SCANNING TUNNELING MICROSCOPY

AND DIGITAL IMAGE CORRELATION

IN NANOMECHANICS INVESTIGATIONS

Thesis by

Ying Huang

In Partial Fulfillment of the Requirements

For the Degree of

Doctor of Philosophy

California Institiute of Technology

Pasadena, California

2001

(Defended March 16, 2001)

Dedicated to my parents

Acknowledgements

First I would like to express my sincere gratitude to my advisor Dr. W.G. Knauss for his guidance, support, patience and confidence in me throughout my entire graduate study. Over the past five and a half years, I have learned a lot of valuable lessons from a true scientist like him, which will definitely have an impact in my future career. Special thanks to Dr. G. Ravichandran, Dr. A. Rosakis and Dr. M. Gharib for serving on my thesis committee and taking the time to review my thesis.

I sincerely thank Dr. Guillaume Vendroux for his pioneer work on the STM and the DIC, and Dr. Mathias Kinzler for training me on the STM and always replying promptly to my questions even after he left for Germany. I wish to thank Allan Goudy, Steve Palm and Dr. Dan Lang for their invaluable help on the STM electronics.

Many thanks are due to my colleagues and friends here at Caltech. I thank my current and previous group mates Dr. Hongbing Lu, Dr. Sanwook Lee, Dr. Sandeep Sane, Ioannis Chasiotis, and my officemates Dr. Pradeep Guduru and Demir Coker for our discussions about work and life. Special thanks to Dr. Claude Seywert and Winston Pun for their help in many little things and the fun things we did together.

Last but not the least, I would like to thank Dr. Giorgio Isella for his support both technically and mentally. Without his love and friendship, my stay at Caltech would not have been so enjoyable.

Abstract

Probe microscopy (scanning tunneling microscopy and atomic force microscopy) and digital image correlation together serve as a potentially powerful tool for experimentally investigating the mechanical behaviors of materials at the sub-micron and nanometer scales. Based on the tunneling effect in quantum physics, the scanning tunneling microscope (STM) records surface topography quantitatively and can achieve angstrom resolution. The digital image correlation (DIC) extracts the displacements and gradients from the undeformed and deformed topographical images.

In this work, a calibration has been performed on the existing STM built "in-house" and the coefficients used in the STM system were confirmed. Major improvements on several components of the system have been made, including constructing a new actuator probe to decouple its in-plane and out-of-plane movements, designing and implementing a new first-stage amplifier to reduce the noise output by a factor of 10 and modeling of the controller in the STM feedback loop.

Further, systematic study of the digital image correlation has been conducted. In the simple case of one-dimensional correlation, key parameters involved are the subset size, variables in the displacement representation, frequency content of the signal and noise. The one-dimensional study was then extended to two dimensions. In addition to those key parameters identified in the one-dimensional study, the sampling rate poses substantial influence on the correlation accuracy. Low amplitude, high frequency noise still increases the correlation error significantly.

Finally, a detailed analysis of noise originating from the atomic force microscope (AFM) reveals that white noise can generate an "artificial" displacement that can cloud the displacement calculation when correlation is performed on a deformed specimen. A possible improvement is proposed to compensate for the hysteresis of the AFM piezo actuator.

Table of Contents

List of Figures

List of Tables

1 Introduction

During the last decade, solid mechanics has gradually evolved to the interdisciplinary area of mechanics of materials, which addresses material related problems from the viewpoint of mechanics. Many topics in this area are associated with a small size scale that was not of major concern in traditional solid mechanics. For example, with the rapid development in the semiconductor industry, MEMS (MicroElectroMechanicalSystems) have received increasing attention. MEMS have found applications in many disciplines, from chemical sensors and medical devices to controllers in the auto and aerospace industries. Because of the small size scale concerned in most MEMS (micron to sub-micron), their reliability is a critical issue in mechanical designs. Particularly in the semiconductor industry, recent research has been focused on the characterization and development of the packaging of microelectronic devices. In the field of composite materials, the mechanical strength of the interfacial region (on a sub-micron size scale) determines the strength of the material, where failure typically occurs in the interfacial regions. Therefore, examination of this region will be beneficial in the evolution of high strength composite materials for aircraft design as well as other applications. In fracture mechanics, most failure of materials initiates from a small void inside the material. Characterization of the stress and strain field at the vicinity of the crack tip will help us understand the failure process, so that criteria can be proposed in engineering design. For all of these examples, it is of fundamental importance that research in micro/nano mechanics be supported by experimental investigations.

Traditionally, the experimental solution to address this need for micro-mechanical studies has been via optical methods, e.g., photo-elasticity, holography, moiré interferometry, speckle interferometry, speckle photography, white light speckle, etc. Due to limitations imposed by the wavelength of light (about 500nm), these optical methods cannot serve for measurements in the sub-micron range. In the nineteen thirties, the transmission electron microscope (TEM) and the scanning electron microscope (SEM) were invented. The TEM works much like a slide projector, or a light transmission microscope, except that a focused beam of electrons is used instead of light to "see through" the specimen. The SEM was developed later as a result of the electronics required to scan the beam of electrons across the sample. The SEM is patterned after reflected light microscopes and yields surface features in addition to information on microstructures that can be obtained from a TEM. These electron microscopes magnify a thousand times more than optical microscopes and can reach resolutions of angstroms. However, the surface profiles obtained from electron microscopes are more qualitative than quantitative. Moreover, these microscopes require observation in vacuum, which makes it difficult to conduct in-situ measurements.

In the early eighties, the scanning tunneling microscope (STM) was invented by Gerd Binnig and Heinrich Rohrer at IBM Zurich. Based on the tunneling effect in quantum physics, the STM records surface topography quantitatively and can achieve angstrom resolution. Because the STM does not emit an electron beam, it does not require a vacuum environment and will not damage the specimen. However, the STM has the drawback that it only works with conductive surfaces and may have problems with metals that oxidize fast. Shortly after the invention of the STM, another probe

microscope, the atomic force microscope (AFM), was invented. By indirectly measuring the atomic force between the probe and the specimen, the AFM obtains quantitative surface images and reaches angstrom resolution as well. Compared to the STM, the AFM works with both conductive and non-conductive materials.

During the same time period, pattern recognition algorithms were developed to measure in-plane displacement field from digital photographs of deformed specimen. In 1982, Peters and Ranson[13] proposed a digital image (DIC) correlation method and applied it to laser speckle images to perform stress analysis. Later Sutton et al.[14]-[16] made major improvements to this technique and used it in determining planar deformation. Luo and Chao[17] extended it to measure 3-D displacement field through stereo imaging. In 1990, Vendroux and Knauss[18]-[19] adapted this method to identify displacements from topographical images obtained with a scanning tunneling microscope.

The probe microscopy (STM and AFM) together with the digital image correlation can be a powerful tool to perform mechanical investigations on the sub-micron size scale. The objective of this thesis is to carry on the pioneer work by Vendroux and Knauss, identify the precision and the limitation of the STM and the DIC quantitatively, and improve the method where applicable. It is divided into three chapters.

Chapter 2 describes the working principle and the experimental setup of the STM. A calibration is performed to examine the coefficients used in the current STM system. Improvements are made in various components of the STM.

Chapter 3 is dedicated to the digital image correlation. The chapter is composed of two parts, for studies on one-dimensional and two-dimensional DIC respectively. In both studies, equations are derived and the correlation process is described. The parameters

involved in the correlation are evaluated and their roles are identified. Particularly, the influence of the high frequency noise on the correlation is examined.

In Chapter 4, a detailed analysis on the nature of the noise associated with AFM scans is carried out. The influence of white noise, intrinsic in AFM instrumentation, on the correlation is revealed and a possible improvement to compensate for the hysteresis of the AFM piezo actuator is proposed.

2 The Scanning Tunneling Microscope

Traditionally, measurements of surface deformations have been performed in solid mechanics via optical methods, as, e.g., holography, moiré interferometry, speckle interferometry, speckle photography, white light speckle, etc. Due to limitations imposed by the wavelength of light (about 500nm), these optical methods cannot serve for measurements in the sub-micron range. In the nineteen thirties, the transmission electron microscope (TEM) and the scanning electron microscope (SEM) were invented. The TEM works much like a slide projector, or a light transmission microscope, except that a focused beam of electrons is used instead of light to "see through" the specimen. The SEM was developed later as a result of the electronics required to scan the beam of electrons across the sample. The SEM is patterned after reflected light microscopes and yields surface features in addition to information on microstructures that can be obtained from a TEM. These electron microscopes magnify a thousand times more than optical microscopes and can reach resolutions of angstroms. However, the surface profiles obtained from electron microscopes are more qualitative than quantitative. Moreover, these microscopes require observation in vacuum, which makes it difficult to conduct in-situ measurements. In the eighties, the scanning tunneling microscope (STM) was invented. Based on the tunneling effect in quantum physics, the STM records surface topography quantitatively and can achieve angstrom resolution. Because the STM does not emit an electron beam, it does not require a vacuum environment and will not damage

the specimen. However, the STM has the drawback that it only works with conductive surfaces and may have problems with metals that oxidize fast.

An "in-house" scanning tunneling microscope was designed, developed by Dr. Vendroux[6]. In this work, the STM is calibrated using a step calibration standard. Major improvements have been made to the hardware, including the current-to-voltage converter and the feedback loop controller. More importantly, a new STM head with the in-plane and out-of-place movements decoupled has been designed and implemented.

2.1 The Working Principle

The theory of the scanning tunneling microscope is based on the tunneling effect in quantum physics. In practice, this is implemented via a feedback control loop to bring the probe to a constant distance from the surface to be scanned. The fine movement of the probe is adjusted via piezo-ceramic actuators.

2.1.1 Tunneling Phenomenon

The scanning tunneling microscope is based on the "tunneling" of electrons between an electrically charged surface and a conductive tip.
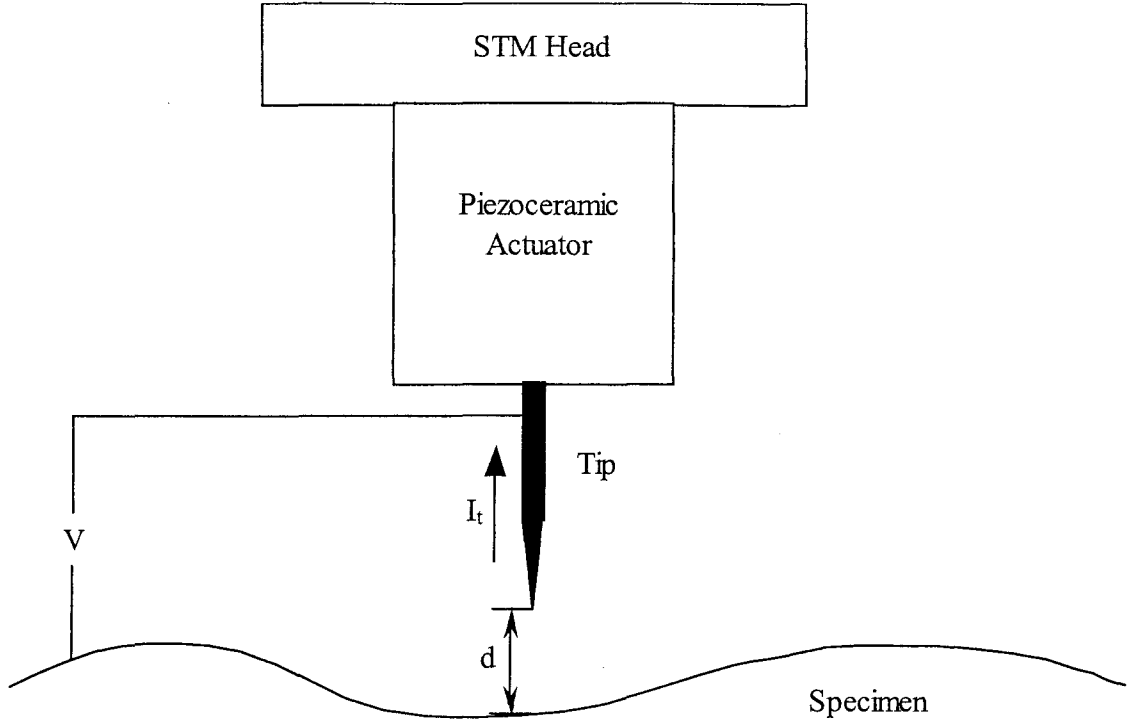
Figure 2.1. STM working principle

When a sharp, conductive tip is brought very close (on a scale of nanometers) to a conductive sample, a bias voltage applied between the tip and the surface causes an electrical current to flow. Such a current, called a tunneling current, is of quantum-mechanical origin and can be described by a one-dimensional model[4] as

$$(2.1.1) \qquad I_t = \frac{e}{2\pi h \Delta d^2} [\bar{\phi} e^{-A\bar{\phi}^{1/2}\Delta d} - (\bar{\phi} + eV)e^{-A(\bar{\phi} + eV)^{1/2}\Delta d}]$$

where for $eV \leq \phi$,

$$\Delta d = d, \qquad \bar{\phi} = \phi - eV/2,$$

and for $eV > \phi$,

$$\Delta d = \frac{d\phi}{eV}, \qquad \bar{\phi} = \phi/2,$$

with

$$A = \frac{4\pi}{h}(2m)^{1/2},$$

$-e$ = charge on an electron,

ϕ = average of the barrier height between the two electrodes,

V = bias voltage applied between the tip and the surface,

d = gap distance,

m = 9.1×10^{-28}g (electron mass),

h = Planck's constant.

