sum} - 1.02(I_4 - I_3)} \times \frac{0.7(I_3 + I_4) + I_{sum}}{I_{sum} + 1.02(I_4 - I_3)} \end{aligned} \quad (3.20)$$

where $I_{sum} = \sum_{i=1}^4 I_i$

3.5 Simulation and Experimental Results

To verify the performance of the new formulae, simulations and an experiment are conducted. In the simulation, the light spot is assumed to move in steps, following the same procedure as that in the previous simulation. The photocurrents are calculated based on Eq. (3.3) and Eq. (3.4). The photocurrents are then used to calculate the

position estimates via our new formulae. The simulation result is shown in Figure 3.4. For a quantitative comparison, the percentage of the effective area with measurement errors of less than 1%, 2%, 3% and 4% are tabulated in Table 3.1. For a fair comparison, the effective area defined here is the same as Wang's definition [44], bounded by a central square with a side length of $\frac{10}{12}$. Measurement errors are defined in [54]: $D = \frac{x_0 - \hat{X}(x_0, y_0)/k_x}{x_0}$ where k_x is the sensitivity defined as $k_x = \partial \hat{X}(x_0, y_0)/\partial x_0|_{x_0=0, y_0=0}$ and $[x_0, y_0]$ is the real position of the light spot in both x and y direction. Comparisons are summarized in the table where some data are from Wang's calculation [44]. It can be seen that the percentage of the area with less than 1% measurement error, considered as error free area, has been greatly enlarged to 32% as compared to 6% based on the conventional formula. This is quite significant and it even has a better performance than clover PSD. We note that in various PSD research areas and industry applications, sensitivity, degree of nonlinearity and maximum measurement errors are three scientific criteria frequently used to judge the PSD performance. Thus, we also compare our new formulae with conventional formula and other improved PSD designs in terms of maximum measurement error within the effective area and sensitivity (Table 3.2). Here, the sensitivity k_x has the same definition as above. It can be seen that the sensitivity has been reduced when our new formulae are used compared to the conventional formulae. This is the trade off made in order to improve the linearity. However, the sensitivity of our formulae is still much higher than any other modified schemes. In addition, the maximum measurement error is significantly reduced by using our new formulae.

Next, a comparison of parameter optimization methods between our linearity indice

Figure 3.4: Estimation result of 2-D tetra-lateral PSD using proposed new formula.

approach and conventional parameter estimation techniques is presented. Conventionally, parameters are optimized to minimize a cost function for different purposes. Since the purpose here is to improvement the measurement linearity and reduce measurement errors, a quadratic cost function is constructed and will be minimized:

$$f = \sum_{i=1}^N [\hat{X}(x_i, y_i) - x_i]^2. \quad (3.21)$$

The data points (x_i, y_i) are chosen to be $(0.05m, 0.05n)$ inside the central square of (0.8×0.8) where m and n are integers. We assume $\hat{X}$ has the same formulae structure as Eq. (3.19) with 16 parameters to be determined. The Levenberg-Marquardt

Table 3.1: Distortion Comparison Among Different Linearity Improvement Schemes

	Percentage of area with distortion less than			
	1.0%	2.0%	3.0%	4.0%
Tetra-lateral	6.0	10.4	13.1	16.2
Pin-cushion	6.9	13.5	20.8	31.6
Clover	17.1	52.7	67.6	76.4
Tetra-lateral (new formula)	31.5	56.4	71.6	84.1

Table 3.2: Comparison of Maximum Measurement Errors and Sensitivity Among Different Linearity Improvement Schemes

	$D_{max}(\%)$	k_x
Tetra-lateral	33	3.3
Pin-cushion	8.4	0.2
Clover	6.0	1.1
Tetra-lateral (new formula)	10.1	2.3

algorithm is be used to optimize these parameters by finding a local minimum of the cost function. Since there is no prior knowledge on the initial guess of the parameters, we assume $a_i = b_i = c_i = d_i = 1$ where $i = 1, 2, 3, 4$ as the initial guess. The optimization process is done in Matlab by calling its function "lsqnonlin" which uses Levenberg-Marquardt algorithm by default. The optimized parameters are obtained as: $a_1 = 0.9401, a_2 = 0.9401, a_3 = 0.9894, a_4 = 0.7032, b_1 = 1.0607, b_2 = 1.0607, b_3 = 1.0107, b_4 = 1.3009, c_1 = 0.9401, c_2 = 0.9401, c_3 = 0.9894, c_4 = 0.7032, d_1 = 1.0607, d_2 = 1.0607, d_3 = 1.0107, d_4 = 1.3009$. The estimate of the light spot position calculated via Eq. (3.19) with these optimized parameters are plotted in Figure 3.5 when the light spot is moved from the center in steps of 0.05 in both directions inside the central square of (0.8×0.8) . It can be seen that the measurement linearity is even worse than the conventional formula (Eq. (3.5)). The reason why conventional parameter estimation techniques like Newton's method and Levenberg-Marquardt method are not suitable for our problem is briefly discussed. In the new formulae (Eq. (3.19)), there are 16 unknown

Figure 3.5: Estimation result of 2-D tetra-lateral PSD using Levenberg-Marquardt algorithm.

coefficients nonlinearly coupled with each other. Our derivative based approach provides an intelligent way to optimize the coefficients for the purpose of linearity improvement. Although it is possible to construct and minimize a cost function in least square sense, the problem is that any conventional searching algorithms like Newton's method and Levenberg-Marquardt method will get stuck in one local minimum of the cost function near the initial guess of the unknown parameters. On the other hand, there is no prior knowledge on how to choose a good initial guess in order to improve the measurement linearity. Thus, the performance of new formulae with optimized parameters via these methods is worse than the formulae obtained based on minimizing linearity indices.

In practice, noise exists when measuring the electrical currents from the four elec-

trodes. Thus, it is important to test the robustness of our proposed formulae subject to measurement noise. Simulations have been done to do such a test. In our simulations, An unbiased Gaussian noise is added to each of the electrical currents measured. We have compared and tabulated the percentage of the effective area with less than 1%, 2%, 3%, and 4% of measurement errors accordingly (Table 3.3 and Table 3.4) between the conventional formulae and our new formulae. Although the complexity of our new formulae is increased, the robustness is not degraded compared to the conventional formulae. In many applications, the estimation formulae are implemented in electrical circuits where the multiplication parameters are implemented by resistors. The resistances of these resistors will become more and more inaccurate after use. Thus, there is a need for our new formulae to be robust against the parameter variations. Another set of simulations have been done for this purpose. The multiplication parameters of our formulae are varied in the range of 2%, 5%, and 10% in each simulation, the percentage of the effective area with distortion of less than 1%, 2%, 3%, and 4% is computed respectively (Table 3.5). It can be seen that our new formulae are robust to parameter variations.

Table 3.3: The Performance of Conventional Formulae and Our New Formulae on Tetra-lateral PSD in the Presence of Gaussian Noise (SNR = 40 dB)

	Percentage of area with distortion less than			
	1.0%	2.0%	3.0%	4.0%
Conventional formulae	4.3±0.5	8.6±0.7	12.9±0.7	17.2±0.8
Our new formulae	17.9±0.9	34.9±1.1	50±1.0	62.7±1.0

In addition to simulations, an experiment was performed to test the new formula in the real environment. The experimental setup consists of a 2-D tetra-lateral PSD

Table 3.4: The Performance of Conventional Formulae and Our New Formulae on Tetra-lateral PSD in the Presence of Gaussian Noise (SNR = 30 dB)

	Percentage of area with distortion less than			
	1.0%	2.0%	3.0%	4.0%
Conventional formulae	2.6 $\pm$ 0.4	5.1 $\pm$ 0.5	7.7 $\pm$ 0.6	10.3 $\pm$ 0.7
Our new formulae	8.1 $\pm$ 0.7	16.1 $\pm$ 0.9	23.8 $\pm$ 1.0	31.2 $\pm$ 1.1

Table 3.5: Robustness Analysis of Our New Formulae with Respect to Multiplication Parameter Variations

	Percentage of area with distortion less than			
	1.0%	2.0%	3.0%	4.0%
2% variation of parameters	30.1%	47.4%	61.9%	84.0%
5% variation of parameters	25.3%	39.1%	52.9%	62.6%
10% variation of parameters	20.8%	30.8%	42.5%	55.7%

(S1200 from Hamamatsu Photonics), a He-Ne laser tube with adjustable intensity, and a motorized X-Y-Z stage. The active area of PSD is 13mm $\times$ 13mm. The motorized stage has a motion accuracy of 50nm. A special lens system is designed to reduce the spot diameter to 100 μ m in front of the laser. The digital image of the experiment setup is shown in Figure 3.6 (a). Before the experiment, the PSD is fixed on a target plane holder which is well aligned with a tilt alignment system. The laser becomes stable after it is turned on for a while. The whole experiment is carried out in dark environment. During the experiment, the center of the PSD is first sought by locating zero value in the position readings. Then, electrode currents are recorded when the laser spot is scanning the PSD line by line with a step size of 0.5mm in both directions. The conventional formula and our new formula are both implemented in electrical circuits with analog devices like operational amplifier for addition and subtraction and analog computational units (AD534) for multiplication and division. The schematic of electrical circuit to implement our new formula is shown in Figure 3.6 (b). The electrical circuit

Figure 3.6: (a) Digital image of the experiment setup. (b) The schematic of electrical circuit to implement our new formula.

to implement our new formula has 6 AD534 units which is slightly more complicated than the 4 AD 534 units deployed in the conventional formula. So the complexity and the additional cost required to achieve linearity improvement are slightly increased. For simplicity, only one quadrant of the PSD is tested as it is center symmetric. Estimates processed via the conventional and new formulae are plotted in Figure 3.7 and Figure 3.8 respectively. To be more quantitative, the composite measurement errors defined as $D = \sqrt{(\hat{X} - k_x x_0)^2 + (\hat{Y} - k_y y_0)^2}$ are computed at each step as well (Table 3.6 and Table 3.7). The result clearly shows that a good linearity is achieved in the real environment by using our new formula.

Figure 3.7: Experiment result for position measurements by using 2-D tetra-lateral PSD with conventional formula.

Figure 3.8: Experiment result for position measurements by using 2-D tetra-lateral PSD with new formula.

3.6 The Effect of Spot Size on Our Proposed Formulae

In the previous sections, all the analyse and simulations on linearity performance of 2-D Tetra-lateral PSD are based on the assumption that the beam projected on the

Table 3.6: Composite measurement errors (mm) in the experiment using conventional formulae

	$x_0 = 0$	$x_0 = 0.5$	$x_0 = 1$	$x_0 = 1.5$	$x_0 = 2$	$x_0 = 2.5$	$x_0 = 3$	$x_0 = 3.5$	$x_0 = 4$
$y_0 = 0$	0.05	0.04	0.04	0.06	0.12	0.21	0.31	0.44	0.60
$y_0 = 0.5$	0.04	0.01	0.03	0.05	0.12	0.20	0.30	0.43	0.59
$y_0 = 1$	0.05	0.02	0.01	0.04	0.10	0.19	0.29	0.42	0.58
$y_0 = 1.5$	0.09	0.06	0.04	0.03	0.09	0.18	0.27	0.41	0.56
$y_0 = 2$	0.15	0.12	0.10	0.08	0.10	0.17	0.26	0.39	0.55
$y_0 = 2.5$	0.25	0.21	0.19	0.16	0.16	0.19	0.27	0.40	0.55
$y_0 = 3$	0.37	0.33	0.30	0.28	0.26	0.27	0.32	0.42	0.57
$y_0 = 3.5$	0.52	0.49	0.46	0.43	0.41	0.40	0.43	0.50	0.62
$y_0 = 4$	0.71	0.67	0.64	0.61	0.58	0.57	0.57	0.63	0.72

Table 3.7: Composite measurement errors (mm) in the experiment using our new formulae

	$x_0 = 0$	$x_0 = 0.5$	$x_0 = 1$	$x_0 = 1.5$	$x_0 = 2$	$x_0 = 2.5$	$x_0 = 3$	$x_0 = 3.5$	$x_0 = 4$
$y_0 = 0$	0.05	0.02	0.01	0.04	0.03	0.02	0.02	0.05	0.08
$y_0 = 0.5$	0.03	0.01	0.04	0.05	0.05	0.04	0.04	0.03	0.06
$y_0 = 1$	0.03	0.02	0.04	0.06	0.06	0.05	0.05	0.04	0.05
$y_0 = 1.5$	0.05	0.05	0.05	0.06	0.06	0.06	0.06	0.06	0.06
$y_0 = 2$	0.07	0.05	0.05	0.05	0.06	0.05	0.06	0.05	0.06
$y_0 = 2.5$	0.08	0.06	0.04	0.05	0.05	0.05	0.05	0.04	0.05
$y_0 = 3$	0.12	0.09	0.07	0.06	0.04	0.04	0.03	0.01	0.03
$y_0 = 3.5$	0.17	0.14	0.11	0.09	0.07	0.05	0.04	0.02	0.04
$y_0 = 4$	0.22	0.19	0.16	0.13	0.11	0.09	0.07	0.06	0.07

sensor has a negligible spot size. Although the simulation process is simplified under this assumption, there is no criterion on when the light spot size can be negligible. In practice, the light beam always has a certain spot size. For example, the light spot size on PSD in AFM systems is around $40\mu m$ [55]. In addition, it is difficult to measure and control the beam size in PSD applications [56]. Thus, there is a strong need to analyze the effect of the beam size on our linearity improvement method for many applications when the light beam has a considerable beam size. The difficulty to analyze the spot size effect on our linearity improvement scheme lies in the complexity of the numerical

calculations of the electrical potential and collected photocurrents.

To address the concern of large spot size effect on our linearity improvement scheme, a new theoretical analysis is presented to simplify the numerical calculation of four photocurrents when the intensity of the light spot is Gaussian distributed. Then, simulations are carried out to estimate the light beam center when the light beam has different beam sizes. Simulation results show that our linearity improvement approach is still effective when the beam size is considered.

3.6.1 Theoretical Analysis

Although the linearity and the measurement accuracy of the 2-D tetra-lateral PSD has been greatly improved by our new formula when the light spot has a negligible beam size, there is a great need to investigate the effect of large beam size on the linearity improvement. In this section, theoretical analysis of the electrical potential and generated photocurrents will be discussed when a light spot with variable beam size is on the PSD. In most applications, a single mode laser is used as the light source where the generated photocurrent on PSD is Gaussian distributed:

$$I_s(x, y) = I_0 \exp\left(\frac{-((x - x_0)^2 + (y - y_0)^2)}{w^2}\right) \quad (3.22)$$

where I_0 is the current density at the spot center (x_0, y_0) and w is the spot radius. In our setting, the PSD center is the origin of the coordinate $(0, 0)$ and the side length of the PSD is assumed to be 1 unit. An electrode at each side imposed a boundary condition for Eq. (3.1) where the electrical potential is zero at electrodes. The solution

to Eq. (3.1) is obtained based on existing techniques to solve the non-homogeneous partial differential equation in [57]:

$$U(x, y) = \frac{\rho_d}{w_d} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \left[\int_{D_1} \int \left(\frac{\sin(m\pi x) \sin(n\pi y) \sin(m\pi x_1) \sin(n\pi y_1)}{(n^2 + m^2)\pi^2} \right. \right. \\ \left. \left. \times I_s(x_1, y_1) \right) dx_1 dy_1 \right] \quad (3.23)$$

where D_1 is the effective area of the PSD. In order to proceed with the calculation, polar coordinate is used instead of Cartesian coordinate. In most applications, the beam size w is much less than the side length of the PSD so the electrical potential can be calculated as:

$$U(x, y) = \frac{I_0 \rho_d}{w_d} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \left[\int_{r=0}^{+\infty} \int_{\theta=0}^{2\pi} \left(\frac{\sin(m\pi x) \sin(n\pi y)}{(n^2 + m^2)\pi^2} \right. \right. \\ \left. \left. \times \sin(m\pi(x_0 + r \cos \theta)) \sin(n\pi(y_0 + r \sin \theta)) r \exp\left(\frac{-r^2}{w^2}\right) d\theta dr \right] \quad (3.24)$$

which could be simplified as:

$$U(x, y) = \frac{2\pi I_0 \rho_d}{w_d} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \left[\int_{r=0}^{+\infty} \frac{\sin(m\pi x) \sin(n\pi y) \sin(m\pi x_0) \sin(n\pi y_0)}{(n^2 + m^2)\pi^2} \right. \\ \left. \times r \exp\left(\frac{-r^2}{w^2}\right) (J_0(m\pi r) J_0(n\pi r) + 2 \sum_{t=1}^{+\infty} (-1)^t (J_{2t}(m\pi r) J_{2t}(n\pi r)) dr \right] \quad (3.25)$$

where J refers to Bessel function of the first kind. Before we calculate the integral, several equations related to the Bessel functions are introduced here [58]:

$$\int_0^{+\infty} r \exp\left(\frac{-r^2}{w^2}\right) J_{2t}(m\pi r) J_{2t}(n\pi r) dr = \frac{w^2 \exp\left(\frac{-(m^2+n^2)\pi^2 w^2}{4}\right) I_{2t}\left(\frac{mn\pi^2 w^2}{2}\right)}{2} \quad (3.26)$$

$$I_0\left(\frac{mn\pi^2 w^2}{2}\right) + 2 \sum_{t=1}^{\infty} (-1)^t I_{2t}\left(\frac{mn\pi^2 w^2}{2}\right) = 1 \quad (3.27)$$

where I refers to modified Bessel function of the first kind. The calculation of electrical potential can be simplified by solving the integral:

$$U(x, y) = \frac{2\pi I_0 \rho_d}{w_d} \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \left(\frac{\sin(m\pi x) \sin(n\pi y) \sin(m\pi x_0) \sin(n\pi y_0)}{(n^2 + m^2)\pi^2} \right. \\ \left. \times \exp\left(\frac{-(m^2 + n^2)\pi^2 w^2}{4}\right) \frac{w^2}{2} \right). \quad (3.28)$$

And the four photocurrents can be calculated as follows:

$$I_1 = I_0 \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \left[\frac{\sin(m\pi x) \sin(n\pi y) \sin(m\pi x_0) \sin(n\pi y_0) n}{m(n^2 + m^2)\pi^2 b_{m,n}} \right. \\ \left. \times (1 - \cos(m\pi)) \right] \quad (3.29)$$

$$I_2 = I_0 \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \left[\frac{\sin(m\pi x) \sin(n\pi y) \sin(m\pi x_0) \sin(n\pi y_0) n}{m(n^2 + m^2)\pi^2 (-1)^{n+1} b_{m,n}} \right. \\ \left. \times (1 - \cos(m\pi)) \right] \quad (3.30)$$

$$I_3 = I_0 \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \left[\frac{\sin(m\pi x) \sin(n\pi y) \sin(m\pi x_0) \sin(n\pi y_0) m}{n(n^2 + m^2)\pi^2 b_{m,n}} \right. \\ \left. \times (1 - \cos(n\pi)) \right] \quad (3.31)$$

$$I_4 = I_0 \sum_{m=-\infty}^{+\infty} \sum_{n=-\infty}^{+\infty} \left[\frac{\sin(m\pi x) \sin(n\pi y) \sin(m\pi x_0) \sin(n\pi y_0) m}{n(n^2 + m^2)\pi^2 (-1)^{m+1} b_{m,n}} \right. \\ \left. \times (1 - \cos(n\pi)) \right] \quad (3.32)$$

where $b_{m,n} = \exp(\frac{(m^2+n^2)\pi^2 w^2}{4})\frac{2}{w^2}$. With specified spot size w , we are able to compute the electrical potential, collected photocurrents and estimates of the light spot center from different estimation formulae efficiently.

3.6.2 Simulation Results

In this section, we will present the simulation results to evaluate the linearity performance of our new formula when the light spots have different spot sizes. In the first simulation, a light spot with a beam size $w = 0.01$ is moved from the center in steps of 0.05 in both directions inside the central square of (0.8×0.8) . The electrical potential and photocurrents are calculated based on the analysis in Woltring's work [40]. Then position estimates are plotted for the 2-D tetra-lateral PSD by using our new formula in Figure 3.9. If the estimation is perfect without inaccuracy, they will form a regular grid pattern. Next, we simulate the same light spot movement on PSD when the spot size is $w = 0.05$. The estimates of the beam center calculated by our new formula are shown Figure 3.10. It is clear that a significant linearity improvement of the 2-D tetra-lateral PSD is still achieved by our formula when the beam size is enlarged. In some PSD applications, the light spot may have an extremely large beam size. To verify whether the linearity improvement of the 2-D tetra-lateral PSD is affected under this extreme scenario, we simulate a light beam with a large beam size ($w = 0.2$) moving from the center in steps of 0.05 in both directions inside the central square of (0.4×0.4) . The area of the movement is reduced because a considerable amount of the light will not be projected on the PSD if the light beam is outside the central square of (0.4×0.4)

Figure 3.9: Estimation result of 2-D tetra-lateral PSD using our new formula when $w = 0.01$.

which will introduce significant errors in the position estimates. The simulation result shows that the linearity improvement is not affected under this scenario. Thus, our linearity improvement scheme for the 2-D tetra-lateral PSD is robust to the beam size and it can significantly improve the measurement accuracy of the 2-D tetra-lateral PSD applications where the spot has a considerable beam size.

3.7 Conclusions

In this chapter, we have introduced two linearity indices as measures of linearity, or equivalently, estimation accuracy for 2-D PSDs. They are very useful for explaining the difference in estimation accuracy of the 2-D tetra-lateral PSD and its modified version.

Figure 3.10: Estimation result of 2-D tetra-lateral PSD using our new formula when $w = 0.05$.