For a typical gap distance of 1nm and a bias voltage of 0.1v, the tunneling current is on the order of 1nA.

2.1.2 Feedback Loop

Equation (2.1.1) describes the relationship between the gap distance d and the tunneling current I_t, from which the gap distance could be obtained by directly measuring the tunneling current. In reality, this difficult procedure can be circumvented by setting up a feedback loop over I_t. In Figure 2.1, a piezoceramic actuator controls the position of the tip to keep the tunneling current I_t at a pre-set steady-state value I_0: if $I_t < I_0$, the actuator bring the tip closer to the specimen; if $I_t > I_0$, the actuator moves the tip farther away from

the specimen surface. This pre-set tunneling current I_0 corresponds to a gap distance d_0. The amount of correction needed from the actuator indicates the relative height of the point on the specimen surface under the tip. If the actuator controls the probe to raster through a certain area of the surface, the topography of the area is mapped out.

Figure 2.2 shows a schematic diagram of the scanning tunneling microscope system that was designed[6] to perform surface measurements on the nanometer scale. For a positive bias voltage applied between the specimen and the tip, $\Delta V = V_{specimen} - V_{tip} \sim 1.5v$, a tunneling current $I_t \sim 1nA$ (corresponding to a gap distance of about 1nm) is amplified and converted to a voltage $V_t \sim -0.1v$ (to be explained in more detail in section 2.3.2). This voltage V_t is amplified further and digitized in an analog-to-digital converter before being fed into a digital signal processor (DSP). As a stand alone computer designed for real time digital processing, the DSP calculates the correction voltage V_c' required to bring the tunneling current I_t back to the pre-set value I_0. The frequency at which these corrections occur is software adjustable via a personal computer. Then the digital correction V_c' is converted back to analog V_c through a digital-to-analog converter, amplified through a high voltage amplifier to drive the piezoceramic actuator so as to move the probe to regulate the gap distance.

During scanning, the probe moves through the grid points of a prescribed area to render its surface topography. This in-plane motion of the probe is controlled via a personal computer.
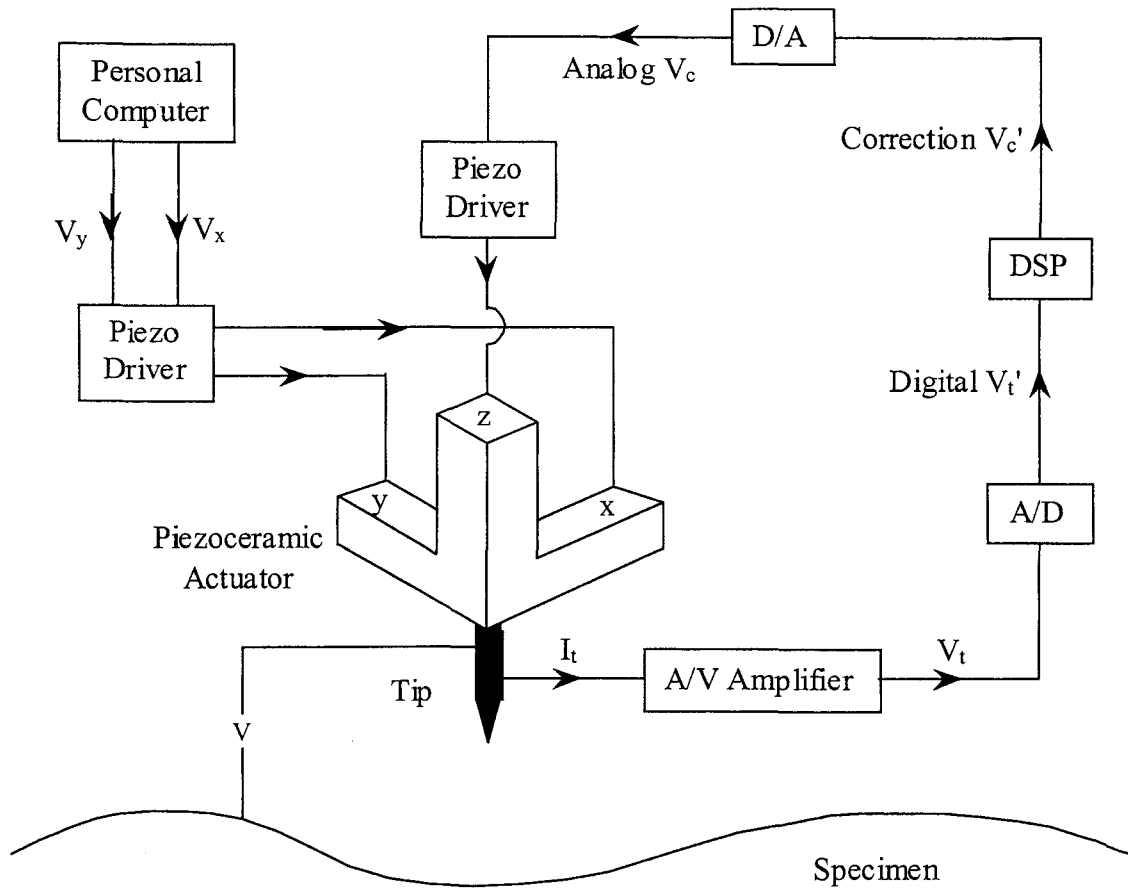
Figure 2.2. STM system schematics

At each of the grid points, the DSP computes the correction through a proportional controller

(2.1.2) $$V_c = V_p + K (I_t - I_0),$$

where

V_c = current correction,

V_p = previous correction,

K = feedback loop gain.

The various parameters involved in the feedback loop, including the frequencies at which data are acquired and corrections are computed, are software adjustable at any time. It is important to note that the feedback loop has the highest priority, because it is the most critical task for the DSP to ensure that the tip does not crash onto the specimen surface. Therefore, the feedback loop frequency F_{fbl} is much higher than the data acquisition frequency F_{acq}. Typically, F_{fbl} is on the order of 100kHz, while F_{acq} ranges from 0 to 3000Hz.

2.1.3   The Piezoceramic Actuator

The precision of the scanning tunneling microscope is determined by how accurately the position of the tip can be controlled with respect to the specimen surface. Therefore, a highly accurate actuator is needed to move the tunneling probe along the three coordinates. For example, in the out-of-plane direction, the gap distance is on the order of 1nm, which requires the actuator to have a resolution of about 10^{-2}nm to regulate the tunneling distance with 1% certainty. In the two in-plane directions, the actuator should be able to achieve a resolution in the nanometer range (1-10nm), because the size scale of the specimen under investigation is 1μm or possibly smaller.

In addition to the fine resolution, other factors need to be taken into consideration as well, namely linear response, repeatability, vibration of the actuator, etc. A solution to satisfy these requirements the best is the piezoelectric or piezoceramic actuator. Piezoceramic material has the characteristics to expand or contract in response to an electric field, with a sensitivity of a few nanometers per volt. Because of the compact size requirement in an

STM system, a tube scanner[6] shown in Figure 2.3 is typically used. It consists of a hollow piezoceramic and nickel-coated tube so as to provide four separate quadrant electrodes on its outside surface and a continuous electrode on the inside.
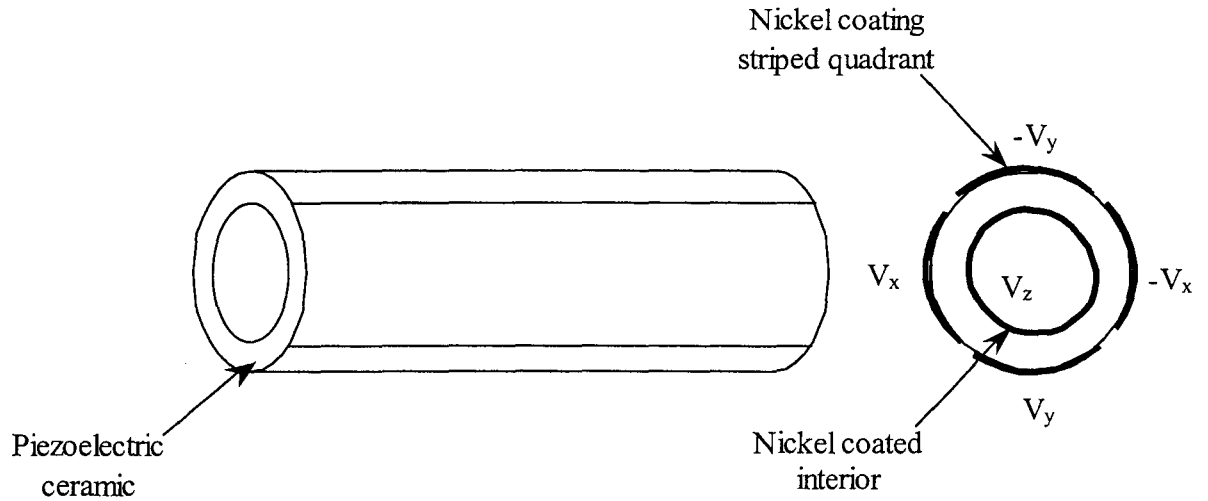
Figure 2.3. Piezoceramic tube actuator

A uniform potential difference V between the inside and all outside electrodes will make the tube expand or contract along its axis, as shown in Figure 2.4.
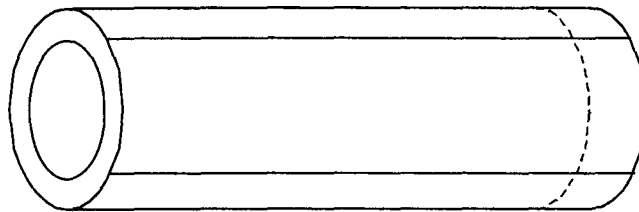
Figure 2.4. Piezoceramic tube elongates when $V_z=V$, $V_x=V_y=0$

Applying voltages V and –V to an opposing pair of electrodes and grounding all other electrodes will make one of the tube quadrants expand and the opposite one contract, so

that the tube is bent as shown in Figure 2.5. The displacement of the points on the centerline is not entirely in-plane, but has a small out-of-plane component. However, the coupling is small compared to the amplitude of the in-plane displacements.
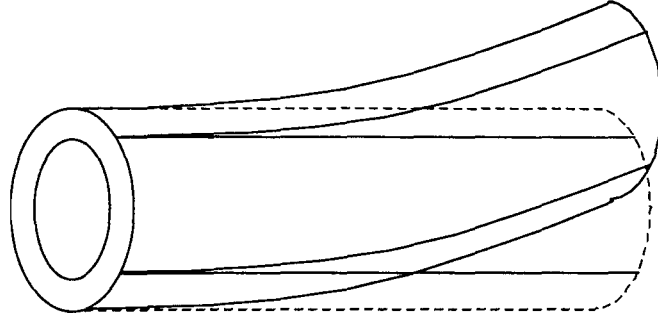
Figure 2.5. Piezoceramic tube bends when $V_x=V$, $V_y=V_z=0$

Mathematically, the motion of the end of the tube is described by

1) Along the axis,

(2.1.3)
$$\Delta Z = 2d_{31}V_z \frac{L}{(OD - ID)},$$

2) In the x-y plane,

(2.1.4)
$$\Delta X = 4d_{31}V_x \frac{L^2}{OD(OD - ID)}$$
$$\Delta Y = 4d_{31}V_y \frac{L^2}{OD(OD - ID)}$$

where

L = length of the piezoceramic tube,

OD = outer diameter of the tube,

ID = inner diameter of the tube,

d_{31} = piezo-electric charge constant, and Vx, Vy, Vz are the voltages applied on the electrodes on the tube as shown in Figure 2.3.

From equations (2.1.3) and (2.1.4), the displacements can be maximized if one can choose the length of the tube large while keeping the two diameters small and d_{31} as large as possible. Under this consideration, piezoceramic tubes made from lead zirconate tinanate have been selected with $d_{31} \approx 250 \times 10^{-12}$ m/V, L = 2", OD = 0.5", ID = 0.44".

2.2 Calibration

After the scanning tunneling microscope was constructed "in-house"[6], it has been used in mechanical investigations for PolyVinylChloride (PVC)[8], fiber/matrix composite materials[8] and thin films[9]. In the study[9] on the compressive properties of thin films such as mica, the property measured using the STM on roughly 500μm thick mica sheets were low by about a factor of two compared to those quoted in the literature. To examine the calibrated coefficients used in the STM system, a calibration is conducted using a standard as shown in Figure 2.6. The specimen* is a precisely fabricated silicon dioxide pitch cluster, coated with a very uniform layer of platinum. The grid pattern consists of an array of alternating bars and spaces with uniform pitch of 10μm width in both x and y direction. The steps have a width of 4μm and a height of 180nm.

---

* Surface Topography Reference (STR10-1800) was purchased from VLSI Standard Incorporated, (408) 428-1800.
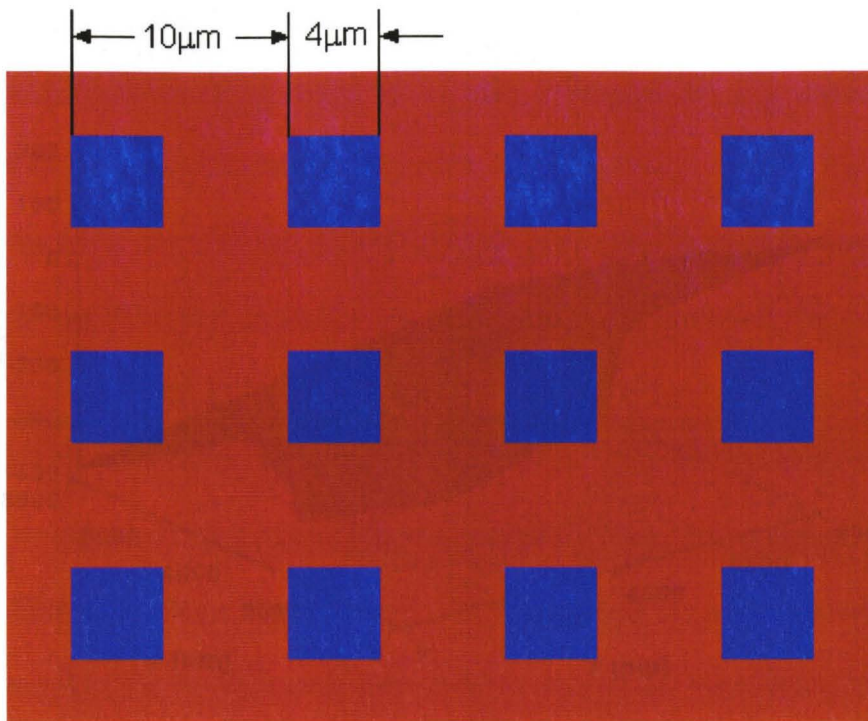
Figure 2.6. Calibration standard

STM scans of the calibration standard on an area of 8µm x 8µm are shown in Figure 2.7 and Figure 2.8.
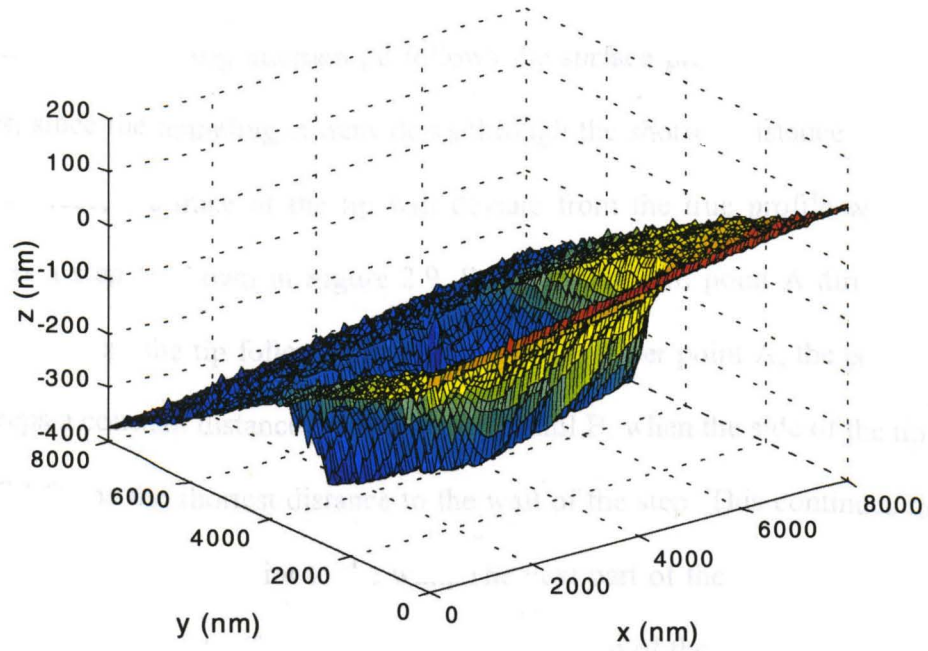
Figure 2.7. Example STM scan of the calibration standard
Note: The inclination appears amplified as a result of the scale difference.
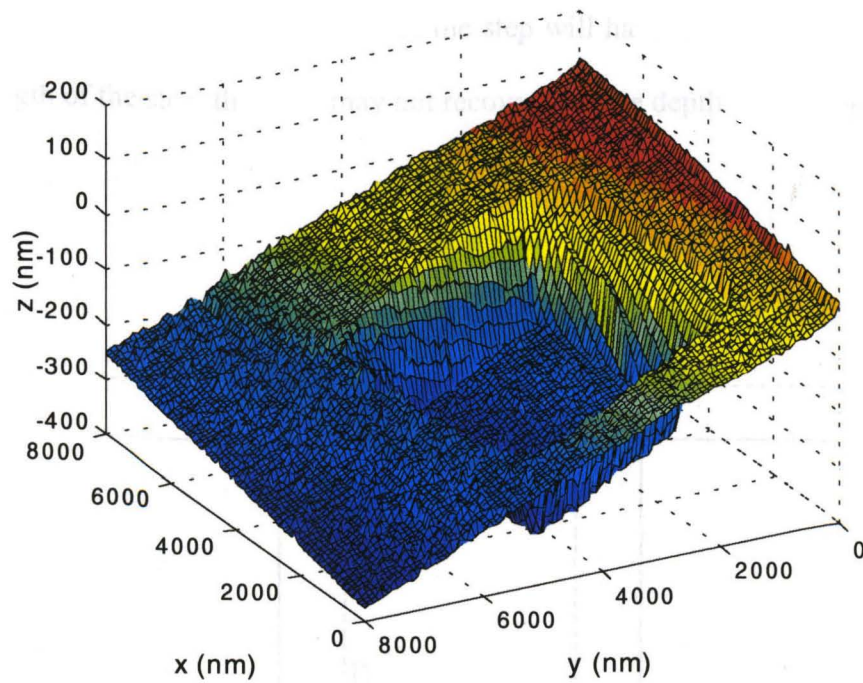
Figure 2.8. A differently oriented view of the STM scan in Figure 2.7

When scanning a smooth surface with little roughness as shown in Figure 2.1, the probe of the scanning tunneling microscope follows the surface profile at a constant distance. However, since the tunneling current flows through the shortest distance between the tip and the surface, the trace of the tip will deviate from the true profile when the probe encounters a step as shown in Figure 2.9. Specifically, until point A directly above the corner of the step, the tip follows the surface profile. After point A, the end point of the tip T_1 keeps a constant distance from the corner until B, when the side of the tip (between T_1 and T_2) forms the shortest distance to the wall of the step. This continues to C where T_2 turns to be the closest point to the wall. The next part of the trace is a vertical line to D, when the end of the tip T_1 comes close to the bottom of the step. After the tip follows the bottom of the step, the ascending trace is the reverse of the descending one. In most cases, the dimension of the tip is larger than the roughness of the surface, thus point C is never reached and the scanned profile of the step will have an inclined wall. Depending on the length of the step, the trace may not recover the true depth of the step.
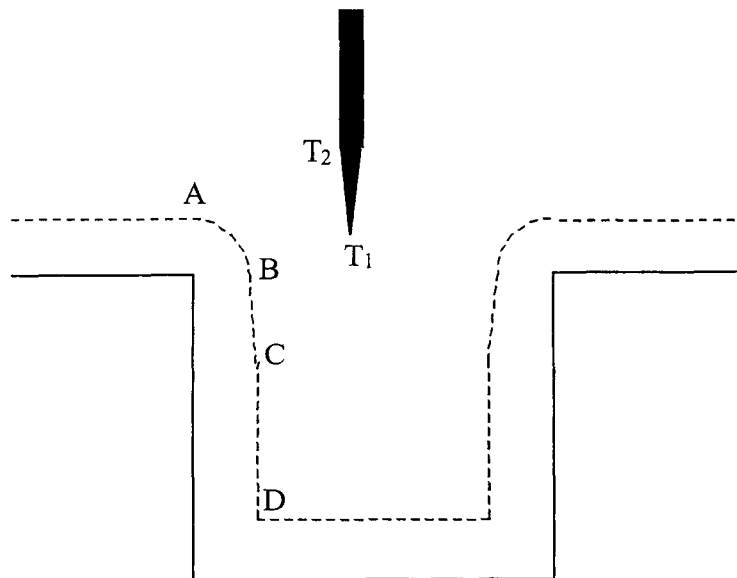
Figure 2.9. Trace of the tip when scanning a step

In practice, the scanned surface is usually inclined to some small degree, as shown in Figure 2.7. This inclination may derive from several sources: positioning of the specimen, misalignment of the piezo-ceramic tube, non-linear behavior of the piezo-ceramic material, etc. Inclination will have an effect on the trace of the tip in addition to the interaction between the side of the tip and the wall of the step as mentioned above. As shown in Figure 2.10, after passing the corner of the step, the tip will move to the next point, which makes the trace move to the bottom of the step. The corner at point D in the scanning trace does not represent the true geometry of the corner, because the real corner is "hidden" and not reachable by the tip. This does not happen during the ascending phase in this example, so that the scanned image is distorted. If the surface inclines the other way, the same phenomenon will be observed and the "hidden" corner appears at ascending.
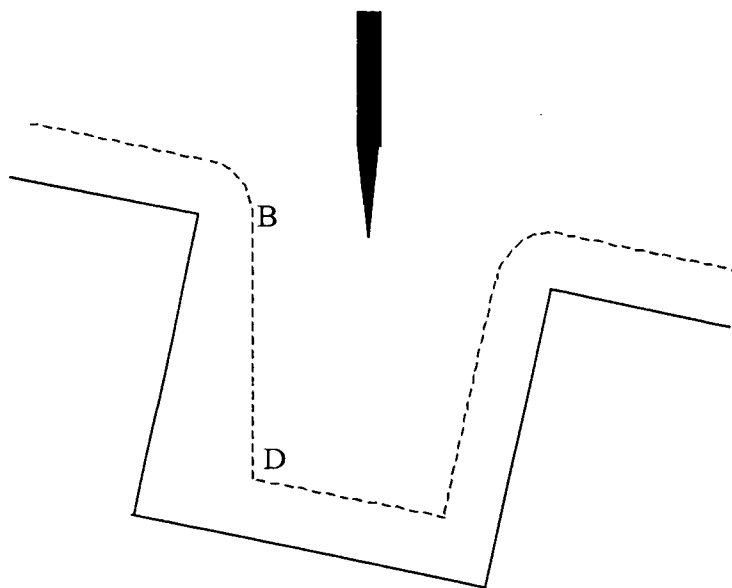
Figure 2.10. Trace of the tip when scanning an inclined step