More importantly, a systematic procedure is proposed to design new estimation formulae for the 2-D tetra-lateral PSD to achieve any degree of linearity characterized by the linearity indices. Following the procedure, a new formula with a simple structure has been derived to achieve fifth order linearity with respect to both x and y directions for $\hat{X}$ and $\hat{Y}$. An experiment was also performed to illustrate very good linearity improvement in the real environment, making it much more superior than any other competitors. In order to further improve the linearity over an even larger range, more low order terms need to be eliminated in sequence. The new formulae structure of Eq. (3.19) is not suitable for this task anymore because there are not enough adjustable parameters to set to zero nonlinear terms simultaneously. More terms with adjustable parameters must

Figure 3.11: Estimation result of 2-D tetra-lateral PSD using our new formula when $w = 0.2$.

be added to Eq. (3.19). Since the accuracy and the resolution of the AFM are limited by the estimation accuracy of the PSD, with our new formula, the 2-D tetra-lateral PSD will offer an effective solution to improve the accuracy and the resolution of the AFM system. More significantly, the design procedure can be applied directly to linearity improvement of other 2-D PSDs where only the coefficients need to be re-customized. Furthermore, our approach is found to be applicable to the more practical case where the light beam has large spot size.

Chapter 4

Improved Accuracy of 2-D

Non-contact Range Sensors Using

2-D Tetra-lateral PSD

4.1 Introduction

In principle, the AFM sensing system is same as a triangulation based 2-D non-contact range sensor. The triangulation based non-contact range sensors are capable of providing non-invasive measurements of surface profile, object position and velocity [59]. Compared to other techniques, the triangulation based non-contact range sensor has the advantage of simplicity and robustness [60] and they are widely used in the area of alignment and robot navigation. However, the nonlinearity problem also exists in triangulation principle [61]. In many triangulation based optical non-contact measurements, the position measured from the sensor is assumed to have a linear relationship with the distance between the object and the sensor. This is true if the object is far away from the sensor and the movement of the object is in a small range. However, when the object

is near to the sensor or the movement of the object is in a big range, the geometric non-linearity appears obvious and it will degrade the accuracy of the measurement. In that case, the accuracy of the AFM sensing system is affected by the measurement errors brought by both the PSD position estimation and the triangulation principle. Thus, the spot movement on PSD is not purely proportional to the deflection of the cantilever. In order to improve the total accuracy of the AFM sensing system, inaccuracies from these two sources should be considered together. In this chapter, another set of formula is proposed to compensate for the total measurement errors in the cantilever deflection. It is shown in simulations that the accuracy of the whole sensing system is improved greatly.

4.2 2-D Non-contact Range Sensor Based on Triangulation Principle

As mentioned previously, the cantilever deflection is in the nanometer range. The most sensitive method to measure the cantilever deflection is optical detection method (Figure 4.1). The cantilever will be bent if there is any force acting on the probe. The bent cantilever will alter the path of reflected light which could be detected by 2-D PSD [62]. In principle, the AFM sensing mechanism is the same as 2-D displacement sensor based on triangulation principle.

An optical non-contact 2-D displacement sensor which is based on the principle of triangulation consists of a light source, a PSD, and some other accessory optics as

Figure 4.1: The AFM sensing principle.

illustrated in Figure 4.2. Using the sensor, the object pasted with a small piece of reflector is illuminated by the light source. The reflected light is collected by the 2-D PSD through the collimation lens. A reference point denoted by o is first calibrated by observing the light spot reflected from o at the center of the PSD. Then the two dimensional movement of the object causes the light spot's deflection on the PSD, and the 2-D displacements of the object in the plane perpendicular to the optical axis of the light source are calculated from the deflection measurements based on the triangulation principle. Here, the x axis is assumed to be normal to the reflection surface. y is the distance between the x axis and reflecting point. In case of AFM sensing mechanism, the cantilever is the reflector. The bent of the cantilever will change x and y accordingly. Besides, the reflector will be slanted with respect to $x - y$ plane when the cantilever is bent. But the slanted reflector will not alter the light spot position on 2-D PSD compared to the scenario where the reflector is not slanted. The twist of the cantilever caused by the lateral force like friction will also change x and y . Thus, the AFM sensing mechanism is an example of this 2-D non-contact range sensors. In this chapter, we will focus on the general model of a 2-D non-contact range sensor (Figure 4.2) and how x and y relate to the incident light position on 2-D PSD.

From the geometry, the relative 2-D displacements of the object with respect to the reference point can be related to the light spot deflection on the PSD by:

Figure 4.2: Side View And Top View of 2-D Range Sensor.

$$\frac{x}{d} = \frac{u}{\cos^2 \theta (F \tan \theta - u)} \quad (4.1)$$

$$\frac{y}{d} = \frac{v(1 + \frac{x \cos^2 \theta}{d})}{F \cos \theta} \quad (4.2)$$

where F denotes the focal length, θ is the observation angle of the reference point, u and v represent the light spot position on the PSD, and x and y represent the 2-D displacement of the object from the reference point. The light spot deflections u and v are measured by a 2-D tetra-lateral PSD.

4.3 Proposal of New Formula and Simulation Study

The new set of formulae to compensate the total inaccuracy of 2-D non-contact range sensors has been derived based on our linearity indices:

$$\begin{aligned}\hat{X} &= \frac{d(I_4 - I_3)}{2.25F \sin \theta \cos \theta (I_1 + I_2 + I_3 + I_4 - 1.02(I_2 - I_1))} \\ &\times \frac{(I_4 + I_3 + 1.7(I_1 + I_2) + \frac{0.8(I_4 - I_3)}{F \tan \theta})}{(I_1 + I_2 + I_3 + I_4 + 1.02(I_2 - I_1))}\end{aligned}\quad (4.3)$$

$$\begin{aligned}\hat{Y} &= \frac{d(I_2 - I_1)}{2.25F \cos \theta (I_1 + I_2 + I_3 + I_4 - 1.02(I_4 - I_3))} \\ &\times \frac{(I_1 + I_2 + 1.7(I_4 + I_3) + \frac{0.6(I_4 - I_3)}{F \tan \theta})}{(I_1 + I_2 + I_3 + I_4 + 1.02(I_4 - I_3))}.\end{aligned}\quad (4.4)$$

It can be verified that $\hat{X}$, the estimate of the object movement in the x direction, has achieved *3rd* order linearity in the x direction and *5th* order linearity in the y direction respectively. And $\hat{Y}$, the estimate of the object movement in the y direction, has achieved *2nd* order linearity in the x direction and *6th* order linearity in the y direction respectively.

Simulations have been conducted to demonstrate the performance of the new formula in compensating the nonlinearities. The performance of the conventional formula is also presented as comparison. In the simulation, we consider two scenarios where the object under measurement is far or near to the sensor respectively. To this end, simulations are carried out by choosing system parameters respectively as ($L = 1cm$, $F = 10cm$, $\theta = 45^\circ$, $d = 25cm$) and ($L = 1cm$, $F = 5cm$, $\theta = 45^\circ$, $d = 8cm$). The object is simulated to move with a step size of $0.2cm$ in a range specified by $-1cm$ to $1cm$ in two dimensions with respect to the reference point. The measured displacements against

the ideal displacements are plotted for the two sets of system parameters respectively in Figure 4.3. It can be seen from the figures that the displacement measurements under the first set of system parameters have good linearity. This is because the object is far from the sensor and the light spot deflection on the PSD is confined in the center of the active area of PSD. Thus, both the calculation of displacement from the triangulation and the responses of the PSD can be approximated very well by linear functions. However, when the displacement measurements are carried out in the second case, the linear approximation gives large errors. But in both scenarios, the accuracy of 2-D non-contact range sensors have been improved significantly by using our new estimation formula.

Figure 4.3: Estimation of Displacement ($F=10\text{cm}$, $d=25\text{cm}$) Using The New Formula (Top Left) And The Conventional Formula (Top Right); ($F=5\text{cm}$, $d=8\text{cm}$) Using The New Formula (Bottom left) And The Conventional Formula (Bottom Right).

As mentioned in the previous chapter, the degree of nonlinearity, together with

sensitivity and maximum measurement error are three key indicators to evaluate the performance of PSD. Thus, it is important to also compare the sensitivity and maximum measurement error for this 2-D non-contact range sensor. Since the measurement results are different in X and Y directions, we compute the two indicators in both directions separately when $F = 5cm$ and $d = 8cm$. Results are tabulated below (Table 4.1). It can be seen that the maximum measurement errors are significantly reduced by our new formulae. But the sensitivity of the measurement is reduced. This is the trade-off made for linearity improvement.

Table 4.1: Comparison of Maximum Measurement Errors and Sensitivities by Using Different Formulae

	$D_{max}(\%)$	k_x
Conventional formulae (X direction)	12.9	1.08
New formulae (X direction)	4.93	0.33
Conventional formulae (Y direction)	17.9	1.50
New formulae (Y direction)	4.95	0.46

Our linearity indices approach serves as an intelligent way to optimize undetermined coefficients for the purpose of linearity improvement. Here, a comparison study is conducted between our optimization method and conventional parameter estimation methods. Similar to the previous chapter, parameters are optimized to minimize a quadratic cost function. Since the measurement results are different for x and y directions, the optimization process will be carried out with respect to each direction accordingly. The

quadratic cost functions are constructed below:

$$f_x = \sum_{i=1}^N [\hat{X}(x_i, y_i) - x_i]^2 \quad (4.5)$$

$$f_y = \sum_{i=1}^N [\hat{Y}(x_i, y_i) - y_i]^2. \quad (4.6)$$

The displacement estimates $\hat{X}$ and $\hat{Y}$ have the same formulae structure as Eq. (3.19) with 16 undermined coefficients. The data points (x_i, y_i) are chosen to be $(0.2m, 0.2n)$ inside the central square of (2×2) where m and n are integers. Again, Levenberg-Marquardt algorithm will be used to optimize these parameters by finding a local minimum of the cost function. As there is no prior knowledge on the initial guess of the parameters, it is assumed that $a_i = b_i = c_i = d_i = 1$ where $i = 1, 2, 3, 4$ as the initial guess. Same as the parameter estimation work done in Chapter 3, the optimization process is conducted in Matlab by calling its function "lsqnonlin" which uses the Levenberg-Marquardt algorithm by default. The optimized parameters obtained for $\hat{X}$ are: $a_1 = 0.9505, a_2 = 0.9505, a_3 = 0.9286, a_4 = 0.9874, b_1 = 1.0499, b_2 = 1.0499, b_3 = 1.0719, b_4 = 1.0126, c_1 = 0.9505, c_2 = 0.9505, c_3 = 0.9286, c_4 = 0.9874, d_1 = 1.0499, d_2 = 1.0499, d_3 = 1.0719, d_4 = 1.0126$. The optimized parameters obtained for $\hat{Y}$ are: $a_1 = 0.0488, a_2 = 0.2968, a_3 = -0.3261, a_4 = 0.7653, b_1 = 1.7225, b_2 = 1.5318, b_3 = 2.0064, b_4 = 1.1781, c_1 = 0.0488, c_2 = 0.2968, c_3 = -0.3261, c_4 = 0.7653, d_1 = 1.7225, d_2 = 1.5318, d_3 = 2.0064, d_4 = 1.1781$. In the same way, the object is simulated to move with a step size of $0.2cm$ in a range specified by $-1cm$ to $1cm$ in two dimensions with respect to the reference point. Then the estimated displacements are plotted in Figure 4.4. It

could be seen that there are significant measurement errors between estimated displacements and real displacements. The reason is similar to the reason given in Chapter 3. Since 16 unknown parameters are nonlinearly coupled together, there are many local minima in the cost functions. Conventional search algorithm will be terminated once a local minimum near the initial guess is found. Since there is no prior knowledge on how to choose a good initial guess near the global minimum, the measurement errors will be enlarged if the initial guess is far away from the optimum value.

Figure 4.4: Estimation result using Levenberg-Marquardt method ($F = 5cm, d = 8cm$).

4.4 Conclusion

In conclusion, the total nonlinearity of the AFM sensing system is considered in this chapter. New calculation formulae have been designed to compensate for the nonlinearity from both the triangulation principle and the 2-D tetra-lateral PSD. Simulations in various scenarios show that our new calculation formulae have greatly improved the measurement accuracy of the AFM sensing system.

Chapter 5

Improved Measurement Accuracy of the Quadrant Detector Through Improvement of Linearity Index

5.1 Introduction

The concept of linearity indices can be applied to design new calculation formulae for many other types of PSDs to improve their measurement accuracies. In this chapter, we will present the application of linearity indices to linearity improvement of the quadrant detector. The quadrant detector (QD) is used to determine the relative position of the light spot projected on its surface [63]. It is made of four identical p-n junction photodiodes placed symmetrically around its center and the photodiodes are separated by small gaps [34]. When there is an illumination on the quadrant detector, a photocurrent is produced. Under certain conditions, the strength of the photocurrent is proportional to the total light intensity on its quadrant if the light has only a single wavelength [64]. Based on the photocurrent generated on each quadrant, the position of the spot center

with respect to the QD center can be estimated by using a conventional formula [35]. Compared to the 2-D lateral-effect PSD, QD has the advantages of lower inherent noise, higher resolution and lower cost [65]. At the same time, QD can provide a continuous measurement and it has a much faster response than its competitor, the charged coupled device (CCD). Due to these advantages, QD has become a mature technology in many non-contact dimensional measurements at the micrometer scale like the optical triangulation displacement measurement [66, 67], the laser beam pointing system [68], the surface profile measurement [69] and the cylinder gauge measurement [70, 71]. Recently, QD is becoming more and more important in the area of nanotechnology. QD is widely deployed in AFM [43], and it has also been applied to membrane motion study at the leading edge of biological cells [72], the curvature measurement of the microcantilever [73] and a microelectromechanical-based magnetostrictive magnetometer [74]. In fact, the performance of all these applications relies heavily on the measurement accuracy of the QD. Unfortunately, the QD measurement is only accurate when the light beam is near the QD center. The measurement error will be significant when the light spot is further away from the QD center. This seriously reduces the percentage of the active area of the QD where the estimate of the light spot can be trusted. Thus, the improvement in the accuracy of the QD is highly desirable to enhance the application of the QD technology.

The implicit nonlinear relationship between the light spot position and its estimate makes the accuracy improvement extremely difficult. Further, the nonlinear relationship will be different if there is a change in the beam profile [75]. To reduce the measurement

errors, it has been suggested that an array of QDs can be used in the measurement rather than a single QD, which is similar to a CCD system, and each QD has an adjustable gain factor to weigh its contribution [76]. Other than using an array of QDs, one may compensate the nonlinearity of a single QD measurement by fitting the measurement data into a polynomial model [45]. Although these methods have improved the accuracy of the measurement, it is time consuming to obtain a large set of data with respect to different light spot positions in order to optimize the model parameters or the gain factors. There are also difficulties in implementations and mass productions. To provide for the linearity improvement of the QD with a simple approach, a new set of formulae to estimate the beam position from the photocurrent outputs are proposed in this chapter. The parameters of the formulae can be readily determined for different beam profiles without the need to obtain a large set of training data. Simulation results show that the measurement accuracy of the QD has been greatly improved. To verify the robustness of our approach, an experiment was performed and the obtained result shows an impressive improvement.

5.2 Theoretical Modelling

Before we present our new method, a brief review of the QD will be necessary. The QD consists of four photodiodes placed symmetrically with respect to the center (Figure 2.4). They are separated by small gaps. When it is reverse-biased, the photocurrents will be produced if there is an illumination on the QD. The conventional formula to

estimate the beam position is given by [35, 65]:

$$\hat{X} = k \frac{(I_A + I_C) - (I_B + I_D)}{I_A + I_B + I_C + I_D}, \quad \hat{Y} = k \frac{(I_A + I_B) - (I_C + I_D)}{I_A + I_B + I_C + I_D} \quad (5.1)$$

where $\hat{X}$ and $\hat{Y}$ are the estimates of the beam position in x and y directions, k is the slope constant whose value is dependent on the beam profile, and I_A , I_B , I_C and I_D are the photocurrents measured at each quadrant respectively. For a single wavelength laser beam, the strength of the photocurrent is proportional to the light beam intensity on the quadrant which could be expressed by:

$$I_i = \frac{\eta q}{h\nu} \oint_{S_i} D(x_0, y_0) dx_0 dy_0, \quad (i = A, B, C, D) \quad (5.2)$$

where S_i refers to the area of each quadrant, $D(x_0, y_0)$ is the power density of the beam at the location $[x_0, y_0]$, η is the quantum efficiency, q is the electronic charge, h is the Planck constant and ν is the light frequency. For different beam intensity profiles, the distribution of $D(x_0, y_0)$ is different. In that case, the accuracy of the position estimates will also be different. The theoretical analysis on the measurement linearity of the QD with respect to a uniform circular laser beam has been studied in Golnabi's paper [77]. In case of a uniform laser beam, the photocurrent from each quadrant is proportional to the beam area on the quadrant. However, in most applications where a single mode laser is deployed as the light source, the beam density profile is Gaussian distributed

[78]:

$$D(r) = D_0 \exp\left(\frac{-2r^2}{w^2}\right) \quad (5.3)$$

where D_0 is the power density at the beam axis, r is the distance from the beam axis, w is the radial distance at which the power density $D(r)$ drops to $1/e^2$ (13.5%) of its axial value D_0 and it is often regarded as the spot radius of a Gaussian beam. The measurement accuracy of the QD with respect to a Gaussian beam profile can be analyzed based on Ma's work [75]. Based on these models, we will present a new set of formulae to estimate the light spot position with respect to different beam profiles in the next section.

5.3 The New Estimation Formulae

In QD applications, there is an approximately linear relationship between the light spot position and its estimate when the light spot is near the QD center. But the approximation will become increasingly inaccurate when the light spot is further away from the QD center. This could be explained by Taylor series expansion. The details can be found in Chapter 3.

In our analysis, the center of the QD is taken as the origin of the coordinate: $[0, 0]$. The beam center is assumed to be located at $[x_0, y_0]$. For center symmetric beam profiles like the uniform circular spot and the Gaussian circular spot, it can be shown that $\frac{\partial \hat{X}}{\partial y_0}|_{x_0=0, y_0=0} = 0$ when it is calculated by using the conventional formula in Eq.

(5.1). When the beam is near the QD center, x_0 and y_0 are small. The first term in Eq. (3.3) dominates and those higher order terms are small. So it is observed that $\hat{X}$ is linear with respect to x_0 approximately. When the light beam is away from the center, the contribution from those higher order terms will be increased so that the linear approximation is no longer accurate. The accuracy of the measurement will be perfect if all the other terms except $\frac{\partial \hat{X}}{\partial x_0}|_{x_0=0, y_0=0}$ are zero. In that case, the estimate $\hat{X}$ will be proportional to the real position in the x direction. On the other hand, the nonlinear relationship will be very complex if $\hat{X}$ is dependent on both x_0 and y_0 . Thus, it is desirable that $\hat{X}$ is only dependent on x_0 which requires:

$$\frac{\partial^{m+n} \hat{X}(x_0, y_0)}{\partial y_0^m \partial x_0^n} \Big|_{x_0=0, y_0=0} = 0, \quad (m = 1, 2, 3, \dots, n = 0, 1, 2, \dots) \quad (5.4)$$

This requirement can be fulfilled if the following relationship is satisfied:

$$\hat{X} = f((I_A + I_C), (I_B + I_D)) \quad (5.5)$$

In other words, $\hat{X}$ must only be a function of the two variables $I_A + I_C$ and $I_B + I_D$ in order to satisfy Eq. (5.4). The linearity indices proposed in Chapter 3 are used to evaluate the degree of linearity of $\hat{X}$:

$$LH_x^n \triangleq \left| \frac{\frac{\partial^n \hat{X}}{\partial x_0^n} \Big|_{x_0=0, y_0=0}}{\frac{\partial \hat{X}}{\partial x_0} \Big|_{x_0=0, y_0=0}} \right| \quad (5.6)$$

where n is an integer such that all $LH_x^i = 0$ for $2 \leq i \leq n-1$ and $LH_x^n \neq 0$. The position

measurement $\hat{X}$ satisfying this property is referred to as n th order linearity with respect to the x direction. For simplicity, only the analysis on $\hat{X}$ will be given in details in this paper. The design of our new formulae aims to achieve the highest order linearity by simultaneously setting to zero the dominant higher order nonlinear terms in Eq. (3.3). The new formulae are specifically designed to improve the linearity performance of the QD with respect to the uniform circular beam profile and the Gaussian circular beam profile. The procedure can be readily adapted for other beam profiles.

5.3.1 The Uniform Circular Beam Profile

When the beam profile is uniform circular, the power density of the beam is uniform throughout the circular region. In our analysis, the gap between quadrants will not be considered since the gap is small. It is assumed in our paper that each quadrant of the QD is a square of $L \times L$. For a beam diameter less than L , the measurement accuracy of the QD is independent of the beam radius. When the light beam is illuminating the QD with its center coordinate $[x_0, y_0]$, the photocurrent on each quadrant can be expressed

as:

$$I_A = \frac{\eta q D_0}{h\nu} \left[\frac{\pi w^2}{4} + x_0 y_0 + \frac{x_0 \sqrt{w^2 - x_0^2} + y_0 \sqrt{w^2 - y_0^2} + (\sin^{-1}(\frac{x_0}{w}) + \sin^{-1}(\frac{y_0}{w})) w^2}{2} \right] \quad (5.7)$$

$$I_B = \frac{\eta q D_0}{h\nu} \left[\frac{\pi w^2}{4} - x_0 y_0 + \frac{y_0 \sqrt{w^2 - y_0^2} - x_0 \sqrt{w^2 - x_0^2} + (\sin^{-1}(\frac{y_0}{w}) - \sin^{-1}(\frac{x_0}{w})) w^2}{2} \right] \quad (5.8)$$

$$I_C = \frac{\eta q D_0}{h\nu} \left[\frac{\pi w^2}{4} - x_0 y_0 + \frac{x_0 \sqrt{w^2 - x_0^2} - y_0 \sqrt{w^2 - y_0^2} + (\sin^{-1}(\frac{x_0}{w}) - \sin^{-1}(\frac{y_0}{w})) w^2}{2} \right] \quad (5.9)$$

$$I_D = \frac{\eta q D_0}{h\nu} \left[\frac{\pi w^2}{4} + x_0 y_0 - \frac{x_0 \sqrt{w^2 - x_0^2} + y_0 \sqrt{w^2 - y_0^2} + (\sin^{-1}(\frac{x_0}{w}) + \sin^{-1}(\frac{y_0}{w})) w^2}{2} \right] \quad (5.10)$$

where w is the beam radius and it is assumed that $w < L$, and D_0 is the power density.

By using the conventional formula in Eq. (5.1), it can be verified that:

$$\begin{aligned} \frac{d^2 \hat{X}}{dx_0^2} \Big|_{x_0=0} &= 0, & \frac{d^3 \hat{X}}{dx_0^3} \Big|_{x_0=0} &\neq 0 \\ \frac{d^2 \hat{Y}}{dy_0^2} \Big|_{y_0=0} &= 0, & \frac{d^3 \hat{Y}}{dy_0^3} \Big|_{y_0=0} &\neq 0. \end{aligned} \quad (5.11)$$

So it can be concluded that the conventional formula has achieved a third order linearity with respect to both the x and y directions. The aim of our new formula is to achieve

a 5th order linearity in both directions with a new formula structure:

$$\begin{aligned}\hat{X} &= k \left(\frac{(I_A + I_C) - (I_B + I_D)}{a(I_A + I_C) + (I_B + I_D)} + \frac{(I_A + I_C) - (I_B + I_D)}{a(I_B + I_D) + (I_A + I_C)} \right) \\ \hat{Y} &= k \left(\frac{(I_A + I_B) - (I_C + I_D)}{a(I_A + I_B) + (I_C + I_D)} + \frac{(I_A + I_B) - (I_C + I_D)}{a(I_C + I_D) + (I_A + I_B)} \right).\end{aligned}\quad (5.12)$$

The coefficient a will be optimized for linearity improvement. It can be calculated by our new formulae that $\hat{X}$ and $\hat{Y}$ will satisfy:

$$\begin{aligned}\frac{d^3 \hat{X}}{dx_0^3} \Big|_{x_0=0} &\neq 0, & \frac{d^4 \hat{X}}{dx_0^4} \Big|_{x_0=0} &= 0 \\ \frac{d^3 \hat{Y}}{dy_0^3} \Big|_{y_0=0} &\neq 0, & \frac{d^4 \hat{Y}}{dy_0^4} \Big|_{y_0=0} &= 0.\end{aligned}\quad (5.13)$$

In order to achieve a 5th order linearity, the following equation must be satisfied:

$$\begin{aligned}\frac{d^3 \hat{X}}{dx_0^3} \Big|_{x_0=0} &= 0 \\ \frac{d^3 \hat{Y}}{dy_0^3} \Big|_{y_0=0} &= 0.\end{aligned}\quad (5.14)$$

And it will be satisfied only if $a = 1.94$.

5.3.2 The Gaussian Circular Beam Profile

In most applications, the single mode laser is deployed as the laser source. In this case, the beam profile will be a Gaussian circular. Although the analysis is more complicated in this scenario, it is a better fit to real applications. For simplicity, we assume $w = 1$ in Eq. (5.3) and $L \gg 1$. Then, the light intensity outside the four quadrants is negligible.

Thus, the light intensity on the quadrants can be calculated as follow:

$$I_A + I_C = \begin{cases} \int_0^{+\infty} D_0 \exp(-2r^2) 2\pi r dr - \int_{x_0}^{+\infty} D_0 \exp(-2r^2) (2r \cos^{-1}(x_0/r)) dr \\ (x_0 > 0) \\ \int_{x_0}^{+\infty} D_0 \exp(-2r^2) (2r \cos^{-1}(x_0/r)) dr \\ (x_0 < 0) \end{cases} \quad (5.15)$$

$$I_B + I_D = \begin{cases} \int_{x_0}^{+\infty} D_0 \exp(-2r^2) (2r \cos^{-1}(x_0/r)) dr \\ (x_0 > 0) \\ \int_0^{+\infty} D_0 \exp(-2r^2) 2\pi r dr - \int_{x_0}^{+\infty} D_0 \exp(-2r^2) (2r \cos^{-1}(x_0/r)) dr \\ (x_0 < 0) \end{cases} \quad (5.16)$$

$$I_A + I_B = \begin{cases} \int_0^{+\infty} D_0 \exp(-2r^2) 2\pi r dr - \int_{y_0}^{+\infty} D_0 \exp(-2r^2) (2r \cos^{-1}(y_0/r)) dr \\ (y_0 > 0) \\ \int_{y_0}^{+\infty} D_0 \exp(-2r^2) (2r \cos^{-1}(y_0/r)) dr \\ (y_0 < 0) \end{cases} \quad (5.17)$$

$$I_C + I_D = \begin{cases} \int_{y_0}^{+\infty} D_0 \exp(-2r^2) (2r \cos^{-1}(y_0/r)) dr \\ (y_0 > 0) \\ \int_0^{+\infty} D_0 \exp(-2r^2) 2\pi r dr - \int_{y_0}^{+\infty} D_0 \exp(-2r^2) (2r \cos^{-1}(y_0/r)) dr \\ (y_0 < 0) \end{cases} \quad (5.18)$$

In the same way, it can be verified that the conventional formula has achieved a third order linearity in both directions. In order to achieve a 5th order linearity in both directions, the parameter a in our new formula is recalculated as $a = 2.97$.

5.4 Simulation Results

We shall now present the simulation study to evaluate and compare the performance of the conventional and the new formulae with respect to the uniform circular and the Gaussian circular beam profiles. In our simulations, it is still assumed that $w = 1$ and $L \gg 1$. The light beam center is assumed to move in steps of 0.1 in both directions from the QD center within the square of 1×1 , one quadrant of the whole effective area. The measurement accuracy of the whole effective area is symmetric to both x and y axes so that the measurement result will be exactly the same if the beam moves to other quadrants. The estimates of the light beam center will be calculated by using both the conventional formula and our new formula. The estimates of the light beam center in case of a uniform circular beam by using the conventional and our new formula are plotted in Figure 5.1 and Figure 5.2, respectively. The estimates of the light beam center in case of a Gaussian circular beam by using the two formulae are plotted in Figure 5.3 and Figure 5.4, respectively. It can be seen that the QD measurements via the conventional formula have severe nonlinearity when the light beam center is away from the QD center. Our new formulae have significantly improved the measurement accuracy, especially when the light beam is towards the periphery of the QD. For a more quantitative comparison, the percentage of the effective area with measurement

errors less than 1%, 2%, 3% and 4% are tabulated in Table 5.1 and Table 5.2. The measurement error is defined in [54] as $E = \frac{\hat{X}/k - x_0}{x_0}$.

Table 5.1: Comparison of Estimation Distortion for Uniform Circular Beam

	Percentage of area with distortion less than			
	1%	2%	3%	4%
Conventional formula	24%	34%	41%	47%
New formula	58%	68%	74%	79%

Table 5.2: Comparison of Estimation Distortion for Gaussian Circular Beam

	Percentage of area with distortion less than			
	1%	2%	3%	4%
Conventional formula	12%	17%	21%	24%
New formula	34%	42%	48%	52%

In general, the percentage of the effective area with less than 4% measurement error has been improved by two to three times by using our new formulae. This is a significant enhancement to the measurement accuracy without any extra complexity in the QD system. In addition, the sensitivities and maximum measurement errors of the measurements by using conventional formulae and our new formulae in both scenarios are computed in Table 5.3. It can be seen that the maximum measurement errors are reduced by nearly 50% in both cases when our new formulae are used. The sensitivities are slightly increased in both scenarios when our new formulae are used. This is different from 2-D tetra-lateral PSD where the measurement sensitivities are reduced by using our new formulae.

Next, a comparison study on linearity improvement by different parameter optimization methods is presented. Our approach is based on improving the linearity indices. In conventional optimization approaches, searching algorithms like Levenberg-Marquardt

Figure 5.1: The estimate of uniform circular beam center by using the conventional formula.

method and Newton's method are applied to find the global minimum of a cost function.

The cost function constructed is the same as Eq. (3.21) in Chapter 3. We assume $\hat{X}$ has the same formula structure as our proposed new formulae (Eq. (5.3.1)) with 16 undetermined coefficients nonlinearly parameterized:

$$\hat{X} = \frac{a_1 I_A + a_2 I_B + a_3 I_C + a_4 I_D}{b_1 I_A + b_2 I_B + b_3 I_C + b_4 I_D} + \frac{c_1 I_A + c_2 I_B + c_3 I_C + c_4 I_D}{d_1 I_A + d_2 I_B + d_3 I_C + d_4 I_D}. \quad (5.19)$$

The optimization process is carried out when the light beam is uniform circular. The data points (x_i, y_i) are chosen to be $(0.1m, 0.1n)$ inside the square (1×1) where m and n are non-negative integers. Since there is no prior knowledge on the initial guess of the parameters, we assume $a_i = b_i = c_i = d_i = 1$ where $i = 1, 2, 3, 4$ as

Figure 5.2: The estimate of uniform circular beam center by using our new formula.

Table 5.3: Comparison of Maximum Measurement Errors and Sensitivities for Uniform Circular Beam and Gaussian Circular Beam

	D_{max} (%)	k_x
Conventional formula (Uniform circular beam)	21.46	1.27
New formula (Uniform circular beam)	12.46	1.73
Conventional formula (Gaussian circular beam)	40.1	1.59
New formula (Gaussian circular beam)	22.7	1.60

the initial guess. The optimization process is done in Matlab by calling its function "lsqnonlin" which uses Levenberg-Marquardt algorithm by default. The optimized parameters are obtained as: $a_1 = 0.4301, a_2 = 0.9646, a_3 = 0.7939, a_4 = 0.9928, b_1 = 1.4701, b_2 = 1.0292, b_3 = 1.1679, b_4 = 1.0059, c_1 = 0.4301, c_2 = 0.9646, c_3 = 0.7939, c_4 = 0.9928, d_1 = 1.4701, d_2 = 1.0292, d_3 = 1.1679, d_4 = 1.0059$. The estimate of the light spot position calculated via Eq. (5.3.1) with these optimized parameters are plotted in Figure 5.5 when the light beam center is assumed to move in steps of 0.1 in both direc-

Figure 5.3: The estimate of Gaussian circular beam center by using the conventional formula.

tions from the QD center within the square of 1×1 . It can be seen that the measurement linearity is worse than the conventional formula. The reason why measurement linearity becomes worse when undermined parameters are optimized by conventional methods has been explained in Chapter 3 and 4.

5.5 Experimental Results

In real QD applications, the single mode laser is mostly employed as the laser source. In practice, besides the fundamental mode, higher order modes also exist in the single mode laser [79, 80]. In this case, the laser beam profile from the laser source is no longer a pure Gaussian. Along the propagation, the air turbulence [75] will also affect the QD

Figure 5.4: The estimate of Gaussian circular beam center by using our new formula.

measurement. In addition, the response of each quadrant may not be uniform. Even if each quadrant receives the same light intensity, the photocurrents may differ due to the inhomogeneous photodiodes [81]. However, to cater to the need of mass production and easy implementation, it is preferred not to measure the response of the photodiodes and the intensity profile of the beam on the QD to recalculate the parameter a in our new formula for every setup. Thus it is essential that our new formula should be robust to these deviations in real applications. To test the accuracy improvement of the QD in practice, an experiment has been conducted. In the experiment, the light source is a single mode He-Ne laser with the laser wavelength of 633 nm . A collimator is connected to the light source to reduce its divergence. A SPOT-9DMI QD from OSI Electronics Inc. is used in our experiment. This QD has an active area of $10\text{ mm} \times 10\text{ mm}$ with a

Figure 5.5: The estimate of uniform circular beam center by using using Levenberg-Marquardt method.

gap width of 0.01 mm . The QD is mounted on a PZT motion stage so that the relative beam position with respect to the QD can be adjusted. The detailed experiment setup is plotted in Figure 5.6. In the experiment, the beam center is assumed to be at the center of the QD when the photocurrents from the four quadrants are equal. The slope constant k is calculated by measuring the change in each photocurrent when there is a small displacement of the light beam from the QD center. To compare the linearity performance, the beam position is changed in steps of 0.2 mm in both x and y directions in one quadrant and the information of photocurrents is recorded. The estimates of the beam position processed via the conventional formula and our new formula with respect to the Gaussian circular beam are plotted in Figure 5.7 and Figure 5.8, respectively. A

Figure 5.6: The experiment setup.

Figure 5.7: The experimental result when the conventional formula is applied.

significant linearity improvement is achieved when our new formula is implemented.

Figure 5.8: The experimental result when our new formula is applied.

5.6 Conclusions

In conclusion, based on the linearity indices proposed in Chapter 3, a new set of formulae has been derived for estimating the relative position of the light spot on the QD PSD in this chapter. It improves the linearity of the QD measurement with respect to different light beam profiles. It has been shown in simulations that the measurement accuracy of the QD has been significantly improved. It is also shown in real experiment that our method is robust and efficient in practice. With the great demand for the linearity improvement in QD measurement without added complexity, our approach would be very useful in QD related researches and industries.

Chapter 6

A Novel Probe-sample Separation

Estimation Scheme for Atomic

Force Microscopy

6.1 Introduction

The atomic force microscopy (AFM) belongs to the scanning probe microscopy (SPM) family and it was invented to measure the surface topography in the nanometer scale [2]. The topography information is obtained from the measurement of the separation dependent atomic force between the AFM probe and the sample. Besides the topography information, material properties like adhesion can also be calculated from the atomic force measurements [82]. Other than passive measurements, researchers are able to use AFM to manipulate the atoms to fabricate nano-sized electronics [83]. Compared to other SPM family members, there is no requirement on the sample conductivity for AFM applications and samples can be prepared in air or liquid environment [84, 85, 86]. Due to these advantages, AFM is widely used in materials science, biological science

and semiconductor industry [87]. The success of AFM relies heavily on its measurement accuracy and resolution. Besides the measurement accuracy of AFM sensing system, the measurement accuracy of the AFM relies heavily on constant separation control between the probe and the sample while moving the probe to follow the profile of the sample [88]. A proper constant separation control can prevent the probe from clashing into the sample during scanning and the profile of the surface is exactly the same as the movement of the probe. To ensure constant separation during scanning, it is essential to obtain an accurate estimation of the probe-sample separation.

Currently, there is no separation estimation scheme that can guarantee good estimation accuracy. The challenge to obtain an accurate separation estimate is the complex nonlinear behavior of the probe-sample interaction. The atomic force, mainly the long range van der Waals' force and the short range repulsive force, is nonlinear with respect to the probe-sample separation [17]. Rather than the static measurement of the atomic force, the probe is excited to vibrate at its resonance to reduce the influence of noise [26]. Thus, the model of the dynamic probe-sample interaction is considered as a second order oscillator subject to the nonlinear atomic force. Unknown parameters like the probe-sample separation and force-related constants are nonlinearly parameterized in the dynamic system which make parameter estimation difficult. In practice, under normal operation condition, it is found that the probe vibration is almost a pure sinusoidal [89]. In the conventional constant separation control scheme, it is assumed that the probe-sample separation and the probe vibration amplitude has a one-to-one relationship during scanning. When the probe is moved to another sample spot and

the probe-sample separation is changed during scanning, the AFM feedback system will adjust the probe to approach or retract from the sample until the vibration amplitude reaches the pre-set reference value. In that case, the probe-sample separation is considered to be maintained at a constant level. However, the one-to-one relationship between the probe-sample separation and probe vibration amplitude is not always valid. First of all, it is under the assumption that the material properties of the sample are homogeneous which is not true for real samples. For inhomogeneous samples, when the vibration amplitude reaches the same reference value during scanning, the phase difference between the external excitation signal and the probe vibration is different [90] which could imply that the probe-sample separation is not maintained at a constant level although the probe vibration amplitudes are the same. Even if the sample is homogeneous, it is found in simulations and experiments that two different separations can lead to the same vibration amplitude [17, 30]. Both factors will introduce tremendous errors in the conventional constant separation control scheme and the topography image obtained. Besides the signal at the fundamental frequency, harmonics due to the nonlinear atomic force do exist and have been measured in many simulations and experiments [91, 92]. The existence of higher order harmonics will make the conventional constant separation control scheme and topography measurement even more inaccurate. Prior to our work, Sebastian *et al.*[91] proposed a very impressive method called harmonic and power balance tools to estimate the unknown parameters in the interaction force based on the signals from both fundamental frequency and harmonics. But their force model is piecewise linear and the unknown parameter are linearly coupled. Thus, their method

is not suitable for our problem.

Here, to improve the measurement accuracy of the AFM and to maintain a constant separation during scanning, an accurate estimation scheme is proposed to calculate the probe-sample separation. Other unknown force-related parameters characterizing the material properties can be calculated once the probe-sample separation is determined. In this chapter, the theoretical modelling of the dynamic probe-sample interaction will first be introduced, followed by the mathematical formulation of our estimation problem. Then, our estimation scheme is proposed with a detailed theoretical proof. Only the amplitude and the phase of the probe vibration at its fundamental frequency and its harmonics, up to the fourth order, are processed in the scheme. To verify the accuracy and robustness of our estimation scheme, simulations have been done. Our simulation results show that the new estimation scheme can provide an accurate estimate even when strong noise exists. Thus, the proposed approach has the potential to achieve reliable and repetitive AFM topography images in practice. It is also capable of providing additional information about the sample material properties once an accurate probe-sample separation has been estimated.

6.2 Theoretical Model

In this section, a comprehensive model of the dynamic probe-sample interaction will be described. The derivation of our estimation scheme and numerical simulations will be based on this model. In the AFM setup, a piezoelectric actuator is attached to the cantilever-probe ensemble. A sinusoidal electrical signal at the frequency near the

Figure 6.1: The schematic of the probe-sample interaction.

resonance of the ensemble will be applied to the actuator which in turn oscillates the ensemble. The schematic of the probe-sample interaction is shown in Figure 6.1.

In our analysis, the cantilever-probe ensemble is considered as a point-mass spring [31]. The equation to describe this dynamic probe-sample interaction is:

$$m\ddot{x} + c\dot{x} + kx = f \sin(\omega t + \phi_1) + F_{int} \quad (6.1)$$

where m is the mass of the probe, c is the damping coefficient, k is the spring constant, f and ω are respectively the amplitude and the frequency of the excitation signal, x is the probe position, and F_{int} denotes the atomic force. The Lennard-Jones model is used to model the atomic force [93]:

$$F_{int} = -\frac{A}{(x+z)^2} + \frac{B}{(x+z)^8} \quad (6.2)$$

where A and B are force-related parameters and they will vary with the sample material properties and z is the probe-sample separation when the probe is at equilibrium. During

scanning, the excitation signal $f \sin(\omega t + \phi_1)$ is available for measurement. System parameters m , c and k can be calculated during the free vibration. The probe trajectory $x(t)$ can be measured from the AFM detection system. The probe-sample separation z is unknown and will be estimated from our proposed estimation scheme. The force-related parameters A and B are also unknown which makes the separation estimation even more challenging. But once the separation z has been successfully estimated, A and B can be easily calculated based on linear identification methods [82].

Under normal operation condition, the probe vibration is dominated by the resonant sinusoidal and its higher order harmonics also exist. In that case, the probe trajectory $x(t)$ can be expressed as:

$$x(t) = d_0 + d_1 \sin(\omega t) + \sum_{n=2}^{\infty} d_n \sin(n\omega t + \phi_n) \quad (6.3)$$

6.3 Proposed Estimation Scheme

Under normal operation condition, the signal of the main sinusoidal is much more significant than its higher order harmonics. In that case, it is reasonable to neglect the contribution of the higher order harmonics to the atomic force so the system equation can be approximated by:

$$\begin{aligned} m\ddot{x} + c\dot{x} + kx &= f \sin(\omega t + \phi_1) - \frac{A}{(d_1 \sin(\omega t) + d_0 + z)^2} \\ &+ \frac{B}{(d_1 \sin(\omega t) + d_0 + z)^8} \end{aligned} \quad (6.4)$$

It is found that the Fourier spectra of the atomic force can be expressed by Legendre function based on the following equations obtained from the residue theory:

$$\begin{aligned} \int_0^T \frac{\sin(m\omega t)}{(1+b\sin(\omega t))^n} dt &= T \frac{(-1)^{\frac{m+1}{2}} (n-1-m)!}{(1-b^2)^{\frac{n}{2}} (n-1)!} \\ &\times P_{n-1}^m\left(\frac{1}{\sqrt{1-b^2}}\right) \\ &(m < n, m = 1, 3, 5 \dots) \end{aligned} \quad (6.5)$$

$$\int_0^T \frac{\sin(m\omega t)}{(1+b\sin(\omega t))^n} dt = 0 \quad (m = 2, 4, 6 \dots) \quad (6.6)$$

$$\begin{aligned} \int_0^T \frac{\cos(m\omega t)}{(1+b\sin(\omega t))^n} dt &= T \frac{(-1)^{\frac{m}{2}} (n-1-m)!}{(1-b^2)^{\frac{n}{2}} (n-1)!} \\ &\times P_{n-1}^m\left(\frac{1}{\sqrt{1-b^2}}\right) \\ &(m < n, m = 0, 2, 4 \dots) \end{aligned} \quad (6.7)$$

$$\int_0^T \frac{\cos(m\omega t)}{(1+b\sin(\omega t))^n} dt = 0 \quad (m = 1, 3, 5 \dots) \quad (6.8)$$

where T is the period of the signal and $P_{n-1}^m(x)$ is the associate Legendre function of the first kind. For convenience of calculation, it is denoted that:

$$\begin{aligned} L &\triangleq (m\ddot{x} + c\dot{x} + kx - f \sin(\omega t + \phi_1)) \\ &\times (1 + 2r \sin(\omega t) + r^2 \sin^2(\omega t)). \end{aligned} \quad (6.9)$$

It is easy to verify that the system dynamics described in Eq. (6.4) is equivalent to:

$$L = \frac{-A}{z^2} + \frac{B}{z^8(1+r\sin(\omega t))^6}. \quad (6.10)$$

where $r = \frac{d_1}{z+d_0}$. Based on Eq. (6.5)- Eq. (6.8), it is found that:

$$\begin{aligned}
& \frac{TB(-30r^7 - 90r^5 + 72r^3 + 48r)}{16z^8(r^2 - 1)\sqrt{1 - r^2}} \\
&= (n_0r + m_1)(r^2 - 1) \\
&= \int_0^T L\left(\frac{-4\sin(5\omega t) + 18\sin(3\omega t) - \sin(\omega t)}{42}\right)dt \tag{6.11}
\end{aligned}$$

which would lead to the following quadratic equation after simple manipulations:

$$\left(\frac{6}{7}n_0 + \frac{2}{7}n_2 - \frac{1}{7}n_4\right)r^2 + \left(\frac{33}{28}s_1 + \frac{9}{28}s_3\right)r - \frac{3}{7}n_2 = 0 \tag{6.12}$$

where $s_i = \int_0^T [m\ddot{x} + c\dot{x} + kx - f\sin(\omega t + \phi_1)]\sin(i\omega t)dt$, $n_i = \int_0^T [m\ddot{x} + c\dot{x} + kx - f\sin(\omega t + \phi_1)]\cos(i\omega t)dt$, ($i=0,1,2,3..$). In the quadratic equation, r is the ratio of the probe vibration amplitude at the fundamental frequency to the sum of the probe-sample separation and the offset. Once r is obtained, the probe-sample separation z can be easily calculated. s_i and n_i can be calculated by only analyzing the Fourier spectra of the probe vibration $x(t)$. Since the coefficients can be calculated from the measurement outputs, the unknown parameter r can be obtained by solving the quadratic equation and it is given by:

$$r = \frac{-\frac{33s_1+9s_3}{28} - \sqrt{\left(\frac{33s_1+9s_3}{28}\right)^2 + \left(\frac{6n_0+2n_2-n_4}{7}\right)\frac{12n_2}{7}}}{2\left(\frac{6n_0+2n_2-n_4}{7}\right)} \tag{6.13}$$

Thus, the probe-sample separation z can be calculated from the following relationship:

$$z = \frac{2d_1 \left(\frac{6n_0+2n_2-n_4}{7} \right)}{-\frac{33s_1+9s_3}{28} - \sqrt{\left(\frac{33s_1+9s_3}{28} \right)^2 + \left(\frac{6n_0+2n_2-n_4}{7} \right) \frac{12n_2}{7}}} - d_0 \quad (6.14)$$

In order to calculate the probe-sample separation z , only the amplitude and the phase of the probe vibration at the fundamental frequency and its harmonics up to the fourth order are needed. Although fourth order harmonics are difficult to be measured using normal stiff cantilever, special cantilevers like torsional harmonics cantilevers are designed for the purpose of enhancing the higher order harmonics [94]. In the next section, numerical simulation results will be provided to verify the accuracy of the estimation scheme. In addition, noise is added to test its robustness.