For illustration purpose, the inclined image shown in Figure 2.7 is rotated in Figure 2.11, so that the top surface is more or less parallel to the scanning plane.
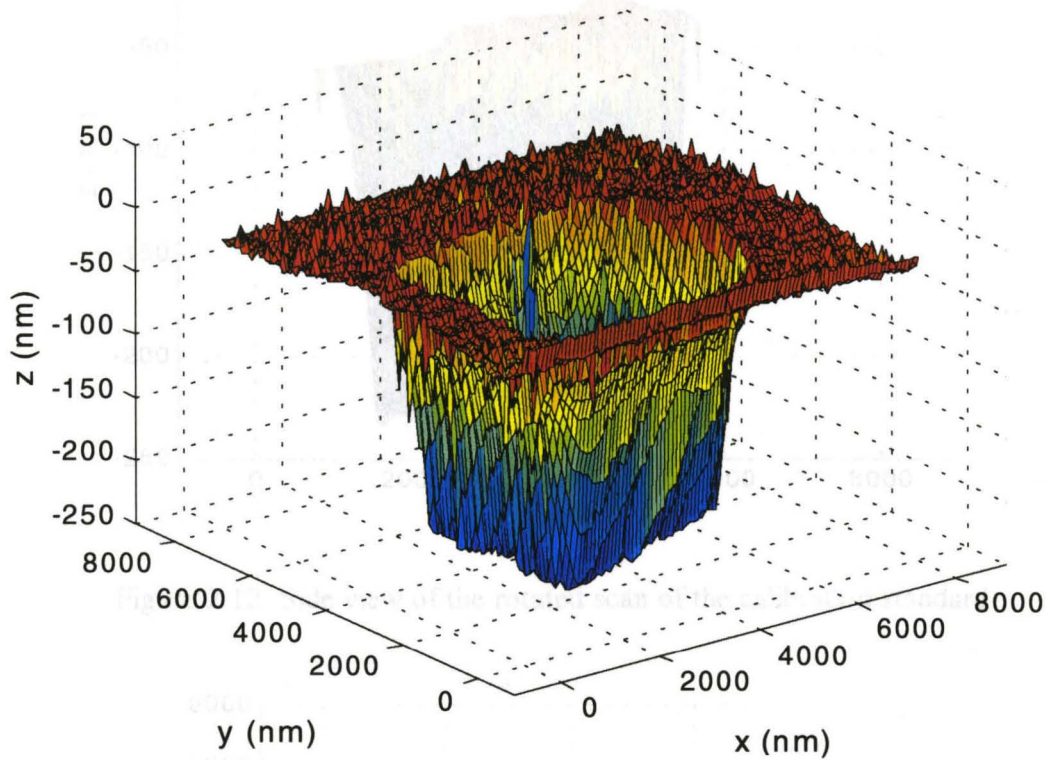
Figure 2.11. Rotated scan of the calibration standard

A side view and a top view are shown in Figure 2.12 and Figure 2.13, respectively. From the side view, the step height is estimated to be 190±5nm. From the top view, the step width is determined to be 4.3±0.1μm in the direction that forms a small angle with the x axis and 4.0±0.1μm in the perpendicular direction.
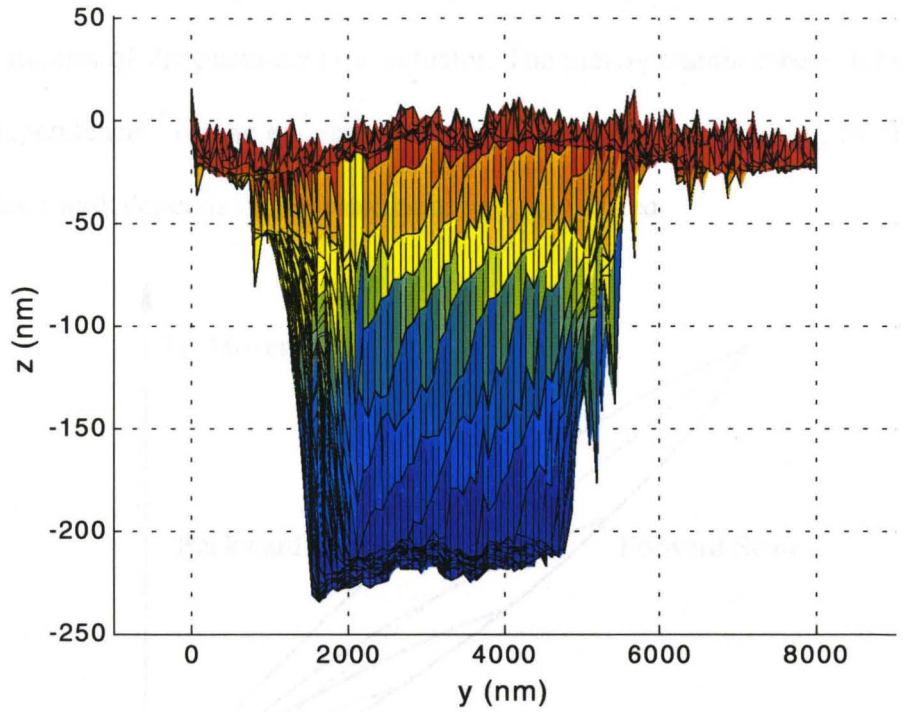
Figure 2.12. Side view of the rotated scan of the calibration standard
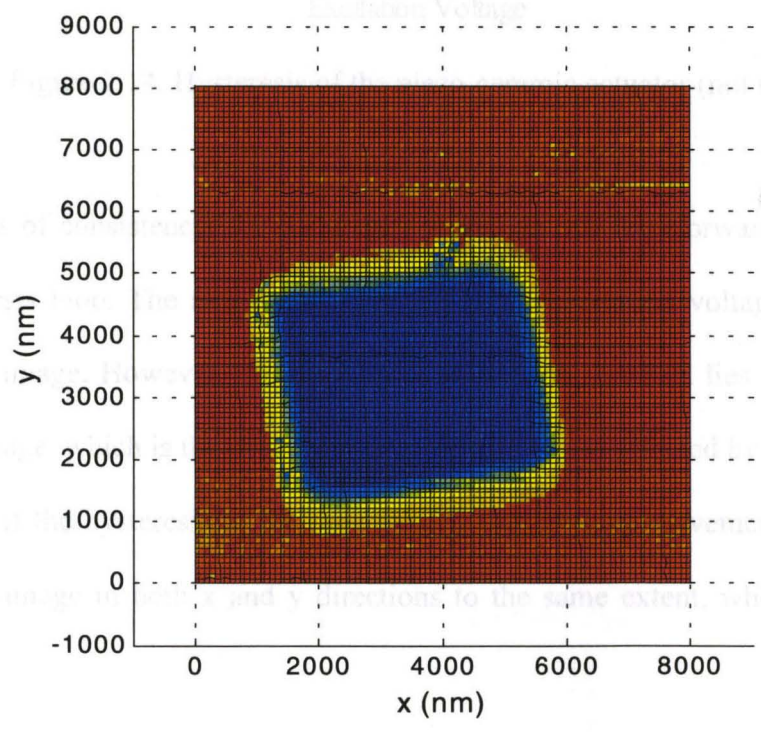
Figure 2.13. Plan view of the rotated scan of the calibration standard

Figure 2.13 shows a slightly distorted square step. A possible source of this distortion is the hysteresis of the piezo-ceramic actuator. The piezo-ceramic tube exhibits a hysteresis type dependence[6] on the excitation voltage as illustrated in Figure 2.14. The size of the hysteresis look depends on the maximum voltage applied.
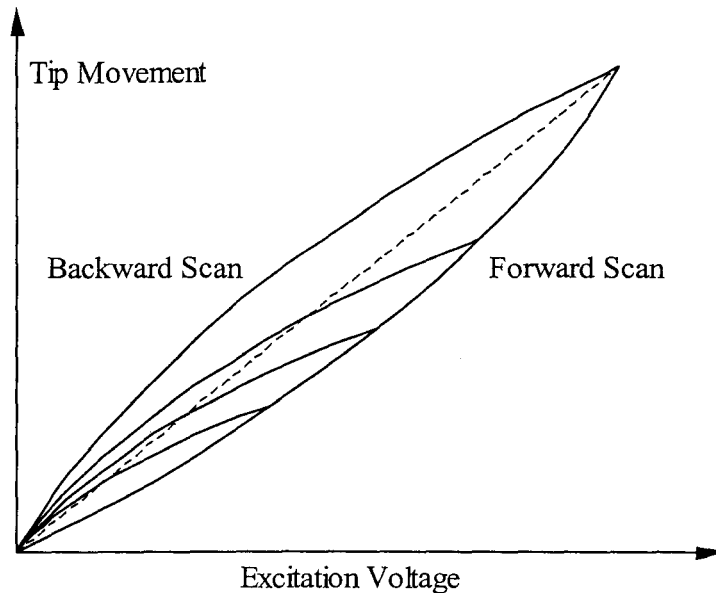
Figure 2.14. Hysteresis of the piezo-ceramic actuator (not to scale)

For reasons of consistency, the STM only works on one leg (forward in our system) of the hysteresis loop. The non-linear response to the excitation voltage on each leg will distort the image. However, the step in the calibration standard lies in the center of the scanned image, which is the portion where linearity is less affected by the hysteresis loop. Moreover, if the hysteresis of the tube distorts the in-plane movement of the tip, it will distort the image in both x and y directions to the same extent, which is not observed here.

Closer investigation of the problem points to a drift problem as the source of the distortion. During tunneling, the tunneling current I_t is amplified and converted to a voltage V_t that is displayed on the computer. When the tip rests, this displayed voltage does not remain constant, but changes monotonically and slowly with time. Because the image acquisition is accomplished in the y direction first as shown in Figure 2.15, it takes longer to scan across the step in the x direction. Therefore, the drift elongates the step in the x direction, but has little influence on the y direction. Unfortunately, the source of this drift-induced noise could not be determined definitely. It might come from the thermal expansion of the probe. If it does, all materials used in the probe need to be carefully selected, to ensure that they have an extremely low thermal expansion coefficients (around 10^{-6}pp/°F).
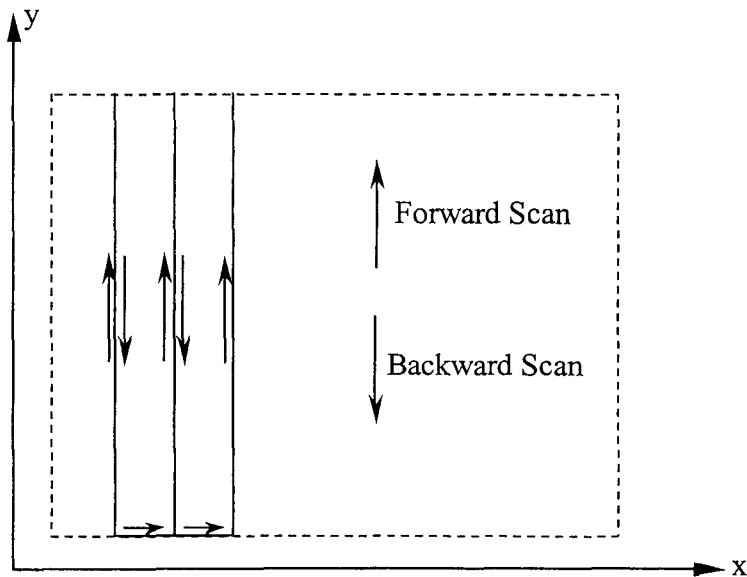
Figure 2.15. Forward and backward scan sequence

In conclusion, the calibration confirmed that the coefficients used in the STM system are accurate within the expected error range. The factor of two in the property measurement

performed on mica sheets[9] is likely to have resulted from calculation errors. Although some problems associated with the hysteresis of the piezo-ceramic actuator and with the drift induced noise remain, the STM has the capability to perform surface measurements with sub-micron resolution.

2.3 Hardware Improvement

The calibration shows that although the STM has the capability to perform surface measurements, its precision can be further improved. The hardware improvement is accomplished through the following components of the STM system: the piezo-ceramic actuator probe, the first stage amplifiers and the feedback controller.