6.4 Simulation Results and Discussions

Numerical simulations have been conducted to simulate the dynamics of the probe-sample interactions. In our simulations, the cantilever-probe ensemble is assumed to be monocrystalline and the sample material is assumed to be silicon. The system parameters of the cantilever-probe ensemble are similar to parameters used in Rützel's simulations [93]: $m = 2.093 \times 10^{-11} kg$, $c = 1.447 \times 10^{-8} kg/s$, $k = 0.1 N/m$, $A = 4.663 \times 10^{-27} kg m^3/s^2$ and $B = 1.133 \times 10^{-79} kg m^9/s^2$. The excitation frequency is set at the resonant frequency of the cantilever: $f_0 = 11 kHz$. In our simulations, only the probe-sample dynamics under normal operation condition will be simulated while chaos and other complex nonlinear dynamics are avoided. The typical probe trajectory is shown in Fig-

Figure 6.2: The simulated probe trajectory under normal operation condition.

Figure 6.2 when the excitation amplitude and the probe-sample separation are chosen as $f = 4.093 \times 10^{-11} N$ and $z = 16 \text{ nm}$. From Figure 6.2, it could be seen that the probe vibration is periodic. To investigate the probe vibration in more details, its Fourier spectra are calculated and plotted in Figure 6.3. From its Fourier spectra, it could be seen that the signal at the fundamental frequency is much stronger than its harmonic and the offset. Based on the simulation data, our proposed estimation scheme is tested. In practice, the data of the probe trajectory is not clean and the measurement can be significantly influenced by random noise. Hence, our proposed estimation scheme must be robust and needs to be accurate even when noise is present. To test the robustness, we add Gaussian noise with different signal to noise ratio (SNR) to the measurement of the probe vibration $x(t)$. The Gaussian noise added in the simulation is quite strong compared to the weak harmonics. The estimation results are summarized in Table 6.1.

Figure 6.3: The Fourier spectra of the probe vibration.

To be more convincing, we have tested our scheme under different probe-sample separations. The simulation results are summarized in Table 6.2 when the probe-sample separation $z = 12 \text{ nm}$ and other conditions remain unchanged.

Table 6.1: The summary of the estimation results when $z=16\text{nm}$

	z=16nm	
The estimate of z	Mean (nm)	Standard derivation (nm)
Noise free	16.14	-
SNR = 50dB	16.13	0.09
SNR = 40dB	16.17	0.35
SNR = 30dB	16.31	1.23

From the simulation results shown in Table 6.1 and Table 6.2, it is found that the estimate of the probe-sample separation is very close to the true value with negligible errors even under severe noise condition. The small estimation error is due to the

Table 6.2: The summary of the estimation results when $z=12\text{nm}$

	$z=12\text{nm}$	
The estimate of z	Mean (nm)	Standard derivation (nm)
Noise free	11.78	-
SNR = 50dB	11.78	0.04
SNR = 40dB	11.78	0.13
SNR = 30dB	11.77	0.36

approximation used in Eq. (6.4) in the derivation of our proposed scheme and it is the only approximation used in the analysis. Since our estimation scheme only relies on the information of the Fourier spectra, the noise will be averaged out.

6.5 Conclusions

In conclusion, we have presented a novel probe-sample separation scheme in this chapter to address the problem of the probe-sample separation estimation. It has been shown in both theory and simulations that the proposed scheme is able to calculate the unknown probe-sample separation accurately even in severe noisy condition. In addition, the proposed estimation scheme is easy to apply in practice because only the Fourier spectra of the probe vibration need to be measured and processed. In the future, experiments will be carried out to implement the scheme to yield a reliable topography image of the sample. Other than the sample topography, material properties can also be calculated once the probe-sample separation is estimated, which will equip AFM with more functionalities.

Chapter 7

An Accurate Separation Estimation Algorithm for the Casimir Oscillator

7.1 Introduction

Casimir force exists between two conductive metals when their separation is comparable with the electromagnetic wavelengths. In principle, the Casimir force has the same origin as the van der Waals' force, that is it arises from the imposed boundary condition of the electromagnetic field by the metals [95]. The imposed boundary condition leads to a different distribution of the electromagnetic mode density as compared to the free space. When two perfectly conductive rectangular metal plates are placed in parallel, the Casimir force is attractive and its strength is inversely proportional to the fourth power of the separation between the two plates [95]. The Casimir force was mostly measured experimentally between a metal sphere and a metal plate to avoid the difficulty in plate alignment [96, 97]. Recently, more and more Casimir force measurements are conducted by AFM where the probe is replaced by a sphere with a bigger radius [98, 99]. Between a perfect sphere and a plate, the attractive Casimir force is found to

be inversely proportional to the cubic power of their separation [100] and it will become significant when their separation is around 100 nm [101]. The Casimir force is playing an increasingly important role in microelectromechanical systems (MEMS) [102] as the separation between the components in MEMS is in the sub-micrometer regime [103]. On the other hand, the Casimir force can also be utilized as an actuation force in nanoelectromechanical systems (NEMS) which are increasingly important in sensors, data storage and medical diagnostics at the nanometer scale [104]. A typical electromechanical system actuated by the Casimir force consists of a polysilicon plate suspended by thin torsional rods, two fixed polysilicon electrodes on a piezo to detect the rotation of the plate and a sphere metallic surface [105]. When the piezo extends or retracts, the change in the separation between the sphere surface and the plate leads to a different strength of the Casimir force in between. Then the plate will rotate according to the change of the Casimir force. A sinusoidal driving voltage is applied to one of the electrodes to oscillate the plate [106]. When the driving voltage is small and fixed, it is found that the oscillation of the plate is almost sinusoidal and the oscillation amplitude changes with respect to the separation between the plate at equilibrium and the sphere surface. Compared to static measurements, the dynamic oscillator has the advantage of high sensitivity with respect to the separation [106]. This motivates the use of the oscillator as a separation sensor at the nanometer scale [107]. As the size of MEMS is becoming smaller and smaller, there is a high demand for the separation sensor at the nanometer scale. For example, in electrostatic actuators, the separation gap between electrodes determines the pull-in voltage which needs to be measured properly [108, 109].

However, the difficulty of achieving an accurate estimate of the separation gap in the Casimir oscillator is a bottleneck to its development as a separation sensor. As an actuation force in the oscillator, the Casimir force is nonlinearly dependent on the separation gap. Besides the unknown separation gap, the lateral distance between the sphere and the plate center is difficult to measure and it should be treated as an unknown parameter too. The existence of two unknown parameters nonlinearly parameterized in the dynamic system makes the estimation challenging. It is found that when a strong external excitation is applied, the hysteresis phenomenon is observed where two stable plate oscillations with different amplitudes co-exist [106]. This nonlinear phenomenon makes the task of estimating the separation even more difficult but the sensitivity of the measurement is higher under this condition. In many parameter estimation problems, algorithms originated from Kalman filters [111] and adaptive least squares algorithms [112, 113] are applied. However, accurate estimates of the unknown parameters from these two methods can only be guaranteed when the unknown parameters are linearly parameterized. Steepest descent method and Newton's method are designed for solving nonlinear parameterized estimation problems in static systems [114]. Direct applications of these methods to our dynamic system require the data of the plate acceleration which is difficult to measure. And the estimation accuracy is not guaranteed because the estimation process will terminate unsuccessfully in one of the local minima of the cost function. Currently, for our problem there is no algorithm yet that can guarantee the convergence of their estimates to the true value [115]. In order to obtain an accurate estimate of the separation gap, a cost function with respect to the estimates of the

unknown parameters must be constructed and the computation of this cost function must be based on the available measurement data. The cost function is also required to have no local maximum and only one local minimum where the estimates of the unknown parameters are equal to their true values. Then, the steepest descent method, Newton's method or other conventional nonlinear optimization techniques can be used to search for the local minimum and the accurate estimates. Based on this rationale, we first simplify our problem of estimating multiple unknown parameters to one of a single unknown parameter estimation problem by constructing a cost function with only one local minimum and no local maximum with respect to the difference between the single unknown parameter and its estimate. In the second step, the golden section search method is applied to search for the local minimum of the cost function and the accurate estimate. Theories and extensive simulations are given in this paper to prove the accuracy and robustness of our method. Other than the separation gap estimation in the Casimir oscillator, our scheme is also applicable to many other parameter estimation problems with similar mathematical models like the gap separation estimation problem in electrostatically actuated MEMS [10]-[12].

In this chapter, a literature review on the Casimir force is presented first. Then the mathematical model of the Casimir oscillator is discussed, and various nonlinear phenomena are observed in our simulations under different conditions. Our nonlinear estimation algorithm is proposed to estimate the gap separation. To verify the accuracy of our method, simulations have been conducted and a theoretical proof of the estimation accuracy of our algorithm is provided.

7.2 Review of the Casimir Force

In 1948, Casimir has conducted a theoretical study on the distribution of the electromagnetic waves in a perfectly conducting cavity [95]. He considered a cubic cavity of volume L^3 bounded by perfectly conducting walls. And a perfectly conducting square plate with side length L is placed in this cavity parallel to the xy face (Figure 7.1). The distance between the plate and the xy face is assumed to be a . The conductive plate and the xy face have imposed the boundary conditions for the vibrated electromagnetic waves in between. Only the electromagnetic modes that have nodes on both the plate and the xy face can exist in between. This reduces the density of the electromagnetic waves. The wave numbers between the plate and the xy face are:

$$k_x = \frac{\pi}{L}n_x, \quad k_y = \frac{\pi}{L}n_y, \quad k_z = \frac{\pi}{a}n_z \quad (7.1)$$

where n_x , n_y and n_z are positive integers. When L is large enough, k_x and k_y can be regarded as continuous variables. When a is small, the interaction energy between the plate and the xy face can be calculated as:

$$E = \frac{\hbar\omega}{2} = \hbar c \frac{L^2}{\pi^2} \int_0^\infty \int_0^\infty \left[\frac{1}{2} \sqrt{k_x^2 + k_y^2} + \sum_{n=1}^\infty \sqrt{n^2 \frac{\pi^2}{a^2} + k_x^2 + k_y^2} \right] dk_x dk_y \quad (7.2)$$

where $\hbar$ is the reduced Planck constant and c is the light velocity. Casimir found that this interaction energy is different if the plate is placed far away from the xy face where k_z is considered as a continuous variable. And the difference in the interaction energy

Figure 7.1: The schematic of the Casimir cavity.

Figure 7.2: The density of electromagnetic waves in free space and between conductors.

is calculated as:

$$\delta E = \hbar c \frac{\pi^2 L^2}{720 a^3}. \quad (7.3)$$

This introduces an attractive force between the plate and the xy face when their separation a is small. The force per unit area is calculated as:

$$F = -\hbar c \frac{\pi^2}{240a^4} \quad (7.4)$$

where the minus sign denotes that the force is attractive. It is very difficult to verify the Casimir force between plates because large errors will be produced from maintaining a high degree of parallelism [116]. To remove the difficulty in plate alignment, most measurements of the Casimir force are carried out between a sphere and a plate where the strength of the attractive Casimir force is calculated as [100]:

$$F = \frac{-\pi^3 \hbar c R}{360d^3} \quad (7.5)$$

where R is the radius of the sphere and d is the minimum distance between the sphere and the plate. In essence, the Casimir force can be understood in a unified way with the well-known van der Waals' force [117]. Both the van der Waals' force and the Casimir force originate from quantum fluctuation [118]. The key difference is that the Casimir force considers the retardation of the electromagnetic fluctuating interaction due to the finiteness of the velocity of light when the separation between two interaction bodies is large enough [117].

In practice, the Casimir force is weaker in non-ideal metals. Lifshits developed a general theory to calculate the Casimir force between macroscopic bodies [119, 120] based on their dielectric functions. For example, the Casimir force between a sphere

and a plate of the same material can be calculated from [119]:

$$F(z) = \frac{\hbar R}{2\pi c^2} \int_0^\infty \int_1^\infty p \xi^2 \left[\ln \left[1 - \frac{(s-p)^2}{(s+p)^2} e^{-2pz\xi/c} \right] + \ln \left[1 - \frac{(s-p\varepsilon)^2}{(s+p\varepsilon)^2} e^{-2pz\xi/c} \right] \right] dp d\xi \quad (7.6)$$

where $s = \sqrt{\varepsilon - 1 + p^2}$, ε is the dielectric function at the imaginary frequency $i\xi$. Besides the imperfectness of the metal, the surface roughness will also alter the strength of the Casimir force [121].

Although most of the Casimir forces generated by interacting materials are attractive, a repulsive Casimir force is possible if suitable interacting materials and the medium in between are chosen. When the dielectric functions of the interacting materials and the medium satisfy the following relationship, the Casimir force is repulsive [120]:

$$\varepsilon_1(i\xi) < \varepsilon_3(i\xi) < \varepsilon_2(i\xi)$$

or

$$\varepsilon_2(i\xi) < \varepsilon_3(i\xi) < \varepsilon_1(i\xi) \quad (7.7)$$

where ε_1 and ε_2 are the dielectric functions of the interacting materials and ε_3 is the dielectric function of the medium. On the other hand, Kenneth [122] developed a condition for the repulsive Casimir force which relies on the permeability and permittivity of the interacting materials and the medium. Recently, the repulsive Casimir force has

been measured between gold and silica materials where both materials are immersed in a fluid [123].

7.3 Mathematical Model of the Casimir Oscillator

The Casimir oscillator fabricated in Chan's work [106] consists of a polysilicon plate free to rotate about the rods, a metallic sphere and two electrodes. Besides this setup, the Casimir oscillator can also be realized by a modified AFM where the probe is replaced by a bigger sphere. Our proposed algorithm is based on Chan's setup. In Chan's oscillator, a driving sinusoidal voltage can be applied to one of the electrodes to excite the plate rotation. The separation of the plate and the sphere can be adjusted by the piezo movement. When there is no external force on the plate, the plate rotation is sinusoidal and proportional to the amplitude of the driving voltage. In the Casimir oscillator, a metallic sphere is used to interact with the plate. Besides the Casimir force between the sphere and the plate, the electrostatic force also exists if there is a residual voltage on the sphere. The electrostatic force can be eliminated if the residual voltage is neutralized when the plate is connected to the ground. The rotation of the plate can be measured indirectly by measuring the change of the capacitance between the detection electrode and the plate. Since the rotation is small compared to the size of the plate, the change of the capacitance is proportional to the plate rotation. The velocity of the plate rotation can be measured by a heterodyne laser Doppler interferometer which has already been implemented in AFM successfully [124]. The schematic of the setup is shown in Figure 7.3. The dynamics of the plate rotation is modelled as:

Figure 7.3: The Casimir oscillator setup (The figure is adapted from Fig. 2 in [105]).

$$\ddot{\theta} + 2\gamma\dot{\theta} + \omega_0^2\theta = \frac{\tau}{I} \sin \omega t + F(z - b\theta) \frac{b}{I} \quad (7.8)$$

where θ is the rotation angle, γ is the damping coefficient, I is the moment of inertia of the plate, ω_0 is the fundamental frequency of the oscillator, τ is the amplitude of the driving torque, ω is the excitation frequency, b is the lateral distance between the sphere and the plate center and F denotes the Casimir force. As mentioned in the previous section, the Casimir force between the sphere and the plate is modelled as [100]:

$$F(z - b\theta) = \frac{-\pi^3 \hbar c R}{360(z - b\theta)^3} \quad (7.9)$$

where z is the separation between the sphere and the plate when the plate is at equilibrium. Since the Casimir force is nonlinear with respect to θ , various complex nonlinear dynamics will be observed under different conditions and they will be discussed in the

next section.

7.4 Dynamics of The Casimir Oscillator

In this section, numerical simulations are performed to present various dynamics of the oscillator. In our simulations, system parameters are chosen to be exactly the same as the parameters in Chan's setup [106]: $\omega_0 = 17198 \text{ rad/s}$, $R = 100 \text{ }\mu\text{m}$, $b = 131 \text{ }\mu\text{m}$ and $I = 7.1 \times 10^{-17} \text{ kg m}^2$. The quality factor of the system is chosen to be $Q = 71.5$ which is equivalent to $\gamma = 120$. All the system parameters above will remain unchanged in our study. The fourth order Runge-Kutta numerical method with a step size of 1×10^{-7} second is used to simulate the dynamics of the Casimir oscillator.

7.4.1 Small Nonlinearity

When the external excitation is small and fixed, the oscillation is a pure sinusoid and the oscillation amplitude A is linear with respect to the gap separation z . This scenario is simulated in Figure 7.4 while the excitation frequency and excitation torque are chosen as $\omega = 17091 \text{ rad/s}$ and $\tau = 5 \times 10^{-13} \text{ kg m}^2$.

7.4.2 Hysteresis

Even when the nonlinearity is small, the one-to-one linear relationship between the oscillation amplitude and the separation gap is not always valid. Under certain conditions, there exists two stable oscillations with different amplitudes for a single separation gap z . This is found in Chan's experiment [106] where the orbit having a larger oscillation

Figure 7.4: The relationship between the separation gap and the plate oscillation amplitude when the nonlinearity is not severe.

amplitude is observed during the retraction of the piezo and the orbit having the lower oscillation amplitude is observed during the approaching of the piezo. This phenomenon is called "hysteresis" and it is also observed in many other MEMS applications [125]. We have simulated this phenomenon when the excitation frequency and the excitation torque are chosen to be $\omega = 17491 \text{ rad/s}$ and $\tau = 3 \times 10^{-13} \text{ kg m}^2$. The simulation result is plotted in Figure 7.5. From the simulation, it can be seen that it is not possible to estimate the separation gap z from the information of the oscillation amplitude only. For example, when the oscillation amplitude is $A = 400 \text{ } \mu\text{rad}$, there are two separation gaps corresponding to it.

Figure 7.5: Simulation of hysteresis phenomenon in the Casimir oscillator.

7.4.3 Period Doubling and Chaos

The dynamics of the plate oscillation will be even more complicated if the excitation amplitude is increased. Beyond a certain limit, period doubling will happen where the period of the plate oscillation will be twice that of the excitation period. This is a completely different phenomenon from the higher order harmonics. Usually, period doubling is considered as a transition stage to chaos [126]. Period doubling and chaos are observed in other MEMS applications [127]-[129]. Period doubling is observed in our simulation when the excitation frequency and excitation torque are chosen as $\omega = 17491 \text{ rad/s}$ and $\tau = 215 \times 10^{-13} \text{ kg m}^2$ at the separation gap: $z = 120 \text{ nm}$. We have plotted the plate oscillation trajectory in Figure 7.6 and the displacement-velocity phase plot in Figure 7.7. It can be seen that the period of the oscillation is exactly twice that of the period of the excitation signal. The chaos phenomenon is observed when the separation

Figure 7.6: The period doubling of the plate oscillation (T is the period the plate oscillation and t is the period of the excitation).

gap is changed to $z = 60 \text{ nm}$ while the excitation frequency and excitation torque are the same. The plate oscillation trajectory and the displacement-velocity phase plot are shown in Figure 7.8 and Figure 7.9 respectively. It can be seen that the plate oscillation appears to be random although it is absolutely deterministic. This makes the separation estimation even more challenging.

7.5 Gap Separation Estimation Scheme

In this section, an accurate separation estimation algorithm based on the plate oscillation trajectory and its velocity will be proposed for estimating the gap separation. The strategy of our method is to construct a cost function with respect to the estimates of the unknown parameters. The computation of this cost function must be based on

Figure 7.7: The displacement-velocity phase plot: Period doubling.

Figure 7.8: The chaos behavior of the plate oscillation.

Figure 7.9: The displacement-velocity phase plot: Chaos.

the available measurement data. The cost function is also required to have no local maximum and only one local minimum where the estimates of the unknown parameters are equal to their true values. Then, the steepest descent method, Newton's method or other conventional nonlinear optimization techniques can be used to search for the local minimum and locate the accurate estimates. In the system dynamics described in Eq. (7.8), the damping coefficient γ , moment of inertia I and fundamental frequency ω_0 can be easily measured during the free vibration. The radius of the sphere R can be measured via metrology tools. The driving torque τ and the excitation frequency ω are the control inputs which are also available for measurement. In practice, it is not easy to measure the lateral distance b as shown in Chan's work [106]. In our separation estimator, the parameter b is assumed to be unknown. The purpose of our

new separation estimation scheme is to estimate the separation gap z based on the plate oscillation $\theta(t)$ and its velocity $\dot{\theta}(t)$ from the time period $[0, T_0]$ while the parameter b is unknown. For simplicity, the system dynamics are rewritten as:

$$\ddot{\theta}(t) + 2\gamma\dot{\theta}(t) + \omega_0^2\theta(t) = \frac{\tau}{I}\sin\omega t + \frac{-B}{(z_1 - \theta(t))^3} \quad (7.10)$$

where $B = \frac{\pi^3\hbar c R}{360Ib^2}$ and $z_1 = \frac{z}{b}$. z_1 will be estimated directly in our scheme. B can be calculated by least squares method based on the estimate of z_1 . Once B and z_1 are obtained, the separation gap z and the lateral distance b can be calculated by simple substitutions. The time period $[0, T_0]$ is divided into M intervals $[t_i, t_{i+1}]$ with $t_0 = 0$ and $t_M = T_0$ based on the criterion below:

$$\theta(t_i) = c_0, \quad i = 1, 2, 3..M - 1. \quad (7.11)$$

where c_0 is a suitably defined constant. This can always be done because of the oscillatory nature of $\theta(t)$. Let $\{t_{j_i}; i = 1, 2, \dots, N\}$ be a subsequence of $\{t_i; i = 1, 2, \dots, M - 1\}$ which is chosen to satisfy:

$$\theta(t) < c_0, t \in [t_{j_{i-1}}, t_{j_i}]. \quad (7.12)$$

The cost function we propose is given by:

$$F(\hat{z}_1) = \sum_{i=1}^N [(\zeta(t_{j_{i-1}}) - \eta(t_{j_{i-1}}))\hat{\Xi}(t_{j_i}) - (\zeta(t_{j_i}) - \eta(t_{j_i}))\hat{\Xi}(t_{j_{i-1}})]. \quad (7.13)$$

ΔT ($\Delta T > 0$) is chosen to satisfy:

$$\Delta T \leq \min(t_{j_i-1} - t_{j_i-2}, t_{j_i-2} - t_{j_i-3}), \quad i = 1, 2, \dots, N. \quad (7.14)$$

$\zeta(t)$, $\eta(t)$, and $\hat{\Xi}(t)$ are computed from the available control inputs, the plate trajectory $\theta(t)$ and its velocity $\dot{\theta}(t)$:

$$\dot{\zeta}(t) + \gamma_1 \zeta(t) = \frac{\tau}{I} \sin \omega t \quad (7.15)$$

$$\dot{\eta}(t) + \gamma_1 \eta(t) = \ddot{\theta}(t) + 2\gamma \dot{\theta}(t) + \omega_0^2 \theta(t) \quad (7.16)$$

$$\dot{\hat{\Xi}}(t) + \gamma_1 \hat{\Xi}(t) = \frac{-0.5}{(\hat{z}_1 - \theta(t))^2} \quad (7.17)$$

where $\hat{z}_1$ is the estimate of the unknown parameter z_1 . To avoid the sphere from clashing into the plate, it is required that $(\hat{z}_1 - \theta(t)) > \theta_0 > 0$. The first order filters are for the purpose of reducing the transient effect. Even if different initial conditions are chosen in the estimation process, their effects will decay exponentially which will not alter the properties of the cost function. For the Casimir oscillator described in this chapter, it is suitable to assume $\theta_0 = 10^{-5} \text{ rad}$. It is proved in Appendix A that if ΔT and γ_1 are chosen to satisfy Eq. (A.10), the following relationships are valid:

$$\frac{dF(\hat{z}_1)}{d\hat{z}_1} < 0 \quad (\hat{z}_1 < z_1) \quad (7.18)$$

$$\frac{dF(\hat{z}_1)}{d\hat{z}_1} = 0 \quad (\hat{z}_1 = z_1) \quad (7.19)$$

$$\frac{dF(\hat{z}_1)}{d\hat{z}_1} > 0 \quad (\hat{z}_1 > z_1). \quad (7.20)$$

In other words, $F(\hat{z}_1)$ has no local maximum and only one local minimum at $\hat{z}_1 = z_1$. Thus, the only local minimum can be found by direct application of conventional optimization techniques like Newton's method or golden section search algorithm. Once the local minimum is found, $\hat{z}_1 = z_1$. And the other parameter B can be estimated accurately by least squares method whose objective is to minimize the following function:

$$P(\hat{B}) = \int_{\tau=0}^{T_0} (\zeta(\tau) - \eta(\tau) - \hat{B}\hat{\xi}(\tau))^2 d\tau. \quad (7.21)$$

It can be verified easily that $B\xi(t) = \zeta(t) - \eta(t)$. When $\hat{z}_1 = z_1$, $\xi(t) = \hat{\xi}(t)$. In that case, $P(\hat{B})$ has only one local minimum and no local maximum. When $P(\hat{B})$ has reached the local minimum, $\hat{B} = B$ which means the estimate of B is accurate at this point. Thus, the accurate estimate of B can be found via searching for the only local minimum of $P(\hat{B})$. This can be done by setting the first derivative of $P(\hat{B})$ with respect to $\hat{B}$ to zero which is equivalent to the following equation:

$$\hat{B} = \frac{\int_{\tau=0}^{T_0} (\zeta(\tau) - \eta(\tau))(\hat{\xi}(\tau)) d\tau}{\int_{\tau=0}^{T_0} (\hat{\xi}(\tau))^2 d\tau} \quad (7.22)$$

7.6 Simulation Results

To verify the effectiveness and robustness of our estimation algorithm, extensive simulations have been conducted and results will be discussed in this section. The estimation algorithm will be tested in all scenarios described in Section 7.4. In practice, the data measurement is subject to noise and errors. In our simulations, Gaussian noise has been

added to test the robustness of our method.

7.6.1 Small Nonlinearity

In our first simulation, the excitation frequency and torque are chosen as $\omega = 17091$ *rad/s* and $\tau = 5 \times 10^{-13}$ *kg m²* and it has already been found in Section 7.4.1 that the nonlinearity is small in this scenario. The parameter estimation is carried out when the separation gap is changing from $z = 70$ *nm* to $z = 120$ *nm* in steps of 10 *nm*. Zero initial conditions are assumed for the system dynamics and the time interval in the estimation is chosen as $[0, 0.01]$. To guarantee the convergence of our scheme, the parameters related to our algorithm are chosen as: $c_0 = 0$, $\Delta T = 1.6 \times 10^{-4}$ and $\gamma_1 = 7 \times 10^4$. As an example, the cost function with respect to the estimate $\hat{z}_1$ has been calculated and shown in Figure 7.10 when the separation gap is $z = 120$ *nm*. The true value for the unknown parameter is $z_1 = 9.16 \times 10^{-4}$. It can be seen that the cost function has only one local minimum with respect to the estimate $\hat{z}_1$ and the cost function will reach the local minimum when $\hat{z}_1 = z_1$. It can be calculated that even when the separation gap is changed, this property of the cost function is still preserved. Thus, the accuracy of the estimation will be guaranteed when Newton's method or the golden section search algorithm is applied in our estimation process. The golden section search algorithm will be used to estimate z_1 where an initial parameter range $[a_1, a_2]$ is required. By comparing the function values at two points x_1 and x_2 where $x_1 = a_1 + 0.382(a_2 - a_1)$ and $x_2 = a_1 + 0.618(a_2 - a_1)$. Either the subinterval $[a_1, x_1)$ or $(x_2, a_2]$ will be discarded to form a new and narrower range. The range will be narrowed down iteratively until

Figure 7.10: The cost function F with respect to the estimate $\hat{z}_1$ when the nonlinearity is small.

it is smaller than the prescribed tolerance. A reasonable initial parameter range will be chosen for $\hat{z}_1$ by considering the facts that there is no clashing between the sphere and the plate and the separation gap is only a few hundred nanometers. The iteration of the parameter range will be stopped when the length of the range is smaller than 4×10^{-6} and the estimate will be equal to x_1 or x_2 whichever has the smaller value in the cost function. Once the estimate of z_1 is obtained, the estimate of B can be calculated through Eq. (7.22) and we are able to estimate the separation gap by simple substitution. In practice, the measurement is subject to noise and errors. To test the robustness of our algorithm in the real environment, Gaussian noise is added to the measured data $\theta(t)$ and $\dot{\theta}(t)$ in our simulated estimation processes. The measurement error of $\theta(t)$ is less than $0.1 \mu rad$ in Chan's static experiment [105] and it will be smaller in the dynamic oscillator. Hence, the standard deviation of the noise added to $\theta(t)$ is

0.1 μrad in our simulations. The standard deviation of the noise added to the measured plate velocity is 10 $\mu rad/s$. The estimation process is carried out when the separation gap changes from 70 nm to 170 nm in steps of 10 nm . The result is summarized in Table 7.1 where we tabulate the initial parameter range $[a_1, a_2]$, numbers of iterations and the estimate of the separation gap at each step.

Table 7.1: The gap separation estimation result when the nonlinearity is small

z (nm)	Initial parameter range for $\hat{z}_1$	numbers of iterations	$\hat{z}$ (nm)
70	$[5 \times 10^{-4}, 17 \times 10^{-4}]$	12	70.2
80	$[5 \times 10^{-4}, 17 \times 10^{-4}]$	12	80.1
90	$[6 \times 10^{-4}, 18 \times 10^{-4}]$	12	89.5
100	$[7 \times 10^{-4}, 19 \times 10^{-4}]$	12	99.1
110	$[8 \times 10^{-4}, 20 \times 10^{-4}]$	12	109.8
120	$[8 \times 10^{-4}, 20 \times 10^{-4}]$	12	119.6
130	$[8 \times 10^{-4}, 20 \times 10^{-4}]$	12	129.2
140	$[10 \times 10^{-4}, 22 \times 10^{-4}]$	12	139.5
150	$[11 \times 10^{-4}, 23 \times 10^{-4}]$	12	148.5
160	$[11 \times 10^{-4}, 23 \times 10^{-4}]$	12	157.5
170	$[11 \times 10^{-4}, 23 \times 10^{-4}]$	12	169.5

7.6.2 Hysteresis

Our algorithm is tested when the hysteresis phenomenon is observed. The excitation frequency and the torque are chosen as $\omega = 17491 \text{ rad/s}$ and $\tau = 3 \times 10^{-13} \text{ kg m}^2$ which are exactly the same as the parameters in section 7.4.2. Under this condition, hysteresis is observed. In our simulation, zero initial conditions are assumed for the system dynamics and the time interval in the estimation is chosen as $[0, 0.01]$. The parameters related to our algorithm are chosen as: $c_0 = 0$, $\Delta T = 1.6 \times 10^{-4}$ and $\gamma_1 = 7 \times 10^4$. The standard deviation of the noise added to $\theta(t)$ is 0.1 μrad and the standard deviation of the noise added to the measured plate velocity is 10 $\mu rad/s$. The estimation process is carried out when the separation gap changes from 70 nm to 170

nm in steps of 10 nm . The result is summarized in Table 7.2. It can be seen that accurate estimates can be obtained from our method.

Table 7.2: The gap separation estimation result when the hysteresis is observed

z (nm)	Initial parameter range for $\hat{z}_1$	numbers of iterations	$\hat{z}$ (nm)
70	$[5 \times 10^{-4}, 17 \times 10^{-4}]$	12	70.2
80	$[5 \times 10^{-4}, 17 \times 10^{-4}]$	12	80.1
90	$[6 \times 10^{-4}, 18 \times 10^{-4}]$	12	89.8
100	$[7 \times 10^{-4}, 19 \times 10^{-4}]$	12	100.2
110	$[8 \times 10^{-4}, 20 \times 10^{-4}]$	12	109.9
120	$[8 \times 10^{-4}, 20 \times 10^{-4}]$	12	120.5
130	$[8 \times 10^{-4}, 20 \times 10^{-4}]$	12	130.7
140	$[10 \times 10^{-4}, 22 \times 10^{-4}]$	12	141.4
150	$[11 \times 10^{-4}, 23 \times 10^{-4}]$	12	150.4
160	$[11 \times 10^{-4}, 23 \times 10^{-4}]$	12	160.5
170	$[11 \times 10^{-4}, 23 \times 10^{-4}]$	12	171.7

7.6.3 Period Doubling and Chaos

The algorithm is also tested when severe nonlinearity occurs. In the simulation, the excitation frequency and the torque are chosen as $\omega = 17491 \text{ rad/s}$ and $\tau = 215 \times 10^{-13} \text{ kg m}^2$. It has been predicted that period doubling will happen when the separation gap is $z = 120 \text{ nm}$ and chaos will be observed if the separation gap is $z = 60 \text{ nm}$. Similarly, zero initial conditions are assumed for the system dynamics and the time interval in the estimation is chosen as $[0, 0.01]$. The parameters for our scheme are chosen as: $c_0 = 0$, $\Delta T = 5 \times 10^{-5}$ and $\gamma_1 = 3 \times 10^5$. The standard deviation of the noise added to $\theta(t)$ is $0.1 \text{ } \mu\text{rad}$ and the standard deviation of the noise added to the measured plate velocity is $10 \text{ } \mu\text{rad/s}$. The estimation process is carried out when the separation gap changes from 60 nm to 160 nm in steps of 10 nm . It is found from the calculation that our algorithm can give an accurate estimate of z_1 but severe errors are encountered when computing the estimate of B from Eq. (7.22). Here, we propose to calculate $\hat{z}$ from the

relationship: $\hat{z} = \hat{z}_1 \times \hat{b}$ where $\hat{b} = 130.86 \mu m$ is obtained from the previous simulation in Section 7.6.2 when the separation gap is $z = 70 nm$. The result is summarized in Table 7.3. It can be seen that we are still able to obtain accurate estimates even if period doubling and chaos happen.

Table 7.3: The gap separation estimation result when the nonlinearity is severe

z (nm)	Initial parameter range for $\hat{z}_1$	numbers of iterations	$\hat{z}$ (nm)
60	$[4.5 \times 10^{-4}, 16.5 \times 10^{-4}]$	12	60.0
70	$[5.3 \times 10^{-4}, 17 \times 10^{-4}]$	12	69.6
80	$[6.05 \times 10^{-4}, 17 \times 10^{-4}]$	12	80.3
90	$[6.8 \times 10^{-4}, 18 \times 10^{-4}]$	12	90.0
100	$[7.6 \times 10^{-4}, 19 \times 10^{-4}]$	12	100.4
110	$[8.5 \times 10^{-4}, 20 \times 10^{-4}]$	12	111.4
120	$[9.1 \times 10^{-4}, 20 \times 10^{-4}]$	12	120.0
130	$[9.85 \times 10^{-4}, 20 \times 10^{-4}]$	12	129.8
140	$[10.6 \times 10^{-4}, 22 \times 10^{-4}]$	12	139.9
150	$[11.4 \times 10^{-4}, 23 \times 10^{-4}]$	12	149.8
160	$[12.2 \times 10^{-4}, 23 \times 10^{-4}]$	12	159.8

7.7 Conclusions

To conclude, the Casimir force which is part of the atomic force between the probe and sample in AFM, is introduced and reviewed in this chapter. Various nonlinear phenomena have been simulated in the Casimir oscillator which is actuated by the Casimir force only. The complex nonlinear dynamics create difficulties for an accurate estimation of the separation gap but it yields a much higher sensitivity for the measurement. To get out of this dilemma, we have proposed an accurate estimation algorithm for estimating the separation gap in the Casimir oscillator regardless of the complex nonlinear dynamics. Theoretical proof is given to guarantee its accuracy. Simulations have been performed to verify its effectiveness. This novel estimation scheme makes the applica-

tion of the Casimir oscillator possible as a separation sensor and with high resolution. Further, the algorithm can be applied to other problems with similar mathematical modelings like the gap separation estimation in electrostatically actuated MEMS.

Chapter 8

Improved Separation Gap

Estimation for Dynamic Systems

Actuated by Casimir Force

8.1 Introduction

In the previous chapter, the atomic interaction caused by the Casimir force is discussed. The Casimir oscillator, which can also be realized by a modified AFM, can be used as a separation sensor with high sensitivity at the nanometer scale. But nonlinear dynamics like the hysteresis are present in the Casimir oscillator because the Casimir force is highly nonlinear with respect to the separation between the interacting materials. This creates the difficulty to estimate the separation gap from the measurement data. A novel algorithm is proposed in the previous chapter to estimate the separation gap accurately regardless of various complex nonlinear dynamics. It is believed that it will have important impacts in both basic researches and industry applications as the Casimir force has become important when designing nanoelectromechanical systems [130]

and understanding fundamental physics such as unification theories of four fundamental interactions [8].

In this chapter, a modified separation estimation algorithm will be proposed for the dynamic system actuated by the Casimir force. Real conditions like the finite conductivity of the interaction surfaces and the surface roughness are considered in the algorithm. The new estimation method still involve a two-step approach. Firstly, we transform the multidimensional parameter estimation problem to a one-dimensional parameter estimation problem by constructing a different cost function with respect to the estimate of a single unknown parameter. This cost function will reach its only local minimum when the estimate is equal to its true value. In the second step, we are able to update the estimate by applying conventional optimization techniques in one dimension until an accurate estimate is found. Compared to the estimation algorithm in the previous chapter, the calculation of the new cost function is easier. And only the plate trajectory measurement is needed instead of measuring both the plate trajectory and the plate velocity in the previous algorithm. In addition, the new algorithm is more robust in noisy environment. Besides the dynamic systems actuated by the Casimir force, the modified algorithm can be applied to a broad class of multidimensional parameter estimation problems in dynamic systems with distance dependent potential such as gap estimation in electrostatic force based MEMS [128] and parameter estimation in gravitational interactions [131, 132]. Our new algorithm can also be applied to parameter estimation problems in CO₂ laser dynamics [133] and bioreactor processes [115].

8.2 Problem Formulation

Our estimation method is derived for a modified atomic force microscopy (AFM) [98]. In the modified AFM, the tiny probe is replaced by a sphere coated with gold. The sphere is glued to a cantilever where a sinusoidal excitation signal is used to excite the cantilever. The deflection of the cantilever is measured from the optical detector. The sample plate is also coated with gold and mounted on a piezoelectric transducer. The schematic of the setup is shown in Figure 8.1. Sample plates with various material structures can be easily replaced in the modified AFM for different experiments. The dynamic equation of this modified AFM is given by:

$$\ddot{x}(t) + 2\gamma\dot{x}(t) + \frac{k}{m}x(t) = \frac{I}{m}\sin\omega t - \frac{R}{m}F_c(z + x(t)) \quad (8.1)$$

where $x(t)$ is the instantaneous position of the sphere, m is the mass of the sphere, γ is the system damping, k is the spring constant of the cantilever, I is the amplitude of the excitation signal, ω is the excitation frequency, R is the sphere radius, $F_c(z + x(t))$ denotes the Casimir force between the sphere and the plate, and z is the separation gap when the sphere is in equilibrium. When the surfaces are smooth and perfectly reflecting at all frequencies, the model of the Casimir force is given by:

$$F_c(z + x(t)) = F(z + x(t)) = \frac{\pi^3\hbar c}{360(z + x(t))^3}. \quad (8.2)$$

where c is the light speed, and $\hbar$ is the reduced Planck constant. In practice, the finite conductivity of the real metal coated on the sphere surface and the sample plate will

reduce the strength of the Casimir force as compared to an ideal metal with a perfect reflectivity. And the surface roughness should also be taken into account. To include the modification of finite conductivity and surface roughness to the Casimir force, the modified Casimir force is given as:

$$F_c(z + x(t)) = F(z + x(t))r_{fc}(z + x(t))r_s(z + x(t)) \quad (8.3)$$

where $r_{fc}(z + x(t))$ is the correction factor to the ideal Casimir force model due to the finite conductivity, and $r_s(z + x(t))$ is the correction factor to the ideal Casimir force due to the surface roughness. It is shown that [134] the following fourth order approximation of the conductivity correction agrees well with the finite conductivity calculation from Liftshitz theory [119]:

$$\begin{aligned} r_{fc}(z + x(t)) = & \left[1 - 4\frac{\delta_0}{z + x(t)} + \frac{72}{5}\left(\frac{\delta_0}{z + x(t)}\right)^2 - \frac{320}{7}\left(1 - \frac{\pi^2}{210}\right)\left(\frac{\delta_0}{z + x(t)}\right)^3 \right. \\ & \left. + \frac{400}{3}\left(1 - \frac{163\pi^2}{7350}\right)\left(\frac{\delta_0}{z + x(t)}\right)^4 \right] \end{aligned} \quad (8.4)$$

where $\delta_0 = l_p/2\pi$ is the effective penetration depth of the electromagnetic zero-point oscillation into the metal, and l_p is the plasma wavelength which depends on the material of the sphere and the plate. It should be pointed out that this fourth order approximation is computed at zero temperature. At room temperature (for example, $T = 300 \text{ K}$), the correction factor deviates from this fourth order approximation. However, this deviation is insignificant at smaller separation [135, 136] when the Casimir force is strong, so the thermal correction will not be considered although our method can also be applied

successfully when thermal correction is included. The correction factor due to surface roughness is calculated as:

$$r_s(z) = [1 + 6(\frac{A_r}{z + x(t)})^2 + 15(\frac{A_r}{z + x(t)})^4] \quad (8.5)$$

where A_r is the roughness amplitude. In the system, $x(t)$ and the excitation signal are available for measurement. System parameters k , γ , and m can be calculated from the frequency response test during free vibration. The roughness amplitude of the sample A_r can be measured by AFM and l_p can be found in [137] with respect to different materials coated on the sphere and the sample. Although the sphere radius R can be measured by metrology tools, it will be demonstrated that the separation gap z can be estimated accurately by our method without measuring R . z is unknown in our setup and it is the parameter we want to estimate. The Casimir oscillator will function properly when the sphere and the plate do not clash.