2.3.1 A New Actuator Probe

The probe in the current STM system consisted of a single piezo-ceramic tube, which controls the movement of the tip in the out-of-plane z direction as well as in the two in-plane directions, as described in section 2.1.3. However, there exists a coupling between the in-plane and out-of-plane motion of the piezo-ceramic actuator, which depends on the three voltages imposed on the tube. To reduce the coupling, a new design configuration presented in Figure 2.16 is adopted. The improved scanning system decouples the out-of-plane motion from the in-plane motion. It consists thus of not one but two concentrically arranged piezo-ceramic tubes: the outer tube accommodates only the up and down movement of the tip, while the inner tube, similar to that used in the existing design, is

used exclusively for the in-plane scanning movement, because its higher length-to-width ratio (4 as compared to 1.6 for the outer tube) provides less error for the bending of the tube.
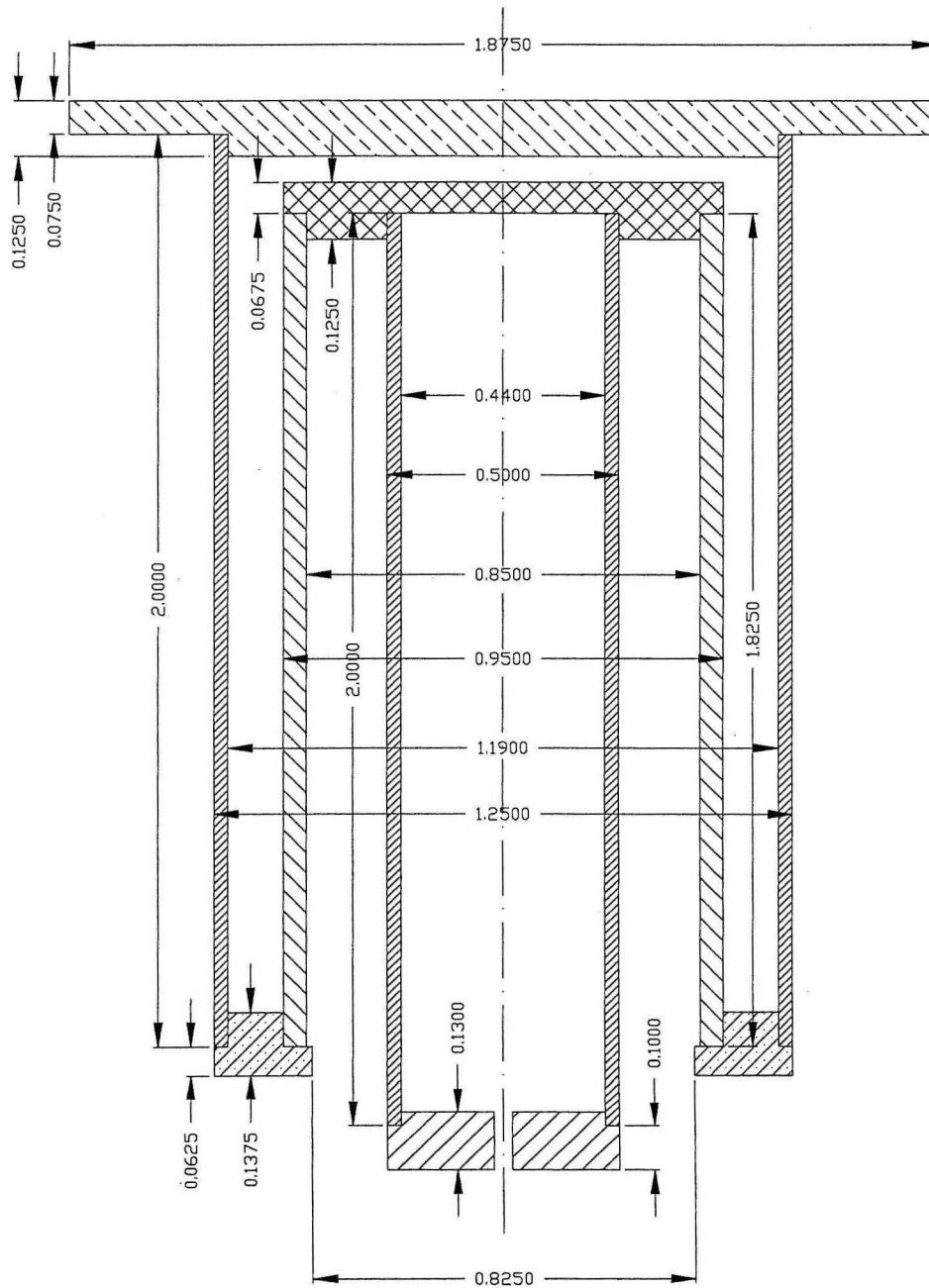
Figure 2.16. Schematic of new STM actuator (to scale, dimensions in inch)

In Figure 2.16, the two tubes with [hatch] hatch are the piezo-ceramic actuators, while all the other parts are made of ceramic that has a thermal expansion coefficient very close to that of the piezo material.

2.3.2   The Current-to-Voltage Converter

After a tunneling current (about 1nA) is obtained, it needs to be converted to a voltage and amplified to approximately 0.1V before being digitized in the analog-to-digital converter. Since the amplification is very high (from 1nA to 0.1V), it is crucial to keep the signal-to-noise ratio as high as possible, because any noise introduced at this stage will be amplified in the operational amplifiers downstream in the electronics system.

2.3.2.1 Design of a New Current-to-Voltage Converter

The previous design for the current-to-voltage converter, as shown in Figure 2.17, is a simple inverting amplifier without filter. The disadvantage of such an amplifier is that when the impedance is high (e.g., in this case), it readily picks up environmental noise like an antenna.
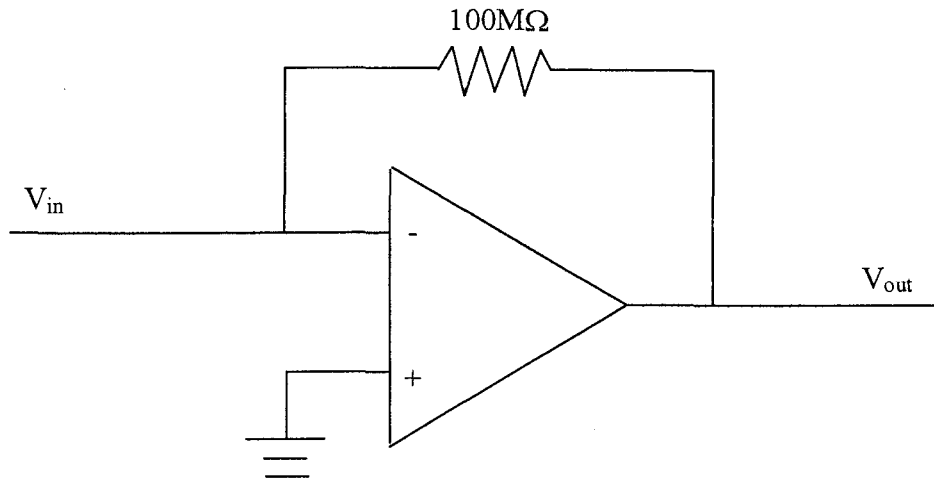
Figure 2.17. Previous design for the current-to-voltage amplifier

To reduce the noise introduced at this stage, a three-stage amplifier shown below is designed to replace the previous one-stage component.
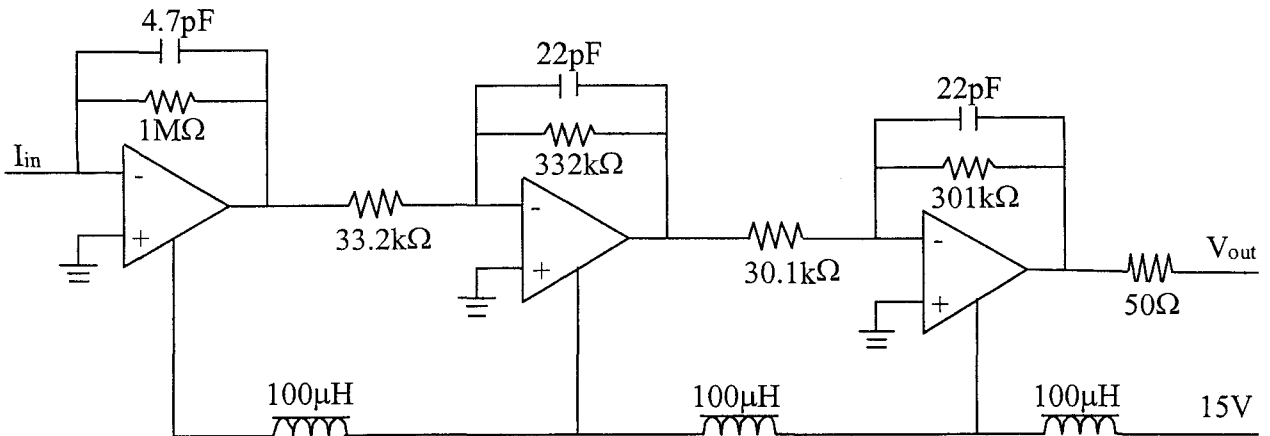
Figure 2.18. New design for current-to-voltage converter

Each stage serves as a low-pass filter to sort out high frequency noise. A low-pass filter reduces the AC signal according to the formula[10]

(2.3.1)
$$\frac{V_{out}}{V_{in}} = \frac{1}{\sqrt{1 + \omega^2 R^2 C^2}} \, .$$

The cut-off frequency corresponding to the –3dB breakpoint is defined by

$$(2.3.2) \qquad f = \frac{1}{2\pi RC} .$$

For each of the stages, the cut-off frequency is

$$f_1 = \frac{1}{2\pi R_1 C_1} = 33.86\text{kHz} ,$$

$$f_2 = \frac{1}{2\pi R_2 C_2} = 21.79\text{kHz} ,$$

$$\text{and } f_3 = \frac{1}{2\pi R_3 C_3} = 24.03\text{kHz} .$$

For example, when the input signal has a dominant frequency of f=67kHz,

$$\frac{V_{out}}{V_{in}} = \frac{1}{\sqrt{1+(\frac{f}{f_1})^2}} \frac{1}{\sqrt{1+(\frac{f}{f_2})^2}} \frac{1}{\sqrt{1+(\frac{f}{f_3})^2}} = 0.0477 ,$$

i.e., the output signal at this frequency drops to approximately 5% after the three-stage "amplification."

Besides the environmental high frequency noise, there also exists Johnson thermal noise, or white noise. An open-circuit noise voltage generated by a resistance R at temperature T is given[10] by

$$(2.3.3) \qquad V_{noise} = \sqrt{4kTRB} ,$$

where

$k = 1.38 \times 10^{-23}$ J/K, Boltzman's constant,

T = 293K, room temperature,

B = bandwidth.

For the amplifier shown in Figure 2.18 with R=1MΩ, B=34kHz, the Johnson noise at the output is estimated to be 23.45μV. Although Johnson noise at any instant is in general unpredictable and follows a Gaussian distribution with an rms value at V_{noise}, its amplitude is thus small enough to be negligible.

2.3.2.2 Circuit Layout and Configuration

In the circuit design, another important consideration is that noise coming from the power supply must be maximally filtered.
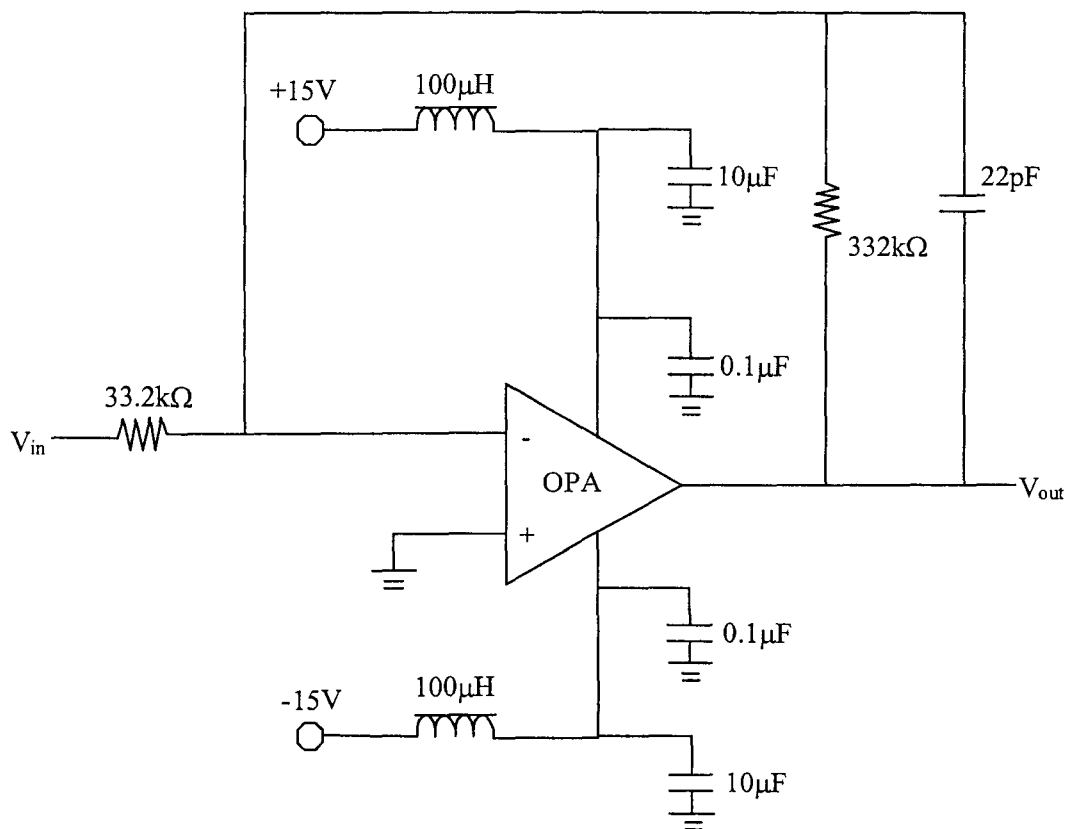
Figure 2.19. Circuit design for second-stage amplifier

This is accomplished with the aid of capacitors. An example circuit design for the second stage amplifier is shown in Figure 2.19.

In the existing setup, the operational amplifier was mounted in an aluminum box to reduce the environmental noise. This aluminum box was then mounted on a cylinder base made of invar, an iron-nickel alloy with a very low thermal expansion coefficient and good stiffness and machinability properties. To accommodate the newly designed three-stage amplifier into the same aluminum box, three circuit boards are used in a stack and fixed between the two ends of the box.

The tunneling current sensed at the tip passes through three amplifiers in sequence. The output from the last amplifier is digitized in the A/D converter. The voltage from the $\pm15V$ power supply is introduced to each circuit board. The excitation voltage to activate the piezo-ceramic tubes is insulated in brass tubing and passes through the aluminum box to the piezo actuators.

Each amplifier is placed on a small, round circuit board. For the purpose of optimal shielding, two-sided copper coated clay boards are used. While the top side of each board is etched to incorporate the circuit, the bottom side is untouched. Thereby the copper coatings and the wall of the aluminum box create shielding for each layer of the amplifier. The aluminum box is grounded from the outside.

The circuit boards are etched out in our laboratory. The etching procedures are described in Appendix A.

2.3.2.3 Result

We examine the performance of the new amplifiers by feeding an input D.C. voltage of 0.1V through a resistor of 100MΩ (which simulates the impedance between the tip and the specimen during tunneling) and measuring the gain as illustrated in Figure 2.20.
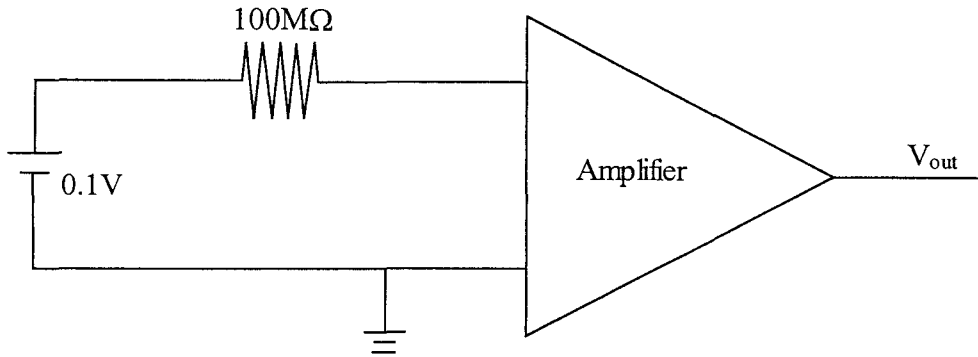
Figure 2.20. Schematic for amplifier testing

The amplifier responds consistently with an output of –0.1V and a noise level of 0.01V. In comparison, the earlier single stage amplifier performed worse in the same test – reading noise level in the output reach as high as 0.5V. Thus, the new three-stage amplifier has reduced the noise level by at least a factor of 10.

2.3.3   Modeling of the STM Controller

The STM is operated in a feedback loop to keep the tip at a constant distance from the specimen, which functions as a proportional controller. In the feedback control system for the STM shown in Figure 2.2, the dynamic behavior of every component is known except that for the piezo-ceramic actuator. How fast the piezo-ceramic tubes respond to a given voltage is crucial in determining the various parameters for the controller. Therefore, the

first step is to investigate and possibly model the control characteristics of the piezo-ceramic actuator.

There is no documentation on the response time of the piezo-ceramic actuator, nor could the manufacturer give definite information. To determine the dynamic behavior of the piezo-ceramic actuator, a frequency response test is conducted: sinusoidal voltage inputs with different frequencies are fed into the interior electrode of the piezo tube V_z while keeping the outer electrodes grounded $V_x=V_y=0$ as shown in Figure 2.4, and the magnitude and phase shift of the sinusoidal output are measured. Due to the difficulty in measuring small deformations of the piezo-ceramic tube directly, this test is conducted during tunneling and the voltage output after the current-to-voltage converter, or V_t as shown in Figure 2.2, is measured instead. In other words, the frequency response test is conducted jointly on the piezoceramic tube and on the current-to-voltage converter. Figure 2.21 shows its BODE diagram[*] of magnitude ratio and phase shift with respect to the frequency of the sinusoidal voltage input.

---

[*] This frequency response plot is named after H.W. Bode for his contribution in synthesizing networks with given characteristics. For more information about BODE diagram and control theory in general, refer to reference [12].
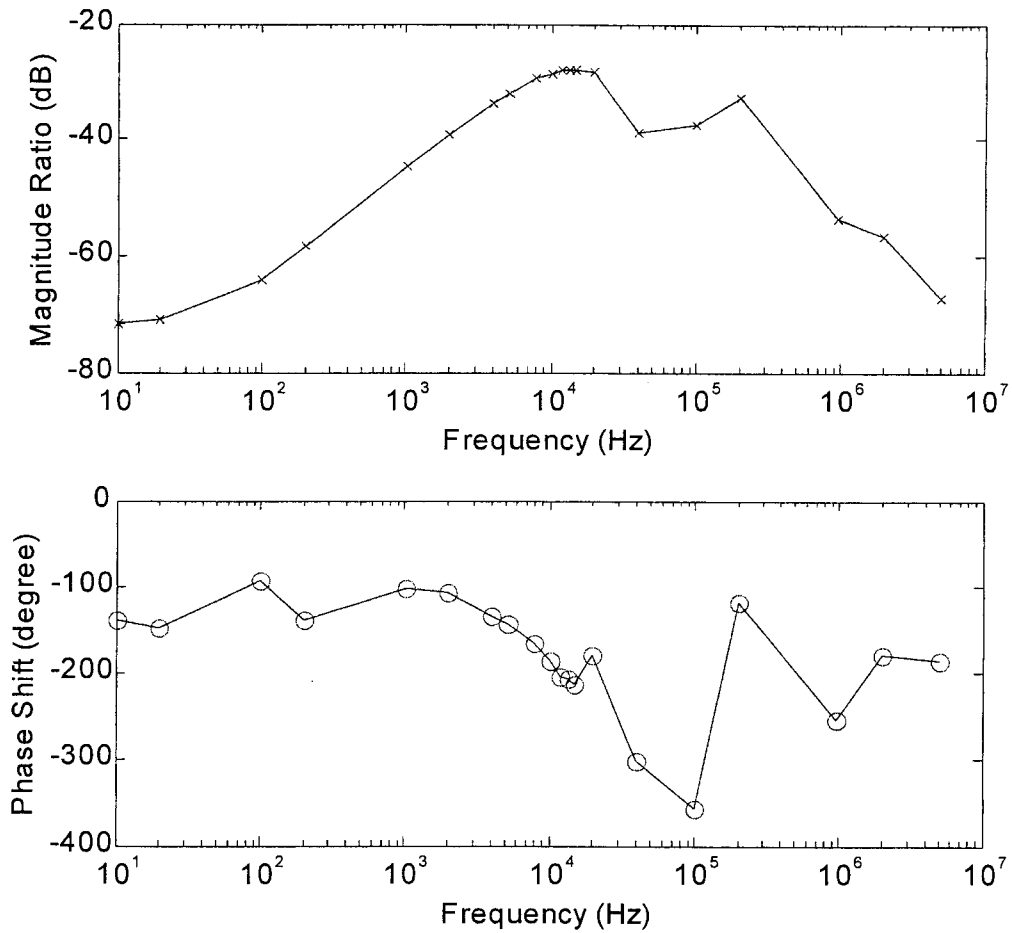
Figure 2.21. BODE diagram of piezo-ceramic actuator based on experimental data

Based on this plot, a mathematical model controller with a transfer function

$$(2.3.4) \qquad F(s) = \frac{2000(s+40)(s+5\times10^4)^4}{(s^2+1.2\times10^5 s+4\times10^{10})(s+1.5\times10^4)^4}$$

is assumed in an attempt to simulate the experimental data. The coefficients are determined by trial and error and the results are as shown in Figure 2.22. The analytical model represents the real controller fairly well.
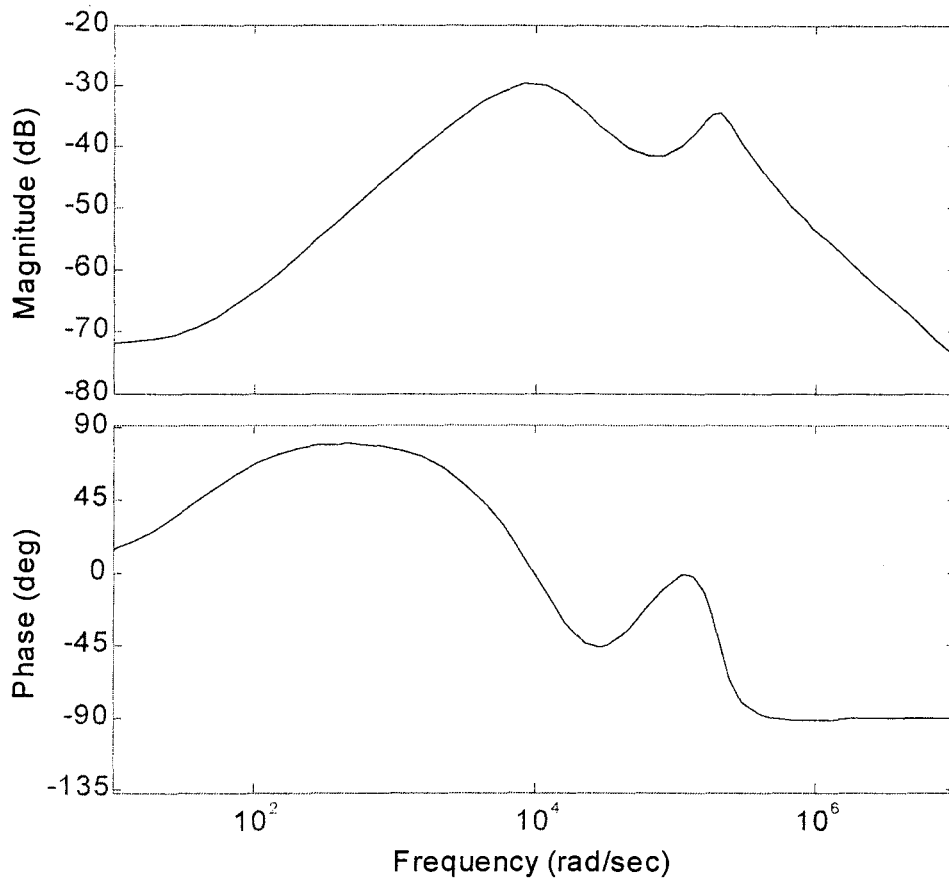
Figure 2.22. BODE diagram of model controller