8.3 Our Estimation Method

The objective of our estimation is to estimate the unknown parameters z and R from the sphere's vibration trajectory $x(t)$ during the time interval $[t_1, t_2]$. Similar to Chapter 7, our strategy is to construct a cost function with respect to the estimates of the unknown parameter z . The cost function is designed to have no local maximum and only one local minimum where the estimates of the unknown parameters are equal to their true values. Then, the steepest descent method, Newton's method or other conventional nonlinear

optimization techniques can be used to search for the local minimum and locate the accurate estimate. Three linear filters are required in our estimation process:

$$\ddot{\xi}(t) + 2\gamma_1\dot{\xi}(t) + \frac{k}{m}\xi(t) = \frac{F_c(\hat{z} + x(t))}{m} \quad (8.6)$$

$$\ddot{\phi}(t) + 2\gamma_1\dot{\phi}(t) + \frac{k}{m}\phi(t) = \ddot{x}(t) + 2\gamma\dot{x}(t) + \omega_0^2x(t) \quad (8.7)$$

$$\ddot{\psi}(t) + 2\gamma_1\dot{\psi}(t) + \frac{k}{m}\psi(t) = \frac{I}{m}\sin\omega t \quad (8.8)$$

where $\hat{z}$ is our estimate of the unknown parameter z . In Eq. (8.6), the force input to the second order system is the estimated Casimir force where the unknown separation gap z is replaced by the estimate $\hat{z}$ in the Casimir force model. The filters chosen here is over-damped which means the transients due to different initial conditions will decay exponentially. Thus, the effect of initial condition will be eliminated quickly. Since the signals chosen for the estimation process consist of thousands of cycles, the computation errors due to the initial condition will not affect our estimation results. Similar filters are proposed in Sahoo's work [138] with the purpose of detecting a change in the sample topography or the interaction force. Since there is no collision between the sphere and the plate, $\hat{z}$ must satisfy: $(\hat{z} + x(t)) > 0$. γ_1 is chosen to satisfy: $\gamma_1 > \sqrt{k/m}$. We denote that $y(t) = \psi(t) - \phi(t)$. The cost function with respect to the single estimate $\hat{z}_1$ in the time interval $[t_1, t_2]$ is constructed as:

$$G(\hat{z}) = \int_{t_1}^{t_2} y^2(t)dt - \frac{(\int_{t_1}^{t_2} y(t)\xi(t)dt)^2}{\int_{t_1}^{t_2} \xi^2(t)dt}. \quad (8.9)$$

It is proved in Appendix B that the cost function has the following properties under

Figure 8.1: The schematic of the Casimir force actuated system.

normal operation conditions:

$$\frac{dG(\hat{z})}{d\hat{z}} < 0 \quad (\hat{z} < z) \quad (8.10)$$

$$\frac{dG(\hat{z})}{d\hat{z}} = 0 \quad (\hat{z} = z) \quad (8.11)$$

$$\frac{dG(\hat{z})}{d\hat{z}} > 0 \quad (\hat{z} > z). \quad (8.12)$$

Thus, the cost function $G(\hat{z})$ has only one local minimum with respect to $\hat{z}$, the estimate of the separation gap, and it will reach its local minimum when $\hat{z} = z$. Once $\hat{z}_1 = z_1$, $\hat{R} = R$ is valid if $\hat{R}$ is calculated from the least-squares algorithm: $\hat{R} = \int_{t_1}^{t_2} \xi(t)y(t)dt / \int_{t_1}^{t_2} \xi^2(t)dt$. The motivation to choose such a cost function is that it is first found that the cost function has a global minimum when $y(t) = \xi(t)$. By a few trials, $y(t)$ and $\xi(t)$ are constructed such that $y(t) = \xi(t)$ is only valid when $z = \hat{z}$ and the cost function has only one local minimum and no local maximum.

8.4 Simulation Results

Here, we shall conduct numerical simulations to verify the accuracy of our method. First of all, we test the accuracy and robustness of our method when the Casimir force is in its ideal form (Eq. (8.2)). Then our method is tested when finite conductivity and surface roughness are considered in the Casimir force model (Eq. (8.3)). In the second scenario, the robustness of our method to parameter variations in finite conductivity and surface roughness are also tested. In all our simulations, system parameters are chosen to be similar to the real experiment data in [98, 139]: $k = 0.1 \text{ N/m}$, $m = 7.0167 \times 10^{-10} \text{ kg}$, $\gamma = 100/s$, and $R = 100 \text{ }\mu\text{m}$. The time interval in the estimation process is chosen as $[t_1, t_2] = [0.07, 0.1]$ and γ_1 in our constructed filters (Eq. (8.6) - Eq. (8.8)) is chosen as $\gamma_1 = 20000/s$. The initial conditions chosen for the system dynamics (Eq. (8.1)) are $[0, 0]$ while the initial conditions chosen in our estimators are $[100\text{nm}, 0]$. This is to test whether our method is affected by different initial conditions chosen.

8.4.1 Ideal Casimir Force Model

First of all, simulations are conducted by assuming perfectly smooth and reflecting surfaces. In this case, the Casimir force is described by the ideal model in Eq. (8.2). Under normal operation conditions, the probe's vibration is found to be a pure sinusoidal [106] which can be modelled as: $x(t) = d_0 + A\sin(\omega t + \phi)$ where A is the probe vibration amplitude, d is the offset and ϕ is the phase difference. Hysteresis is found when the excitation amplitude and frequency are chosen as $I = 0.1 \text{ nN}$ and $\omega = 11000 \text{ rad/s}$. When the separation gap z is in the range of 105 nm to 115 nm , there co-exists two

stable orbits with different amplitudes with respect to a single separation gap z (Figure 8.2a). Depending on the history of the probe position, the probe will enter into one of the orbits. Hysteresis is a typical nonlinear phenomenon which is observed in many other physical systems [140]-[142]. The hysteresis phenomenon can be eliminated by reducing the excitation amplitude to $I = 0.01 \text{ nN}$ but the sensitivity of the measurement is greatly reduced (Figure 8.2b). In both scenarios, it is difficult to estimate the separation gap z based on the vibration amplitude A due to the inherent nonlinearity. To demonstrate the effectiveness of our estimation method, we calculate and plot the cost function $G(\hat{z})$ with respect to different estimates $\hat{z}$ at the separation gap $z = 120 \text{ nm}$ when $I = 0.1 \text{ nN}$ and $\omega = 11000 \text{ rad/s}$ (Figure 8.3a). It is shown that $G(\hat{z})$ has only one local minimum with respect to $\hat{z}$ when $\hat{z} = z = 120 \text{ nm}$. In this case, an accurate estimate of the separation gap can be obtained by applying optimization techniques like Newton's method and golden search algorithm to search for the local minimum of $G(\hat{z})$. In practice, noise exists in the measurement. In our next simulation, the measurement of the probe position $x(t)$ is subject to Gaussian noise where the signal to noise ratio (SNR) is 20 dB. All other conditions remain the same. The probe trajectory when the measurement is subject to noise is plotted in Figure 8.4. The noise added in our simulation is much bigger than the noise observed in AFM experiments when the probe is excited by a sinusoidal signal [143]. But even though the measurement is under such severe noisy conditions, the cost function $G(\hat{z})$ will still reach its only local minimum when $\hat{z}$ is very close to z (Figure 8.3b). Thus, our estimation method is robust to noise in the real environment.

Figure 8.2: (a) The relationship between vibration amplitude A and separation gap z computed by assuming the ideal Casimir force model when $I = 0.1 \text{ nN}$ and $\omega = 11000 \text{ rad/s}$. (b) The relationship between the vibration amplitude A and the separation gap z computed by assuming the ideal Casimir force model when $I = 0.01 \text{ nN}$ and $\omega = 11000 \text{ rad/s}$.

In this subsection, we will demonstrate the estimation process when z changes from 85 nm to 140 nm by using the golden search algorithm. The probe trajectory measurement is still subject to Gaussian noise ($\text{SNR} = 20 \text{ dB}$). It is shown that estimation results are accurate (Figure 8.5) for both scenarios observed in Figure 8.2.

Figure 8.3: (a) The cost function $G(\hat{z})$ with respect to $\hat{z}$ by assuming the ideal Casimir force model when $I = 0.1$ nN, $\omega = 11000$ rad/s, $z = 120$ nm, and the measurement is noise free. (b) The cost function $G(\hat{z})$ with respect to $\hat{z}$ by assuming the ideal Casimir force model when $I = 0.1$ nN, $\omega = 11000$ rad/s, $z = 120$ nm, and the measurement is subject to Gaussian noise (SNR = 20dB).

8.4.2 Modified Casimir Force Model

In this subsection, the accuracy and robustness of our algorithm are tested by considering finite conductivity and surface roughness in the Casimir force model described by Eq. (8.3). Since both the sphere and the sample are coated with gold, the plasma wavelength for gold is $l_p = 132$ nm, and the surface roughness is assumed to be $A_r = 10$ nm. When the excitation amplitude and frequency are chosen as $I = 0.1$ nN and $\omega = 11000$ rad/s, the probe vibration is still a pure sinusoidal. Hysteresis is still found when the separation gap z is in the range of 100 nm to 120 nm (Figure 8.6a). For an easy comparison with

Figure 8.4: The trajectory of the probe when it is subject to Gaussian noise ($\text{SNR} = 20\text{dB}$). It is computed by assuming the ideal Casimir force model when $I = 0.1 \text{ nN}$, $\omega = 11000 \text{ rad/s}$, and $z = 120 \text{ nm}$.

Figure 8.5: The estimation results when $\omega = 11000 \text{ rad/s}$ and the probe trajectory measurement is subject to Gaussian noise ($\text{SNR} = 20\text{dB}$) by assuming the ideal Casimir force model. If $\hat{z} = z$, the data will lie on the blue line.

simulations in the previous subsection, simulations have been done with the excitation amplitude reduced to $I = 0.01 \text{ nN}$. It can be seen that hysteresis is not observed and the sensitivity of the measurement is greatly reduced (Figure 8.6b). These phenomena are similar with the phenomena observed in the previous subsection. To demonstrate the effectiveness of our estimation method when finite conductivity and surface roughness are considered in the Casimir force model, we calculate and plot the cost function $G(\hat{z})$ with respect to different estimates $\hat{z}$ at the separation gap $z = 120 \text{ nm}$ when $I = 0.1$

nN and $\omega = 11000 \text{ rad/s}$ (Figure 8.7a). As expected, $G(\hat{z})$ has only one local minimum with respect to $\hat{z}$ when $\hat{z} = z = 120 \text{ nm}$. Thus, an accurate estimate of the separation gap can be obtained by applying conventional optimization techniques mentioned above to search for the local minimum of $G(\hat{z})$. In our next simulation, the measurement of the probe position $x(t)$ is subject to Gaussian noise where the SNR is still 20 dB. The cost function $G(\hat{z})$ will still reach its only local minimum when $\hat{z}$ is very close to z (Figure 8.7b). The estimation process is computed again when z changes from 85 nm to 140 nm by using golden search algorithm and the probe trajectory measurement is subject to Gaussian noise (SNR = 20 dB). It is shown that estimation results are accurate (Figure 8.8a).

We have also tested the robustness of our method against parameter variations. $A_r = 10 \text{ nm}$ and $l_p = 132 \text{ nm}$ are assumed in our estimation filters (Eq. (8.6) - Eq. (8.8)), but we examine the estimation results when the true surface roughness of the sample is actually $A_r = 12 \text{ nm}$ in the system equation of Eq. (8.1) which is 20% higher. We also examine the estimation results when both the true surface roughness and the true plasma wavelength ($A_r = 12 \text{ nm}$ and $l_p = 120 \text{ nm}$) differ significantly from the assumed values in the system equation. From Figure 8.8b, it is clear that accurate estimates can still be obtained in both cases. This guarantees the success of our method even when errors exist in measuring the surface roughness of the sample and the finite conductivity of the coated metal in practice.

Figure 8.6: (a) The relationship between the vibration amplitude A and the separation gap z computed by assuming modified Casimir force model when $I = 0.1 \text{ nN}$ and $\omega = 11000 \text{ rad/s}$. (b) The relationship between the vibration amplitude A and the separation gap z computed by assuming modified Casimir force model when $I = 0.01 \text{ nN}$ and $\omega = 11000 \text{ rad/s}$.

8.5 Conclusions

In conclusion, we have proposed an improved estimation method to estimate the unknown separation gap in dynamic systems actuated by the Casimir force. It is shown in various simulations that our method is accurate and more robust to noise when an ideal Casimir force model is considered. Our method is also shown to be effective when the finite conductivity of real materials and surface roughness are considered in the Casimir force model. In addition, our estimation method is shown to be robust against the variations in initial conditions chosen, parameter variations in finite conductivity and

Figure 8.7: (a) The cost function $G(\hat{z})$ with respect to $\hat{z}$ by considering the modified Casimir force model when $I = 0.1 \text{ nN}$, $\omega = 11000 \text{ rad/s}$, $z = 120 \text{ nm}$, and the measurement is noise free. (b) The cost function $G(\hat{z})$ with respect to $\hat{z}$ by considering the modified Casimir force model when $I = 0.1 \text{ nN}$, $\omega = 11000 \text{ rad/s}$, $z = 120 \text{ nm}$, and the measurement is subject to Gaussian noise (SNR = 20dB).

surface roughness. This guarantees its success in practice where measurement errors do exist in the finite conductivity and the surface roughness measurements. More generally, besides the separation gap estimation in dynamic systems actuated by Casimir force, our method can also be applied successfully to other nonlinear dynamic systems when unknown parameters are nonlinearly parameterized in a similar way to the Casimir force model of Eq. (8.3). Accordingly, the right hand side of our estimation filter in Eq. (8.6) will be changed to the estimate of the new nonlinear inputs where the single unknown parameter is replaced by its estimate.

Figure 8.8: (a) The estimation results when $\omega = 11000$ rad/s and the probe trajectory measurement is subject to Gaussian noise ($SNR = 20dB$). If $\hat{z} = z$, the data will lie on the blue line. (b) The estimation results when true values of A_r and l_p differ with the values assumed in our method ($A_r = 10$ nm and $l_p = 132$ nm are assumed in our method). The probe trajectory measurement is subject to Gaussian noise ($SNR = 20dB$). If $\hat{z} = z$, the data will lie on the blue line.

Chapter 9

Conclusions and Future Work

9.1 Conclusions

In this thesis, several new research results on improving the measurement accuracy and resolution of the AFM and the Casimir oscillator which is a modified AFM system have been presented. As AFM is a precise measurement and manipulation tool in the nanometer scale which is crucial in many research areas and industry applications, it is highly desirable to improve its measurement accuracy and resolution.

Essentially, the improvement of measurement accuracy of the AFM sensing system is achieved through the understanding of the inaccurate measurement of the AFM sensing system which is mainly due to the measurement inaccuracy of the PSD and the AFM sensing principle. A set of linearity indices has been proposed to evaluate the measurement accuracy of various PSDs in a simple and quantitative way. Based on the linearity indices, a new estimation formula is proposed to give a nearly perfect measurement accuracy of the 2-D tetra-lateral PSD. The percentage of the effective area with less than 1% measurement errors has been enlarged by 6 times. As the 2-D tetra-lateral PSD has many advantages in terms of fast response and low noise, it is believed that the mea-

surement accuracy of the AFM sensing system will be improved if the 2-D tetra-lateral PSD with our proposed formula is implemented. More than that, other applications with stringent requirement on the measurement linearity of the 2-D tetra-lateral PSD will also benefit from our work. We have also investigated the effect of large beam size on our new formula and it is found that the measurement accuracy improvement of the proposed 2-D tetra-lateral PSD is not affected even if the light spot has a considerable beam size. This is very helpful in practice because it is difficult to control the beam size of the light spot projected on the PSD. As the sensing principle also contributes to the measurement inaccuracy, we specifically examined the total measurement errors arising from both the PSD and the sensing principle and they are effectively reduced by modifying our estimation algorithm. Besides the 2-D tetra-lateral PSD, the concept of linearity indices can be readily applied to many other types of PSDs. As an example, it is applied to improve the measurement accuracy of the QD PSD and the percentage of the effective area with less than 1% measurement errors has been enlarged by 2 to 3 times.

Besides the AFM sensing system, measurement errors can also arise from the imprecise probe-sample separation estimation. Indeed, inaccurate estimates will occur under practical conditions due to the complexity of nonlinear dynamics. We proposed a novel algorithm that is based on the information of the main sinusoidal and its harmonics, from which accurate estimates of the probe-sample separation can be obtained regardless of the nonlinear dynamics. And it is shown to be robust under severe noisy environment.

The AFM system is also modified to characterize different atomic forces. Casimir

force is one of the atomic forces acting between the probe and the sample and the Casimir oscillator is actuated by the Casimir force. Compared to AFM, a sphere instead of a tiny probe is used to interact with the sample in the Casimir oscillator. Unknown parameters like the separation gap are nonlinearly parameterized in this dynamic system. The nonlinear dynamics are predicted and observed which create significant difficulties in obtaining an accurate estimate of the gap separation and there is no guarantee from conventional methods to give accurate estimations. An innovative algorithm for the separation gap estimation is proposed and it involves a two-step process. It first transforms the multi-dimensional nonlinear parameter estimation problem into a one-dimensional problem. A cost function with only one local minimum is constructed. Then, conventional searching methods like Newton's method can be applied to search for the true value of the unknown parameter. To achieve a simpler and more robust estimation, we also proposed a modified algorithm based on a similar philosophy. This modified algorithm includes the real experimental conditions like the finite conductivity of the interacting materials and the surface roughness.

It is our belief that the measurement accuracy and resolution of the AFM and the Casimir oscillator can be significantly improved by the systematic work developed in this thesis. Many other research areas and applications can also benefit from our work. For example, various PSD applications in the area of precision measurement will benefit from the results on linearity improvement of the PSDs.

9.2 Future Work

Although our efforts towards performance improvement of the AFM and the Casimir oscillator have addressed several important issues and several significant results have been derived, they are still a lot of important and interesting research issues for future work. Several of them will be discussed in this section.

The first task in the near future is to carry out experiments to test the effectiveness of our algorithms aiming to improve the measurement linearity of AFM sensing system. And it is important to compare the imaging performances between AFM sensing systems employing QD PSD and 2-D tetra-lateral PSD. There are some preparations that need to be done before the experiment. One QD PSD is deployed in the existing AFM in our lab. To implement 2-D tetra-lateral PSD in our AFM, some circuits need to be modified as the bias voltage to power up the PSD is different for different PSDs. In addition, the calculation formulae are implemented in analogue circuits in our AFM. This part of the circuit needs to be redesigned if 2-D tetra-lateral PSD is implemented. All these changes must be compatible with other existing AFM components. It is believed that this comparison study will be a very useful reference for AFM system manufacturers to decide which PSD is more appropriate for AFM.

The second task in the near future is to implement the algorithms proposed to estimate the probe-sample separation in AFM and the gap separation in the Casimir oscillator. Although the accuracy and the robustness are tested in a great number of simulations where severe measurement noise is added, it is still of practical interest to test and validate them in a real implementation. This will require substantial investment

in time and resources to set up appropriate experimental setups.

Although the van der Waals' force is the dominant force between the probe and the sample in a charge-free vacuum environment. Other forces can be significant under certain conditions. For example, the capillary force exists in the humidity environment [18] and it is commonly detected when a biological sample is studied under the AFM. And various hydrodynamics of the cantilever are measured when the sample is a liquid [144]. Even in the vacuum environment, other forces like the adhesion force hysteresis and the velocity dependent viscous damping force are predicted and measured in some scenarios [145]-[146]. The existence of these forces will affect the AFM measurements and degrade its performance. Thus, it is of both theoretical interest and practical importance to study these forces and to find suitable solutions towards a more accurate AFM measurements when these forces exist. The results and techniques reported in this thesis could be useful when investigating these forces.

Currently, the AFM is used to manipulate small particles like single molecules and it is believed to be a dominating manipulator in the nano-world. To be a good manipulator, and to yield fast response, accurate positioning and force control are required. These can be achieved by exploiting advanced control schemes where accurate estimate of unknown force-related parameters is a prerequisite. In this context, our proposed estimation algorithms could be valuable. In the future, exploration in this research direction will be carried out.

Author's Journal Publications

1. Cui, S. and Soh, Y.C., "The effect of spot size on linearity improvement of tetra-lateral position sensitive detector," *Optical and Quantum Electronics*, DOI:10.1007/s11082-011-9462-y (Accepted).
2. Cui, S. and Soh, Y.C., "Separation gap estimation in dynamic systems actuated by Casimir force," *Advanced Materials Research*, vol. 254, pp. 21-24, 2011.
3. Cui, S. and Soh, Y.C., "Analysis and improvement of Laguerre-Gaussian beam position estimation using quadrant detectors," *Optics Letters*, vol. 36, no. 9, pp. 1692 -1694, 2011.
4. Cui, S. and Soh, Y.C., "Highly precise conductivity corrections to the Casimir force via exponential basis functions," *Physical Review A*, vol. 82, no. 6, 062510, 2010.
5. Cui, S. and Soh, Y.C., "An accurate separation estimation algorithm for the Casimir oscillator," *Journal of Microelectromechanical Systems*, vol. 19, no. 5, pp. 1153-1161, 2010.
6. Cui, S. and Soh, Y.C., "Linearity indices and linearity improvement of 2-D tetra-

- lateral position sensitive detector,” *IEEE Transactions on Electron Devices*, vol. 57, no. 9, pp. 2310-2316, 2010.
7. Cui, S. and Soh, Y.C., “A novel probe-sample separation estimation scheme for atomic force microscopy,” *Ultramicroscopy*, vol. 110, no. 6, pp. 622-625, 2010.
 8. Cui, S. and Soh, Y.C., “Improved measurement accuracy of the quadrant detector through improvement of linearity index,” *Applied Physics Letters*, vol. 96, no. 8, 081102, 2010.
 9. Cui, S., Zhang, Y., Lim, S.Y., Soh, Y.C., “Improved accuracy of 2-D non-contact range sensors using 2-D lateral effect position sensitive detector,” *Key Engineering Materials*, vol. 381-382, p333-336, 2008.