Given this mathematical model controller, its step response can be simulated in Matlab, as shown in Figure 2.23.
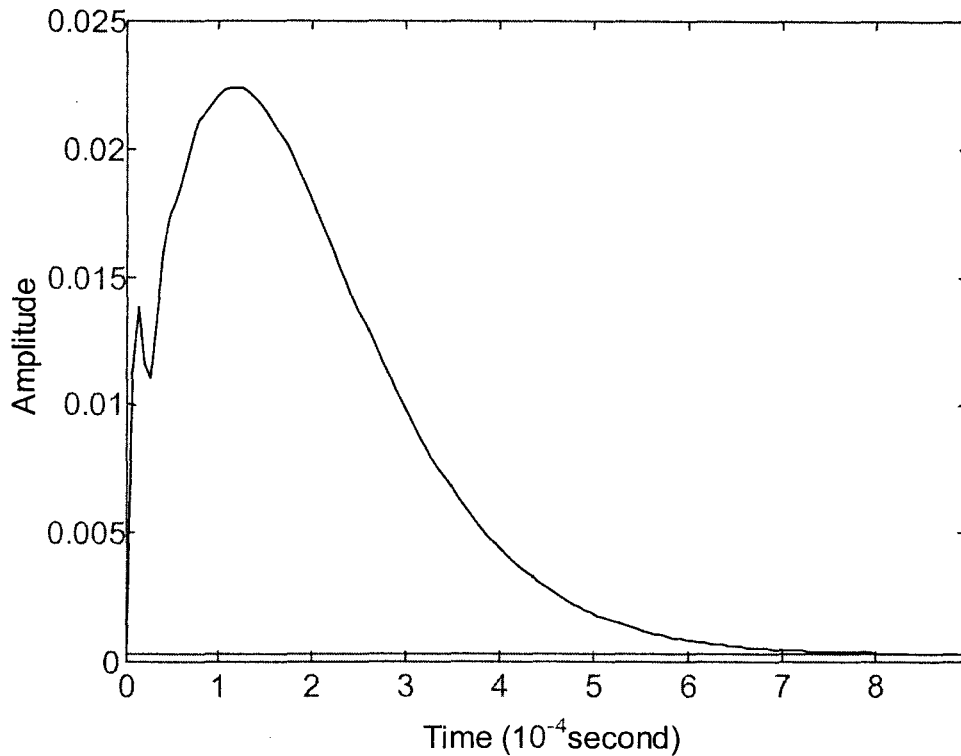
Figure 2.23. Step response of the model controller
Note: non-zero steady-state value

This figure shows that the estimated response time for a step input is approximately 0.5ms. Because the current-to-voltage converter actually responds about 10 times faster, this result represents essentially the response time of the piezo-ceramic actuator.

In date acquisition, the excitation voltage to activate the piezo-ceramic actuator changes at the feedback frequency. To examine the piezo actuator's response at different input frequencies, a periodic pulse with adjustable frequency is fed into the piezo actuator and the response is shown in the following plots.
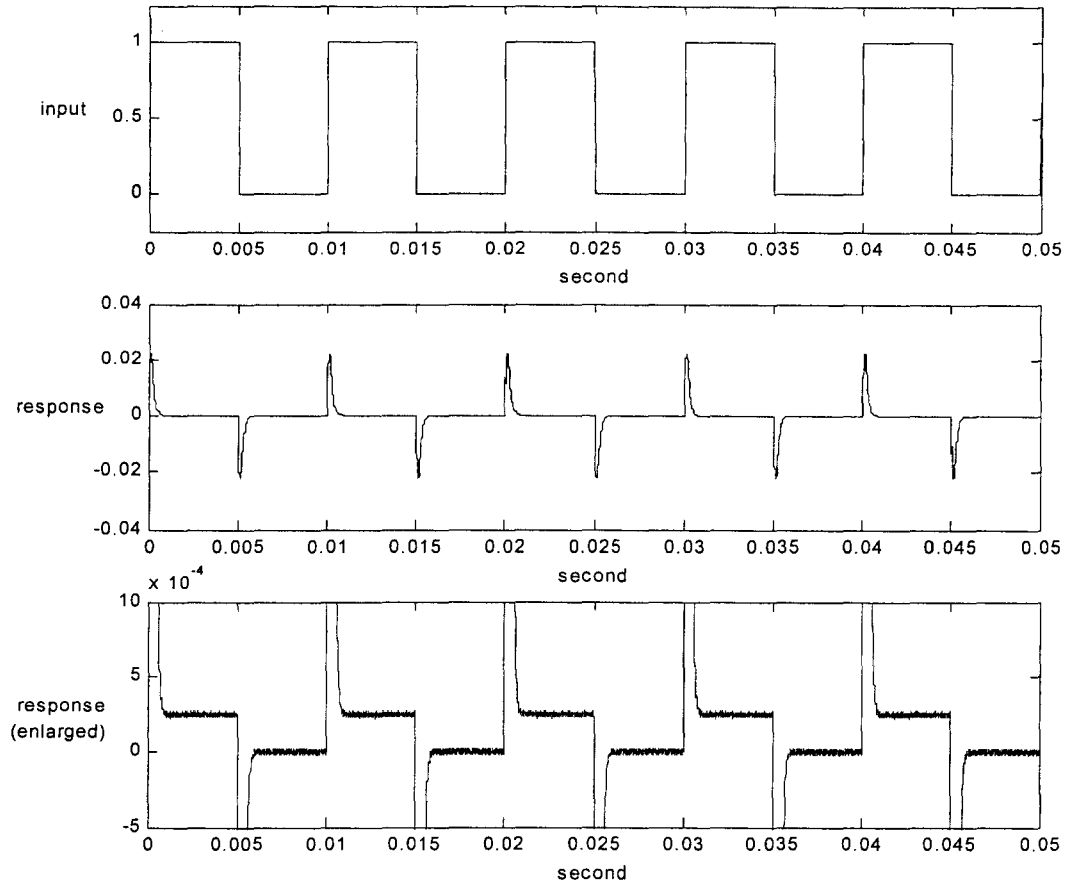
Figure 2.24. Response of actuator model to a pulse input with 100Hz frequency
Note: different scales on the vertical axis for the last two plots

At 100Hz input, the actuator model overshoots before attaining a steady state value (around 2.5×10^{-4}) in a little less than 1ms. The response to a pulse sequence agrees with the step response shown in Figure 2.23. As the input frequency increases beyond 100Hz, there is insufficient time for the actuator to reach the steady state value and the response starts to take various shapes, as shown in Figure 2.25 for a 1kHz input and Figure 2.26 for a 10kHz input.
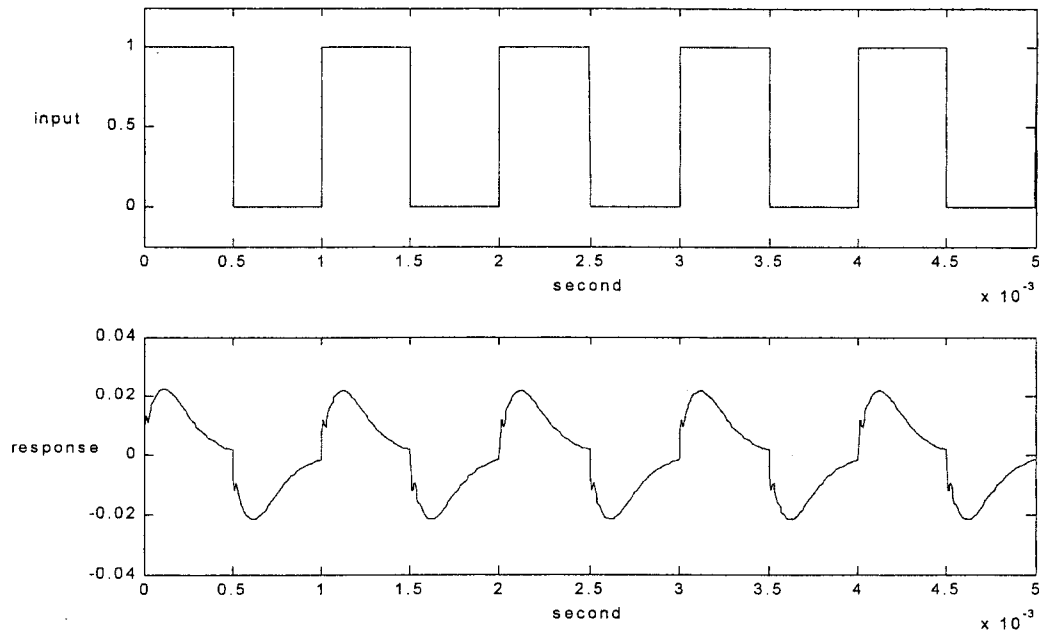
Figure 2.25. Response of actuator model to an input with 1kHz frequency
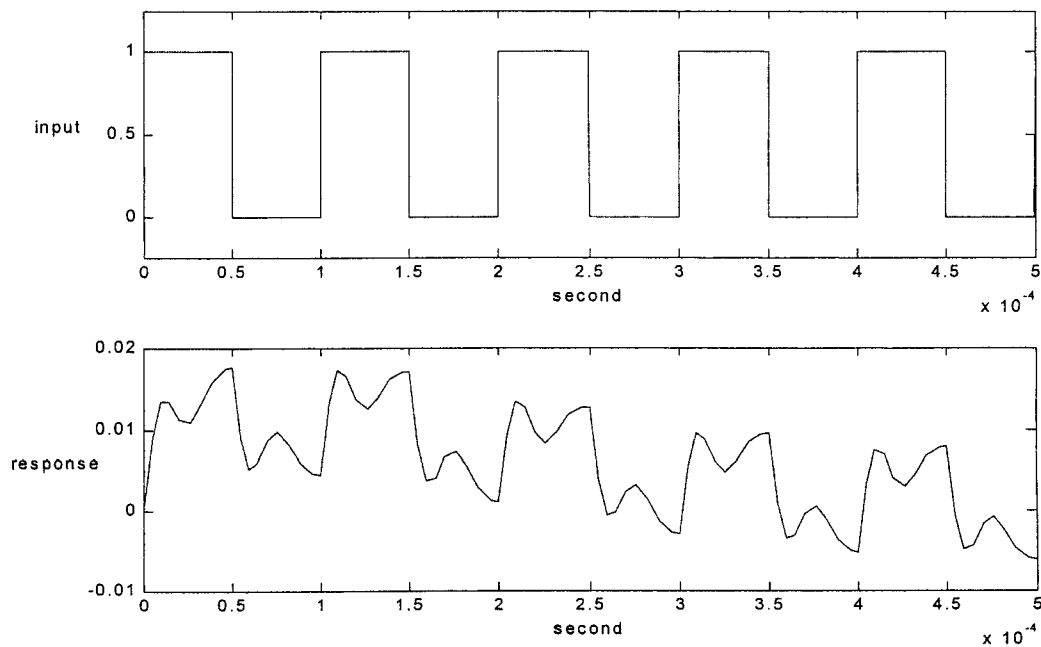
Figure 2.26. Response of actuator model to an input with 10kHz frequency

When the input frequency reaches 100kHz, the response to a square wave becomes a triangular wave. It even takes about 0.5ms for the wave form to stabilize.
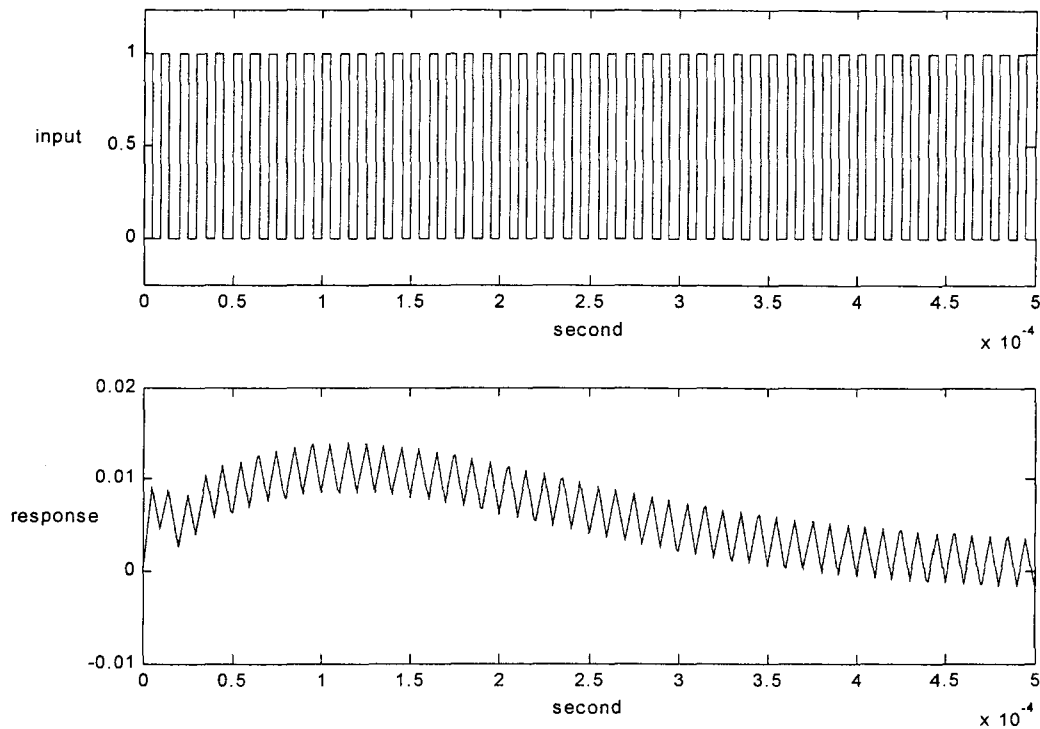
Figure 2.27. Response of actuator model to an input with 100kHz frequency

All these plots suggest that if the linear controller model represents the dynamic behaviors of the actuator, the current feedback frequency set at 100kHz is clearly too fast for the piezo-ceramic actuator to respond. However, the real actuator is unlikely to behave as shown at these high frequencies. For instance, reducing the feedback frequency has not improved the quality of the scan significantly. On the contrary, due to the drift-induced noise, a longer scan has usually resulted in more distortion. Moreover, the real controller is unlikely to be strictly linear, while the mathematical model is a linear controller. In addition, there always exists noise from electronics at various points in the

feedback loop, which is likely to have a major influence on the performance of the controller. Therefore, the modeling of the STM controller in this section is used only for estimation purpose.