Bibliography

- [1] C. J. Chen, *Introduction to Scanning Tunneling Microscopy*. Oxford University Press, 1993.
- [2] G. Binnig *et al.*, “Atomic force microscope,” *Phys. Rev. Lett.*, vol. 56, pp. 930-933, 1986.
- [3] F. J. Giessibl, “Atomic resolution of the silicon (111) – (7x7) surface by atomic force microscopy,” *Science*, Vol. 267, pp. 68-71, 1995.
- [4] R. D. Piner *et al.*, “‘Dip-Pen’ Nanolithography,” *Science*, vol. 283, pp. 661-663, 1999.
- [5] F. J. Giessibl, “High-speed force sensor for force microscopy and profilometry utilizing a quartz tuning fork,” *App. Phys. Lett.*, vol. 73, no. 26, pp. 3956-3958, 1998.
- [6] M. Marth *et al.*, “A unifying view on some experimental effects in tapping-mode atomic force microscopy,” *J. Appl. Phys.*, vol. 85, no. 10, pp. 7030-7036, 1999.
- [7] E. Buks and M. L. Roukes, “Casimir force changes sign,” *Nature*, vol. 419, pp. 119-120, 2002.

- [8] R. S. Decca *et al.*, “Constraining new forces in the Casimir regimes using the isoelectronic technique,” *Phys. Rev. Lett.*, vol. 94, 240401, 2005.
- [9] H. B. Chan *et al.*, “Nonlinear micromechanical Casimir oscillator,” *Phys. Rev. Lett.*, vol. 87, 211801, 2001.
- [10] M. Lu and G. Fedder, “Position control of parallel-plate microactuators for probe-based data storage,” *J. Microelectromech. Syst.*, vol. 13, no. 5, pp. 759-769, Oct. 2004.
- [11] H. Sadeghian, G. Rezazadeh and P. M. Osterberg, “Application of the generalized differential quadrature method to the study of pull-in phenomena of MEMS switches,” *J. Microelectromech. Syst.*, vol. 16, no. 6, pp. 1334-1340, Dec. 2007.
- [12] D. Lee and O. Solgaard, “Pull-In Analysis of Torsional Scanners Actuated by Electrostatic Vertical Combedrives,” *J. Microelectromech. Syst.*, vol. 17, no. 5, pp. 1228-1238, Oct. 2008.
- [13] O. H. Willemsen *et al.*, “Reconstruction of the tip-surface interaction potential by analysis of the Brownian motion of an atomic force microscope tip,” *Langmuir*, vol. 16, pp. 4339-4347, 2000.
- [14] J. N. Israelachvili *et al.*, *Intermolecular and Surface Forces*. Academic Press, 1985.
- [15] P. M. Hoffmann, “Dynamics of small amplitude, off-resonance AFM,” *Appl. Surf. Sci.*, vol. 210, pp. 140-145, 2003.

- [16] L. Wang, “Analytical descriptions of the tapping-mode atomic force microscopy response,” *Appl. Phys. Lett.*, vol. 73, no. 25, pp. 3781-3783, 1998.
- [17] R. Garcia *et al.*, “Dynamics of a vibrating tip near or in intermittent contact with a surface,” *Phys. Rev. B*, vol. 61, no. 20, pp. 13381-13384, 2000.
- [18] L. Zitzler *et al.*, “Capillary forces in tapping mode atomic force microscopy,” *Phys. Rev. B*, vol. 66, no. 15, 155436, 2002.
- [19] K. Fa, A. V. Nguyen, J. D. Milled, “Hydrophobic attraction as revealed by AFM force measurements and molecular dynamics simulation,” *J. Phys. Chem. B*, vol. 109, no. 27, pp. 13112-13118, 2005.
- [20] N. A. Burnham, R. J. Colton, H. M. Pollock, “Work-function anisotropies as an origin of long-range surface forces,” *Phys. Rev. Lett.*, vol. 69, no. 1, pp. 144-147, 1992.
- [21] F. J. Giessible, “Advances in atomic force microscopy,” *Rev. Mod. Phys.*, vol. 75, no. 3, pp. 949-983, 2003.
- [22] B. Cappella *et al.*, “Force-distance curves by AFM,” *IEEE Eng. Med. Biol.*, vol. 16, no. 2, pp. 58-65, 1997.
- [23] N. A. Burnham, R. J. Colton, H. M. Pollock, “Interpretation of force curves in force microscopy,” *Nanotechnology*, vol. 4, no. 2, pp. 64-80, 1993.

- [24] F. J. Giessibl and G. Binnig, “True atomic resolution on KBr with a low-temperature atomic force microscope in ultrahigh vacuum,” *Ultramicroscopy*, vol. 42-44, pp. 281-286, 1992.
- [25] D. F. Ogletree, R. W. Carpick and M. Salmeron, “Calibration of frictional forces in atomic force microscopy,” *Rev. Sci. Instrum.*, vol. 67, no. 9, pp. 3298-3306, 1996.
- [26] Q. Zhong *et al.*, “Fractured polymer/silica fiber surface studied by tapping mode atomic force microscopy,” *Surf. Sci.*, vol. 290, L688-692, 1993.
- [27] J. Tamayo and R. Garcia, “Effects of elastic and inelastic interaction on phase contrast images in tapping-mode scanning force microscopy,” *Appl. Phys. Lett.* vol. 71, pp. 2394-2396, 1997.
- [28] M. Ashhab *et al.*, “Dynamical analysis and control of microcantilevers,” *Automatica*, vol. 35, no. 10, pp. 1663-1670, 1999.
- [29] R. W. Stark and W. M. Heckl, “Fourier transformed atomic force microscopy: tapping mode atomic force microscopy beyond the Hookian approximation,” *Surf. Sci.*, vol. 457, pp. 219-228, 2000.
- [30] R. Garcia and A. San Paulo, “Attractive and repulsive tip-sample interaction regimes in tapping-mode atomic force microscopy,” *Phys. Rev. B*, vol. 60, no. 7, pp. 4961-4967, 1999.
- [31] M. Lee and W. Jhe, “General Theory of Amplitude-Modulation Atomic Force Microscopy,” *Phys. Rev. Lett.*, vol. 97, 036104, 2006.

- [32] D. Ramos, J. Tamayo, J. Mertens and M. Calleja, “Photothermal excitation of microcantilevers in liquids,” *J. Appl. Phys.*, vol. 99, 124904, 2006.
- [33] S. Nishida, D. Kobayashi, T. Sakurada, T. Nakazawa, Y. Hoshi, H. Kawakatsu, “Photothermal excitation and laser doppler velocimetry of higher cantilever vibration modes for dynamic atomic force microscopy in liquid,” *Rev. Sci. Instrum.*, vol. 79, 123703, 2008.
- [34] A. V. Marquina and A. T. Jacome, “Silicon quadrant detector in CMOS technology,” *Proc. SPIE*, vol. 5622, pp. 922-925, 2004.
- [35] A. Mäkynen, J. Kostamovaara and R. Myllyla, “Positioning resolution of the position-sensitive detectors in high background illumination,” *IEEE Trans. Instrum. Meas.*, vol. 45, 1, pp. 324-326, 1997.
- [36] N. Jalili and K. Laxminarayana, “A review of atomic force microscopy imaging systems: application to molecular metrology and biological sciences,” *Mechatronics*, vol. 14, pp. 907-945, 2004.
- [37] J. T. Wallmark, “A new semiconductor photocell using lateral photoeffect,” *Proc. IRE*, vol. 45, pp. 474-483, Apr. 1957.
- [38] W. Schottky, “Ueber den Entstehungsort der Photoelektronen in Kupfer-Kupferoxydulphotozellen,” *Phys. Z.*, vol. 31, pp. 913-925, 1930.
- [39] W. P. Connors, “Lateral photodetector operating in the fully reverse-biased mode,” *IEEE Trans. Electron Devices*, vol. ED-18, pp. 591-596, 1971.

- [40] H. J. Woltring, "Single- and dual-axis lateral photodetectors of rectangular shape," *IEEE Trans. Electron Devices*, vol. ED-22, pp. 581-586, 1975.
- [41] Y. Y. Cha and D. G. Gweon, "A calibration and range-data extraction algorithm for an omnidirectional laser range finder of a free-ranging mobile robots," *Mechatronics*, vol. 6, no. 6, pp. 665-689, 1996.
- [42] H. Hoff and J. Bargon, "A vibrating optical fiber tip as a new sensor to monitor organic vapors," *Sensor. Actuat. B*, 67, pp. 29-35, 2000.
- [43] G. Meyer and N. M. Amer, "Novel optical approach to atomic force microscopy," *Appl. Phys. Lett.*, vol. 53, no. 12, pp. 1045-1047, 1988.
- [44] W. Wang and I. J. Busch-Vishniac, "The linearity and sensitivity of lateral effect position sensitive devices-An improved geometry," *IEEE Trans. Electron Devices*, vol. ED-36, pp. 2475-2480, 1989.
- [45] E. Silva and K. Vliet, "Robust approach to maximize the range and accuracy of force application in atomic force microscopes with nonlinear position-sensitive detectors," *Nanotechnology*, 17, pp. 5525-5530, 2006.
- [46] *Position-Sensitive Detectors*, Hamamatsu Photonics K.K., Solid State Division, Product Catalog, 1987
- [47] W. Wang, "Clover design lateral effect position-sensitive device," U.S. Patent 4 887 140, December 12, 1989.

- [48] J. W. Noorlag, “Quantatative analysis of effects causing nonlinear position response in position-sensitive photodetectors,” *IEEE Trans. Electron Devices*, vol. ED-29, pp. 158-161, 1982.
- [49] X. Wang and M. Ye, “Modeling and non-linear correction of two-dimension photoelectric position sensitive detector,” *Proc. SPIE*, vol. 4919, pp. 452-460, 2002.
- [50] M. Huang *et al.*, “Development of a new signal processor for tetralateral position sensitive detector based on single-chip microcomputer,” *Rev. Sci. Instrum.*, 77, 083301, 2006.
- [51] M. Bruno *et al.*, “Position determination and resolution of two-dimensional position-sensitive solid-state detectors,” *Nucl. Instr. and Meth. A*, vol. 311, pp. 189-194, 1992.
- [52] R. L. Cowin and D. L. Watson, “Surface charge flow in two-dimensional position sensitive silicon detectors,” *Nucl. Instr. and Meth. A*, vol. 399, pp. 365-381, 1997.
- [53] G. Lucovsky, “Photoeffects in nonuniformly irradiated p-n junctions,” *J. Appl. Phys.*, vol. 31, no. 6, pp. 1088-1095, June 1960.
- [54] G. Petersson and L. Linholm, “Position sensitive light detectors with high linearity,” *IEEE J. Solid-State Circuits*, vol. SC-13, pp. 392-399, June 1978.
- [55] C. Han, H. Lee and C. C. Chung, “Automatic approaching method for atomic force microscope using a Gaussian laser beam,” *Rev. Sci. Instrum.*, vol. 80, no. 7, 073705, 2009.

- [56] E. J. Lee, Y. Park, C. S. Kim and T. Kouh, "Detection sensitivity of the optical beam deflection method characterized with the optical spot size on the detector," *Curr. Appl. Phys.*, vol. 10, pp. 834-837, 2010.
- [57] R. Haberman, *Elementary Applied Partial Differential Equations: With Fourier Series and Boundary Value Problems*. Prentice-Hall, 1987.
- [58] A. P. Prudnikov, Y. A. Brychkov, O. I. Marichev, *Integrals and Series, Volume 2: Special Functions*. Gordon and Breach Science Publishers, 1986.
- [59] F. Francini and G. Longobardi, "Automatic inspection with optical position sensors," *Opt. Laser Eng.*, vol. 16, pp. 49-56, 1992.
- [60] R. G. Dorsch, G. Hausler, and J. M. Herrmann, "Laser triangulation: fundamental uncertainty in distance measurement," *Appl. Opt.*, vol. 33, no. 7, 1994.
- [61] X. Zhang, Y. Sui and J. Yang, "Nonlinearity of optical triangulation," *Proc. SPIE*, vol. 3558, pp. 262-265, 1998.
- [62] S. Grafstorm, J. Ackermann, T. Hagen, R. Neumann, O. Probst, "Analysis of lateral force effects on the topography in scanning probe microscopy," *J. Vac. Sci. Technol. B*, vol. 12, no. 3, pp. 1559-1564, 1994.
- [63] B. Schmidt and R. Ross, "Position-sensitive photodetectors made with standard silicon-planar technology," *Sensor. Actuat.*, vol. 4, pp. 439-466, 1983.

- [64] C. Hicks, M. Kalatsky, R. A. Metzler and A. O. Goushcha, "Quantum efficiency of silicon photodiodes in the near-infrared spectral range," *Appl. Opt.*, vol. 42, pp. 4415-4422, 2003.
- [65] A. J. Makynen, J. T. Kostamovaara and R. A. Myllylä, "Positioning resolution of the position-sensitive detectors in high background illumination," *IEEE Trans. Instrum. Meas.*, vol. 45, pp. 324-326, 1996.
- [66] C. H. Liu, W. Y. Jywe and C. K. Chen, "Development of a diffractive-type optical triangulation sensor," *Appl. Opt.*, vol. 43, pp. 5607-5613, 2004.
- [67] C. H. Liu, H. L. Huang and H. W. Lee, "Five-degrees -of-freedom diffractive laser encoder," *Appl. Opt.*, vol. 48, pp. 2767-2777, 2009.
- [68] M. Toyoda, K. Araki and Y. Suzuki, "Wave-front tilt sensor with two quadrant detectors and its application to a laser beam pointing system," *Appl. Opt.*, vol. 41, pp. 2219-2223, 2002.
- [69] D. Y. Lou, A. Martinez and D. Stanton, "Surface profile measurement with a dual-beam optical system," *Appl. Opt.*, vol. 23, pp. 746-751, 1984.
- [70] W. D. S. John, "Cylinder gauge measurement using a position sensitive detector," *Appl. Opt.*, vol. 46, pp. 7469-7474, 2007.
- [71] W. D. S. John, "Determination of the beam centroid of an obstructed focused Gaussian laser beam," *Appl. Opt.*, vol. 48, pp. 4501-4505, 2009.

- [72] M. Gögler, T. Betz and J. A. Käs, "Simultaneous manipulation and detection of living cell membrane dynamics," *Opt. Lett.*, vol. 32, pp. 1893-1895, 2007.
- [73] S. Jeon and T. Thundat, "Instant curvature measurement for microcantilever sensors," *Appl. Phys. Lett.*, vol. 85, no. 6, pp. 1083-1084, 2004.
- [74] R. Osiander *et al.*, "A microelectromechanical-based magnetostrictive magnetometer," *Appl. Phys. Lett.*, vol. 69, no. 19, pp. 2930-2931, 1996.
- [75] J. Ma, H. Sun, S. Wang and D. Yan, "Effects of atmospheric turbulence on photodetector arrays," *Appl. Opt.*, vol. 28, pp. 2123-2126, 1983.
- [76] T. E. Schäffer, M. Richter and M. B. Viani, "Array detector for the atomic force microscope," *Appl. Phys. Lett.*, vol. 76, pp. 3644-3646, 2000.
- [77] H. Golnabi, "Linearity in output signal of optical position-sensing detector systems," *Opt. Laser Technol.*, vol. 32, pp. 311-316, 2000.
- [78] W. Koechner and M. Bass, *Solid-State Lasers: A Graduate Text*. Springer, 2003.
- [79] S. Mailhot, Y. Champagne and N. McCarthy, "Single-mode operation of a broad-area semiconductor laser with an anamorphic external cavity: experimental and numerical results," *Appl. Opt.*, vol. 39, pp. 6806-6813, 2000.
- [80] C. Hsu, D. Cheng, W. Kuo and T. Yen, "Conducting single-transverse -mode operation with a wide tuning range in commercial multimode VCSELs by beam-profile-adapted optical feedback," *Opt. Lett.*, vol. 34, pp. 2861-2863, 2009.

- [81] J. Campos, A. Corróns and A. Pons, "Response uniformity of silicon photodiodes," *Appl. Opt.*, vol. 27, pp. 5154-5156, 1988.
- [82] S. Hu, S. Howell, A. Raman, R. Reifenberger and M. Franchek, "Frequency domain identification of tip-sample van der Waals interactions in resonant atomic force microcantilevers," *J. Vib. Acoust.* vol. 126, pp. 343-351, 2004.
- [83] S. J. Ahn *et al.*, "AFM nanolithography on a mixed LB film of hexadecylamine and palmitic acid," *Ultramicroscopy*, vol. 91, pp. 171-176, 2002.
- [84] W. F. Kolbe, D. F. Ogletree and M. B. Salmeron, "Atomic force microscopy imaging of T4 bacteriophages on silicon substrates," *Ultramicroscopy*, vol. 42, pp. 1113-1117, 1992.
- [85] R. Afrin, T. Yamada and A. Ikai, "Analysis of force curves obtained on the live cell membrane using chemically modified AFM probes," *Ultramicroscopy*, vol. 100, pp. 187-195, 2004.
- [86] C. Espenel *et al.*, "Temperature-dependent imaging of living cells by AFM," *Ultramicroscopy*, vol. 108, pp. 1174-1180, 2008.
- [87] R. García and R. Pérez, "Dynamic atomic force microscopy methods," *Surf. Sci. Rep.*, vol. 47, pp. 197-302, 2002.
- [88] Y. Martin, C. C. Williams and H. K. Wickramasinghe, "Atomic force microscope force mapping and profiling on a sub 100-A scale," *J. Appl. Phys.*, vol. 61, pp. 4723-4729, 1987.

- [89] J. P. Cleveland, B. Anczykowski, A. E. Schmid and V. B. Elings, “Energy dissipation in tapping-mode atomic force microscopy,” *Appl. Phys. Lett.*, vol. 72, pp. 2613-2615, 1998.
- [90] G. Reiter *et al.*, “Direct visualization of random crystallization and melting in arrays of nanometer-size polymer crystals,” *Phys. Rev. Lett.*, vol. 87, 226101, 2001.
- [91] A. Sebastian, M. V. Salapaka, D. J. Chen and J. P. Cleveland, “Harmonic and power balance tools for tapping-mode atomic force microscope,” *J. Appl. Phys.*, vol. 89, pp. 6473-6480, 2001.
- [92] T. R. Rodríguez and R. García, “Tip motion in amplitude modulation (tapping-mode) atomic-force microscopy: Comparison between continuous and point-mass models,” *Appl. Phys. Lett.*, vol. 80, pp. 1646-1648, 2002.
- [93] S. Rützel, S. I. lee and A. Raman, “Nonlinear dynamics of atomic-force-microscope probes driven in Lennard-Jones potentials,” *Proc. R. Soc. Lond. A*, vol. 459, pp. 1925-1948, 2003.
- [94] O. Sahin, S. Magonov, C. Su, C. F. Quate and O. Solgaard, “An atomic force microscope tip designed to measure time-varying nanomechanical force,” *Nature Nanotech.*, vol. 2, pp. 507-514, 2007.
- [95] H. B. G. Casimir, “On the attraction between two perfectly conducting plates,” *Proc. K. Ned. Akad. Wet.*, vol. 51, pp. 793-795, May 1948.

- [96] P. H. G. M. Van Blokland and J. T. G. Overbeek, “Van der Waals forces between objects covered with a chromium layer,” *J. Chem. Soc. Faraday Trans.*, vol. 74, pp. 2637-2651, 1978.
- [97] S. K. Lamoureux, “Demonstration of the Casimir force in the 0.6 to 6 μm range,” *Phys. Rev. Lett.*, vol. 78, pp. 5-8, 1997.
- [98] S. de Man, K. Keck, R. J. Wijngaarden and D. Iannuzzi, “Having the Casimir force with conductive oxides,” *Phys. Rev. Lett.*, vol. 103, 040402, 2009.
- [99] F. Chen, G. L. Klimchitskaya, V. M. Mostepanenko and U. Mohideen, “Demonstration of the difference in the Casimir force for samples with different charge-carrier densities,” *Phys. Rev. Lett.*, vol. 97, 170402, 2006.
- [100] B. V. Derjaguin and I. I. Abrikosova, “Direct measurement of the molecular attraction of solid bodies. Statement of the problem and measurement of the force by using negative feedback,” *Sov. Phys. JETP*, vol. 3, pp. 819-829, 1957.
- [101] T. Ederth, “Template-stripped gold surfaces with 0.4-nm rms roughness suitable for force measurements: application to the Casimir force in the 20100 nm range,” *Phys. Rev. A*, vol. 62, 062104, 2000.
- [102] J. G. Guo and Y. P. Zhao, “Influence of van der Waals and Casimir forces on electrostatic torsional actuators,” *J. Microelectromech. Syst.*, vol. 13, no. 6, pp. 1027-1035, Dec. 2004.

- [103] T. Hirano, T. Furhata, K. J. Gabriel and H. Fujita, "Design, fabrication, and operation of submicron gap comb-drive microactuators," *J. Microelectromech. Syst.*, vol. 1, no. 1, pp. 52-59, Mar. 1992.
- [104] H. G. Craighead, "Nanoelectromechanical systems," *Science*, vol. 290, pp. 1532-1535, Nov. 2000.
- [105] H. B. Chan, V. A. Aksyuk, R. N. Kleiman, D. J. Bishop and F. Capasso, "Quantum mechanical actuation of microelectromechanical systems by the Casimir force," *Science*, vol. 291, pp. 1941-1944, 2001.
- [106] H. B. Chan, V. A. Aksyuk, R. N. Kleiman, D. J. Bishop and F. Capasso, "Non-linear micromechanical Casimir oscillator," *Phys. Rev. Lett.*, vol. 87, 211801, 2001.
- [107] F. Capasso, J. N. Munday, D. Iannuzzi and H. B. Chan, "Casimir forces and quantum electrodynamical torques: physics and nanomechanics," *IEEE J. Sel. Topics. Quantum Electron.*, vol. 13, no. 2, pp. 400-414, Mar. 2007.
- [108] Y. Nemirovsky and O. Degani, "A methodology and model for the pull-in parameters of electrostatic actuators," *J. Microelectromech. Syst.*, vol. 10, no. 4, pp. 601-615, Dec. 2001.
- [109] O. Degani and Y. Nemirovsky, "Design considerations of rectangular electrostatic torsion actuators based on new analytical pull-in expressions," *J. Microelectromech. Syst.*, vol. 11, no. 1, pp. 20-26, Feb. 2002.

- [110] F. M. Serry, D. Walliser and G. J. Maclay, “The anharmonic Casimir oscillator (ACO)- the Casimir effect in a model microelectromechanical system,” *J. Microelectromech. Syst.*, vol. 4, no. 4, pp. 193-205, Dec. 1995.
- [111] V. J. Aidala, “Parameter estimation via the kalman filter,” *IEEE Trans. Auto. Cont.* AC-22, 3, pp. 471-472, June 1977.
- [112] G. Tao, *Adaptive control design and analysis*: John Wiley and Sons, Inc. 2003.
- [113] U. Parlitz, “Estimating Model Parameters from Time Series by Autosynchronization,” *Phys. Rev. Lett.*, vol. 76, pp. 1232-1235, 1996.
- [114] M. T. Heath, *Scientific computing, an introductory survey*: McGraw-Hill, 1997.
- [115] J. Bošković, “Adaptive control of a class of nonlinearly parameterized plants,” *IEEE Trans. Autom. Control*, vol. 43, no. 7, July 1998.
- [116] M. J. Sparnaay, “Measurements of attractive forces between flat plates,” *Physica*, vol. 24, pp. 751-764, 1958.
- [117] G. L. Klimchitskaya and V. M. Mostepanenko, “Experiment and theory in the Casimir effect,” *Contemp. Phys.*, vol. 47, pp. 131-144, 2006.
- [118] H. B. G. Casimir and D. Polder, “The influence of retardation on the London-van der Waals forces,” *Phys. Rev.*, vol. 73, no. 4, pp. 360-372, 1948.
- [119] E. M. Lifshitz, “The theory of molecular attractive forces between solids,” *Sov. Phys. JETP*, vol. 2, pp. 7383, 1956.

- [120] I. E. Dzyaloshinskii, E. M. Lifshitz, and L. P. Pitaevskii, “The general theory of van derWaals forces,” *Adv. Phys.*, vol. 10, pp. 165-209, 1961.
- [121] P. A. M. Neto, A. Lambrecht, and S. Reynaud, “Roughness correction to the Casimir force: Beyond the proximity force approximation,” *Europhys. Lett.*, vol. 69, pp. 924930, 2005.
- [122] O. Kenneth, I. Klich, A. Mann, and M. Revzen, “Repulsive casimir forces,” *Phys. Rev. Lett.*, vol. 89, 033001, 2002.
- [123] J. N. Munday, F. Capasso, and V. A. Parsegian, “Measured long-range repulsive CasimirLifshitz forces,” *Nature*, vol. 457, pp. 170-173, 2009.
- [124] S. Kawai, and H. Kawakatsu, “Atomically resolved dynamic force microscopy operating at 4.7 MHz,” *Appl. Phys. Lett.*, 88, 133103, 2006.
- [125] J. Stulemeijer, J. A. Bielen, P. G. Steeneken, and J. B. van der Berg, “Numerical path following as an analysis method for electrostatic MEMS,” *J. Microelectromech. Syst.*, vol. 18, no. 2, pp. 488-499, April 2009.
- [126] J. M. T. Thompson, and H. B. Stewart, *Nonlinear dynamics and chaos*: Wiley, 1986.
- [127] S. K. De, and N. R. Aluru, “Complex nonlinear oscillations in electrostatically actuated microstructures,” *J. Microelectromech. Syst.*, vol. 15, no. 2, pp. 355-369, April 2006.

- [128] S. K. De, and N. R. Aluru, “Complex oscillations and chaos in electrostatic MEMS under superharmonic excitations,” *Phys. Rev. Lett.*, vol. 94, 204101, 2005.
- [129] Y. C. Wang, S. G. Adams, J. S. Thorp, N. C. MacDonald, P. Hartwell, and F. Bertsch “Chaos in MEMS, parameter estimation and its potential applications,” *IEEE Trans. Circuit Syst. I: Fundam. Theory Appl.*, vol. 45, pp. 1013-1020, 1998.
- [130] M. Serry, D. Walliser, and J. Maclay, “The role of the casimir effect in the static deflection and stiction of membrane strips in microelectromechanical systems (MEMS),” *J. Appl. Phys.*, vol. 84, pp. 2501-2506, 1998.
- [131] D. -H. Wen, B. -A. Li, and L. -W. Chen, “Supersoft Symmetry Energy Encountering Non-Newtonian Gravity in Neutron Stars,” *Phys. Rev. Lett.*, vol. 103, 211102, 2009.
- [132] A. Dekel, M. Lecar, and J. Shaham, “Tidal effects on the mass profile of galactic haloes, *Nature*, vol. 286, pp. 135-136, 1980.
- [133] A. Varone, A. Politi, and M. Ciofini, “CO₂ laser dynamics with feedback,” *Phys. Rev. A*, vol. 52, pp. 3176-3182, 1995.
- [134] V. B. Bezerra, G. L. Klimchitskaya, and V. M. Mostepanenko, “High-order conductivity corrections to the Casimir force,” *Phys. Rev. A*, vol. 62, 014102, 2000.
- [135] M. Bordag, G. L. Klimchitskaya, U. Mohideen, and V. M. Mostepanenko, *Advances in The Casimir Effect*. Oxford University Press, New York, 2009.

- [136] I. Brevik, S. A. Ellingsen, and K. A. Milton, “Thermal corrections to the Casimir effect,” *New J. Phys.*, vol. 8, 236, 2006.
- [137] H. Ehrenreich, P. L. Sagalyn, and H. R. Philipp, *Phys. Rev.*, vol. 128, 1622, 1962.
- [138] D. R. Sahoo, A. Sebastian, and M. V. Salapaka, “Transient-signal-based sample-detection in atomic force microscopy,” *Appl. Phys. Lett.*, vol. 83, no. 26, 5521, 2003.
- [139] R. Castillo-Garza, C. -C. Chang, D. Yan, and U. Mohideen, “Customized silicon cantilevers for Casimir force experiments using focused ion beam milling,” *J. Phys.*, vol. 161, 012005, 2009.
- [140] D. Hua, and Y. Ma, “Hysteresis phenomena in CO catalytic oxidation system in the presence of inhomogeneities of the catalyst surface,” *Phys. Rev. E*, vol. 66, 066103, 2002.
- [141] X. Huang, Y. Qian, X. Zhang, and G. Hu, “Hysteresis and bistability in periodically paced cardiac tissue,” *Phys. Rev. E*, vol. 81, 051903, 2010.
- [142] R. D. Deegan, “Stress hysteresis as the cause of persistent holes in particulate suspensions,” *Phys. Rev. E*, vol. 81, 036319, 2010.
- [143] J. Legleiter, M. Park, B. Cusick, and T. Kowalewski, “Scanning probe acceleration microscopy (SPAM) in fluids: Mapping mechanical properties of surfaces at the nanoscale,” *Proc. Natl. Acad. Sci. USA*, vol. 103, pp. 4813-4818, 2006.
- [144] A. Maali *et al.*, “Hydrodynamics of oscillating atomic force microscopy cantilevers in viscous fluids,” *J. Appl. Phys.*, vol. 97, 074907, 2005.

- [145] M. Balantekin, A. G. Onaran and F. L. Degertekin, “Quantative mechanical characterization of materials at the nanoscale through direct measurement of time-resolved tip-sample interaction forces,” *Nanotechnology*, vol. 19, 085704, 2008.
- [146] R. Garcia, J. Tamayo, M. Calleja and F. Garcia, “Phase contrast in tapping-mode scanning force microscopy,” *Appl. Phys. A*, vol. 66, pp. 309-312, 1998.

Appendix A

Properties of Cost Function Eq.

(7.13)

In this Appendix, we will show that with appropriately chosen ΔT and γ_1 , Eq. (7.18) - Eq. (7.20) are valid. First of all, it is denoted that:

$$\dot{\hat{\xi}}(t) + \gamma_1 \hat{\xi}(t) = \frac{1}{(\hat{z}_1 - \theta(t))^3} \quad (\text{A.1})$$

It is easy to verify that:

$$\begin{aligned} \frac{dF(\hat{z}_1)}{d\hat{z}_1} &= \sum_{i=1}^N [(\zeta(t_{j_i-1}) - \eta(t_{j_i-1}))\hat{\xi}(t_{j_i}) - (\zeta(t_{j_i}) - \eta(t_{j_i}))\hat{\xi}(t_{j_i-1})] \\ &= \sum_{i=1}^N B \left[\int_{\tau=0}^{t_{j_i-1}} \frac{1}{(z_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i-1}-\tau)} d\tau \int_{\tau=0}^{t_{j_i}} \frac{1}{(\hat{z}_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i}-\tau)} d\tau \right. \\ &\quad \left. - \int_{\tau=0}^{t_{j_i}} \frac{1}{(z_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i}-\tau)} d\tau \int_{\tau=0}^{t_{j_i-1}} \frac{1}{(\hat{z}_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i-1}-\tau)} d\tau \right] \\ &= \sum_{i=1}^N B \left[\int_{\tau=0}^{t_{j_i-1}} \frac{1}{(z_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i-1}-\tau)} d\tau \int_{\tau=t_{j_i-1}}^{t_{j_i}} \frac{1}{(\hat{z}_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i}-\tau)} d\tau \right. \\ &\quad \left. - \int_{\tau=t_{j_i-1}}^{t_{j_i}} \frac{1}{(z_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i}-\tau)} d\tau \int_{\tau=0}^{t_{j_i-1}} \frac{1}{(\hat{z}_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i-1}-\tau)} d\tau \right] \quad (\text{A.2}) \end{aligned}$$

Since the filter is designed to have a fast decay, it can be approximated that:

$$\begin{aligned}
\frac{dF(\hat{z}_1)}{d\hat{z}_1} &\simeq \sum_{i=1}^N B \left[\int_{\tau=t_{j_i-2}}^{t_{j_i-1}} \frac{1}{(z_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i-1}-\tau)} d\tau \int_{\tau=t_{j_i-1}}^{t_{j_i}} \frac{1}{(\hat{z}_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i}-\tau)} d\tau \right. \\
&- \int_{\tau=t_{j_i-1}}^{t_{j_i}} \frac{1}{(z_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i}-\tau)} d\tau \int_{\tau=t_{j_i-2}}^{t_{j_i-1}} \frac{1}{(\hat{z}_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i-1}-\tau)} d\tau \\
&+ \int_{\tau=t_{j_i-3}}^{t_{j_i-2}} \frac{1}{(z_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i-1}-\tau)} d\tau \int_{\tau=t_{j_i-1}}^{t_{j_i}} \frac{1}{(\hat{z}_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i}-\tau)} d\tau \\
&- \left. \int_{\tau=t_{j_i-1}}^{t_{j_i}} \frac{1}{(z_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i}-\tau)} d\tau \int_{\tau=t_{j_i-3}}^{t_{j_i-2}} \frac{1}{(\hat{z}_1 - \theta(\tau))^3} e^{-\gamma_1(t_{j_i-1}-\tau)} d\tau \right] \quad (\text{A.3})
\end{aligned}$$

A sufficient condition to satisfy Eq. (7.18) - Eq. (7.20) is: $\forall t_p \in [t_{j_i-1}, t_{j_i}]$, $t_q \in [t_{j_i-2}, t_{j_i-1}]$, $\exists t_r \in [t_{j_i-3}, t_{j_i-2}]$ where $\frac{t_q - t_{j_i-2}}{t_{j_i-1} - t_{j_i-2}} = \frac{t_r - t_{j_i-3}}{t_{j_i-2} - t_{j_i-3}}$, such that the following is valid:

$$g(t_q, t_r) < 0 \quad \hat{z}_1 < z_1 \quad (\text{A.4})$$

$$g(t_q, t_r) = 0 \quad \hat{z}_1 = z_1 \quad (\text{A.5})$$

$$g(t_q, t_r) > 0 \quad \hat{z}_1 > z_1 \quad (\text{A.6})$$

where

$$\begin{aligned}
g(t_q, t_r) &= e^{-\gamma_1(t_{j_i}-t_p)} e^{-\gamma_1(t_{j_i-1}-t_q)} \\
&\times \left(\frac{1}{(z_1 - \theta(t_q))^3} \frac{1}{(\hat{z}_1 - \theta(t_p))^3} - \frac{1}{(z_1 - \theta(t_p))^3} \frac{1}{(\hat{z}_1 - \theta(t_q))^3} \right) \\
&+ e^{-\gamma_1(t_{j_i}-t_p)} e^{-\gamma_1(t_{j_i-1}-t_r)} \\
&\times \left(\frac{1}{(z_1 - \theta(t_r))^3} \frac{1}{(\hat{z}_1 - \theta(t_p))^3} - \frac{1}{(z_1 - \theta(t_p))^3} \frac{1}{(\hat{z}_1 - \theta(t_r))^3} \right) \quad (\text{A.7})
\end{aligned}$$

To satisfy Eq. (A.4) - Eq. (A.6), it requires:

$$\begin{aligned}
& (\theta(t_q) - \theta(t_p)) \\
& \times \left[\frac{(z_1 - \theta(t_q))^2(\hat{z}_1 - \theta(t_p))^2 + (z_1 - \theta(t_p))^2(\hat{z}_1 - \theta(t_q))^2}{(z_1 - \theta(t_p))^3(\hat{z}_1 - \theta(t_q))^3(z_1 - \theta(t_q))^3(\hat{z}_1 - \theta(t_p))^3} \right. \\
& \quad \left. + \frac{(z_1 - \theta(t_p))(\hat{z}_1 - \theta(t_q))(z_1 - \theta(t_q))(\hat{z}_1 - \theta(t_p))}{(z_1 - \theta(t_p))^3(\hat{z}_1 - \theta(t_q))^3(z_1 - \theta(t_q))^3(\hat{z}_1 - \theta(t_p))^3} \right] \\
& - (\theta(t_p) - \theta(t_r)) \\
& \times \left[\frac{(z_1 - \theta(t_r))^2(\hat{z}_1 - \theta(t_p))^2 + (z_1 - \theta(t_p))^2(\hat{z}_1 - \theta(t_r))^2}{(e^{\gamma_1 \Delta T})(z_1 - \theta(t_p))^3(\hat{z}_1 - \theta(t_r))^3(z_1 - \theta(t_r))^3(\hat{z}_1 - \theta(t_p))^3} \right. \\
& \quad \left. + \frac{(z_1 - \theta(t_p))(\hat{z}_1 - \theta(t_r))(z_1 - \theta(t_r))(\hat{z}_1 - \theta(t_p))}{(e^{\gamma_1 \Delta T})(z_1 - \theta(t_p))^3(\hat{z}_1 - \theta(t_r))^3(z_1 - \theta(t_r))^3(\hat{z}_1 - \theta(t_p))^3} \right] \\
& > 0
\end{aligned} \tag{A.8}$$

Since $\theta(t_p), \theta(t_r) < c_0 < \theta(t_q)$, a sufficient condition to satisfy (A.8) is:

$$(\theta(t_q) - \theta(t_p)) - (\theta(t_p) - \theta(t_r))e^{-\gamma_1 \Delta T} > 0 \tag{A.9}$$

which requires the chosen γ_1 and ΔT to satisfy:

$$e^{\gamma_1 \Delta T} > \max\left(\frac{\theta(t_p) - \theta(t_r)}{\theta(t_q) - \theta(t_p)}\right) \tag{A.10}$$

Appendix B

Properties of Cost Function Eq.

(8.9)

In this section, we will prove that the proposed cost function $G(\hat{z})$ satisfy Eq. (8.10) - Eq. (8.12) under normal operation conditions. When $\gamma_1 > \sqrt{k/m}$, the second order filters in Eq. (8.6) - Eq. (8.8) are over-damped which lead to:

$$y(t) = \int_0^t RQ(t-\tau) \frac{F_c(z+x(\tau))}{m} d\tau \quad (\text{B.1})$$

$$\xi(t) = \int_0^t Q(t-\tau) \frac{F_c(\hat{z}+x(\tau))}{m} d\tau \quad (\text{B.2})$$

where $Q(t-\tau) = (\frac{e^{-\alpha(t-\tau)}-e^{-\beta(t-\tau)}}{\beta-\alpha})$, $\alpha = \gamma_1 - \sqrt{\gamma_1^2 - \omega_0^2}$, $\beta = \gamma_1 + \sqrt{\gamma_1^2 - \omega_0^2}$, and $\omega_0 = \sqrt{k/m}$. From the Cauchy-Schwarz inequality, we have:

$$\int_{t_1}^{t_2} y^2(t) dt \int_{t_1}^{t_2} \xi^2(t) dt \geq (\int_{t_1}^{t_2} y(t)\xi(t) dt)^2. \quad (\text{B.3})$$

The equality is valid only when $y(t) = R\xi(t)$. This is equivalent to $\hat{z} = z$. Thus, we can conclude that the cost function $G(\hat{z})$ will be equal to zero only when $\hat{z} = z$ and it will be positive otherwise. The first derivative of $G(\hat{z})$ with respect to $\hat{z}$ can be expressed

as:

$$\begin{aligned} \frac{dG(\hat{z})}{d\hat{z}} &= 2 \frac{\int_{t_1}^{t_2} y(t)\xi(t)dt}{(\int_{t_1}^{t_2} \xi^2(t)dt)^2} \\ &\times \left[\int_{t_1}^{t_2} y(t)\xi(t)dt \int_{t_1}^{t_2} \xi(t) \frac{d\xi(t)}{d\hat{z}_1} dt \right. \\ &\quad \left. - \int_{t_1}^{t_2} \xi^2(t)dt \int_{t_1}^{t_2} y(t) \frac{d\xi(t)}{d\hat{z}_1} dt \right]. \end{aligned} \quad (\text{B.4})$$

From Eqs. (B.1) and (B.2), it can be concluded that $y(t)$ and $\xi(t)$ are non-negative which implies:

$$\frac{\int_{t_1}^{t_2} y(t)\xi(t)dt}{(\int_{t_1}^{t_2} \xi^2(t)dt)^2} > 0. \quad (\text{B.5})$$

So the sign of the first derivative of $G(\hat{z})$ is the same as the sign of $L(t_2, t_1)$ where $L(t_2, t_1)$ is denoted as:

$$\begin{aligned} L(t_2, t_1) &= \left[\int_{t_1}^{t_2} y(t)\xi(t)dt \int_{t_1}^{t_2} \xi(t) \frac{d\xi(t)}{d\hat{z}_1} dt \right. \\ &\quad \left. - \int_{t_1}^{t_2} \xi^2(t)dt \int_{t_1}^{t_2} y(t) \frac{d\xi(t)}{d\hat{z}_1} dt \right]. \end{aligned} \quad (\text{B.6})$$

When $\hat{z} = z$, it is easy to verify that $L(t_2, t_1) = 0$. This leads to Eq. (8.11). Next, we will evaluate $L(t_2, t_1)$ when $\hat{z} \neq z$. Under normal operation conditions, the probe trajectory is periodic [110]: $x(t + T) = x(t)$. The plate's oscillation frequency will be different from the excitation frequency if the system is experiencing severe nonlinearity.

Next, we will prove that $y(t)$ is also periodic in steady state:

$$\begin{aligned} y(t+T) &= \int_T^{t+T} RQ(t+T-\tau) \frac{F_c(z+x(\tau))}{m} d\tau \\ &+ \int_0^T RQ(t+T-\tau) \frac{F_c(z+x(\tau))}{m} d\tau. \end{aligned} \quad (\text{B.7})$$

When the probe vibration is in steady state, t is sufficiently large. It can be calculated that $Q(t+T-\tau) \approx 0$ when $\tau \in [0, T]$. Then the following equation is valid:

$$\int_0^T RQ(t+T-\tau) \frac{F_c(z+x(\tau))}{m} d\tau \approx 0 \quad (\text{B.8})$$

which is equivalent to:

$$\begin{aligned} y(t+T) &\approx \int_T^{t+T} RQ(t+T-\tau) \frac{F_c(z+x(\tau))}{m} d\tau \\ &= \int_0^t RQ(t+T-(\tau+T)) \frac{F_c(z+x(\tau+T))}{m} d(\tau+T) \\ &= \int_0^t RQ(t-\tau) \frac{F_c(z+x(\tau))}{m} d(\tau) \\ &= y(t). \end{aligned} \quad (\text{B.9})$$

It can be proved in the same way that $\xi(t) = \xi(t+T)$. If the estimation period $t_2 - t_1$ is large compared to T or if $(t_2 - t_1) = NT$ where N is an integer, we have the following relationship based on the periodicity of $y(t)$ and $\xi(t)$:

$$\begin{aligned} L(t_2, t_1) = N \quad [& \int_{t_1}^{t_1+T} y(t)\xi(t)dt \int_{t_1}^{t_1+T} \xi(t) \frac{d\xi(t)}{d\hat{z}} dt \\ & - \int_{t_1}^{t_1+T} \xi^2(t)dt \int_{t_1}^{t_1+T} y(t) \frac{d\xi(t)}{d\hat{z}} dt]. \end{aligned} \quad (\text{B.10})$$

Numerical calculations can be easily performed to investigate the relationship among $L(t_2, t_1)$, z and $\hat{z}$. For simplicity, we denote $L_1 = L(t_2, t_1)/N$. Based on our modified AFM setup, numerical calculation results show that:

$$L_1 < 0 \quad (\hat{z} < z) \tag{B.11}$$

$$L_1 = 0 \quad (\hat{z} = z) \tag{B.12}$$

$$L_1 > 0 \quad (\hat{z} > z). \tag{B.13}$$

This is equivalent to Eq. (8.10)- Eq. (8.12) and our estimation method proposed is proved to be accurate.