2.4 Conclusion

This chapter summarizes the application of a "home-built" scanning tunneling microscope for mechanics investigations. The typical tunneling distance (of approximately 1nm) between the tip and the specimen is much less than the average roughness of most materials. To keep the tip from interfering with surface features during scanning, the STM is operated in a feedback loop such as to keep the probe at a constant distance from the surface. The feedback correction to maintain this constant distance (or tunneling current) indicates the height at the scanning point on the surface. The movement of the probe is controlled by piezo-ceramic tube actuators, the bending and length change of which make the probe tip move in-plane and up and down, respectively.

A calibration of the tip motion has been conducted using a pitch calibration standard. Within the error range, calibration results agree with the dimension of the standard, thus confirming the capability of the scanning tunneling microscope to perform surface measurements at the nanometer scale.

Based on the previous STM design, several improvements have been made to further heighten its precision.

- To decouple the in-plane and out-of-plane motion of the piezo-ceramic tube actuator, a new design composed of two concentric tubes is *implemented*. The

inner tube, similar to that used in the existing STM, is dedicated to the scanning motion of the tip, while the outer tube controls the up and down motion of the tip exclusively.

- The first-stage amplifier that converts the tunneling current to voltage has been redesigned to incorporate three amplifiers. The new amplifiers filter out high frequency noise and have reduced the overall noise level in the output by a factor of 10. To keep the translation stage intact, the new amplifiers are fitted into the existing setup in stack form.

- To access the performance of the feedback controller, the dynamic behavior of the piezo-ceramic actuator has been modeled by using a mathematical, linear controller model. Testing of the linear model controller shows that the piezo-ceramic actuator has a response time on the order of 0.5ms, which is slow compared to the feedback frequency of the current setting. However, this model may not fully represent the piezo-ceramic actuator, and thus can only be used as a reference.

3 Digital Image Correlation

Scanning tunneling microscopy offers the capability of performing surface measurements with nanometer resolution. In terms of mechanics, a post-processing technique is needed to extract the deformations from the undeformed and deformed surface measurements with sufficient precision. Because of the high-resolution requirement, many of the traditional optical methods cannot be used. Instead, a two-dimensional digital image correlation method has been explored to determine the displacements and strains from the undeformed and deformed configurations.

This method was originally proposed by Peters, Ranson and Sutton[13]-[16] in 1982 and has been used on laser speckle images to obtain deformations. Later this technique was extended to measure 3-D displacement field through stereo imaging[17]. In 1990, Vendroux and Knauss[18]-[19] adapted this method to identify displacements from topographical images obtained with a scanning tunneling microscope. In the latter work, a least square correlation coefficient for the undeformed and deformed configurations was introduced and minimized to obtain the displacements and displacement gradients. The technique has proven to be effective. However, more quantitative measures are needed to assess the precision of the method. The following study serves that purpose.

We start with the study of digital image correlation in one dimension rather than two dimensions, for the sole purpose of understanding and investigating the precision and limitation of the method with respect to the various parameters involved. In the first part of this chapter, such a complete one-dimensional study is performed. Many of the

parameters involved in the correlation process are evaluated and governing parameters are identified. Especially, high frequency noise is introduced and its influence on the correlation results is examined. A two-dimensional study follows then along the same lines. Key parameters to the correlation process are identified and compared with the one-dimensional study. The influence of high frequency noise is analyzed in more detail and a calibration method is proposed for experimental work.

3.1   One-Dimensional Digital Image Correlation

In one dimension, digital image correlation deals with line instead of surface profiles in two dimensions. The one-dimensional correlation is first derived theoretically, based on the assumption that the deformation is sufficiently small to preserve the characteristic features of the profile. The method is then tested out for the simple case of a continuous sinusoidal signal subject to linear in-line deformation. Once the one-dimensional method has been proven useful, several parameters are considered and their influence on the correlation is examined. For example, the size of the subset chose for the correlation has been recognized by researchers[13]-[16] as a key parameter, but its effect has not been studied closely. In addition to the in-plane deformation, out-of-plane deformation is superposed to illustrate the appropriate displacement representation to be used in the image correlation. To examine correlation on signals with complex frequency content, continuous line profiles with multiple frequencies are explored. Results agree with those for single frequencies. Furthermore, high frequency, low amplitude noise is added to the continuous signal to study the error associated with the introduction of noise. In addition

to continuous signals, discrete data taken from real scans are used and interpolation is applied. Correlation for non-linear in-plane deformations (inhomogeneous strains) applied is examined and correlation results are analyzed. Finally a summary of the important parameters is given and quantitative measures for a better digital image correlation are proposed.

3.1.1 Theoretical Background

Let F be a point of coordinate x in an undeformed one-dimensional configuration and let f(x) be the height of the profile at F.
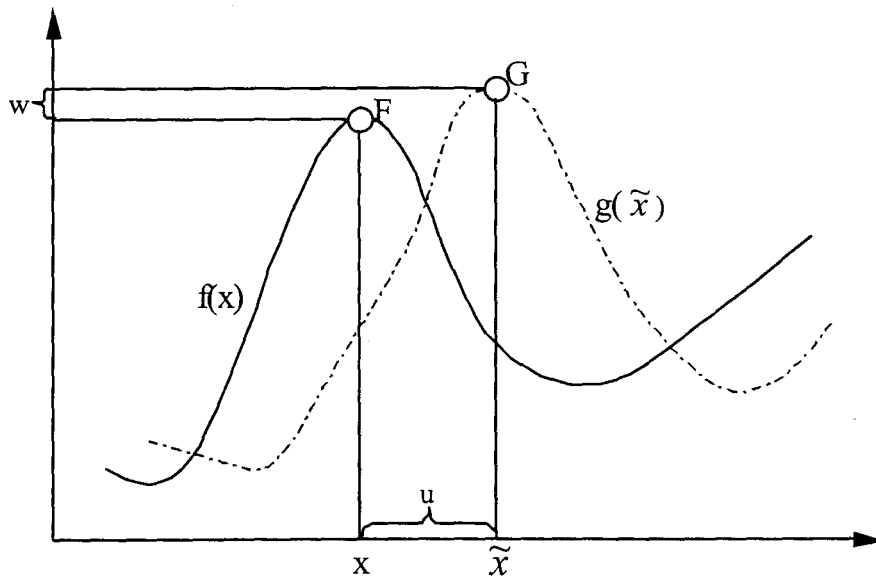
Figure 3.1. Illustration of one-dimensional image correlation

After deformation, the point is mapped into $G = g(\widetilde{x})$, where

$$(3.1.1) \qquad\qquad \widetilde{x} = x + u(x)$$

with u the in-plane displacement of F, and

(3.1.2) $$g(\tilde{x}) = f(x) + w(x)$$

with w denoting the surface-normal displacement.

Let F_0 of coordinate x_0 be mapped to G_0 of coordinate $\tilde{x}_0$ and S be a subset around point F_0. Assuming that S is sufficiently small, the deformation equation (3.1.1) can be expanded as

(3.1.3) $$\tilde{x} = x + u(x_0) + \frac{du}{dx}(x_0)(x - x_0) + \frac{1}{2}\frac{d^2u}{dx^2}(x_0)(x - x_0)^2 + \dots.$$

We shall first deal with situations requiring terms to first order.

Define a least square coefficient C on the subset S by

(3.1.4) $$C = \frac{\int_S [f(x) - g(\tilde{x})]^2 \, dS}{\int_S f^2(x) dS}.$$

It is obvious that C is zero when the mapping is exact. Assuming linear deformation within the subset and considering only out-of-plane translation, the coefficient at point F_0 can be rewritten as

(3.1.5) $$C(x_0) = \frac{\int_S \{f(x) + w(x_0) - g[x + u(x_0) + \frac{du}{dx}(x_0) * (x - x_0)]\}^2 \, dx}{\int_S f(x)^2 \, dx}.$$

Define the three-dimensional vector

(3.1.6) $$P(x) = \{u(x), \frac{du}{dx}(x), w(x)\},$$

so that C=C(x, P) or equation (3.1.5) becomes

$$(3.1.7) \qquad C(x_0, P) = \frac{\int_S \{f(x) + P_3(x_0) - g[x + P_1(x_0) + P_2(x_0) * (x - x_0)]\}^2 \, dx}{\int_S f(x)^2 \, dx}.$$

Expanding C(P) as a truncated Taylor series around P_0 leads to

$$(3.1.8) \qquad C(P) = C(P_0) + \nabla C(P_0)^T (P - P_0) + \tfrac{1}{2}(P - P_0)^T \nabla\nabla C(P_0)(P - P_0).$$

Differentiating equation (3.1.8) with respect to P on both sides yields

$$(3.1.9) \qquad \nabla C(P) = \nabla C(P_0) + \nabla\nabla C(P_0)(P - P_0).$$

When C reaches a minimum at P, $\nabla C(P) = 0$. Thus equation (3.1.9) becomes

$$(3.1.10) \qquad \nabla\nabla C(P_0)(P - P_0) = -\nabla C(P_0).$$

Solving for P iteratively from equation (3.1.10) renders the solution of this minimization process. This scheme is called the Newton-Raphson optimization method, and the double gradient $\nabla\nabla C(P)$ is called the Hessian matrix

$$(3.1.11) \qquad \nabla\nabla C(P) = \left(\frac{\partial^2 C}{\partial P_i \partial P_j} \right)_{i=1,2,3; j=1,2,3}.$$

Because C may have multiple minima, results might converge to a local minimum instead of the absolute minimum. Therefore, the initial guess needs to be "sufficiently close" to the solution. Typically, a coarse correlation is deployed either by hand or by a simple program before correlation. Results from the coarse correlation are used as initial values for the first point to be correlated. Then, correlation results of this point are used as initial values for the next point, and so on.

3.1.2 Continuous Signals

In order to evaluate the effectiveness of the minimization algorithm alone, the correlation technique is applied first to a simple case involving well-defined continuous line profiles and linear deformation. Various parameters associated with the image correlation can thus be examined and their influence on the correlation results evaluated quantitatively. In addition, more complex continuous signals are deployed and random noise is introduced to examine the influence of experimental noise on the correlation results.

3.1.2.1 Subset Size

According to work by Vendroux and Knauss[19] and Sutton et al.[15][16], subset size is always a critical parameter in the correlation. On one hand, one assumes that the subset is small enough so that the deformation within the subset can be represented as being linear, which implies that smaller subsets should be preferable. On the other hand, the subset has to include some feature of the signal, so that it can be readily correlated with the deformed subset, which argues for a larger subset. Therefore, there is reason for a trade-off with an optimum subset for specific signals and deformations.

In the following example, a uniform in-plane strain is applied to a single sinusoidal trace. The "topographical" original signal (as shown in Figure 3.2) and deformed signal are, respectively,

$$f(x) = \sin(2\pi \frac{x}{r})$$

$$g(x) = \sin(\,2\pi \frac{x}{r(1+\varepsilon)}\,)$$
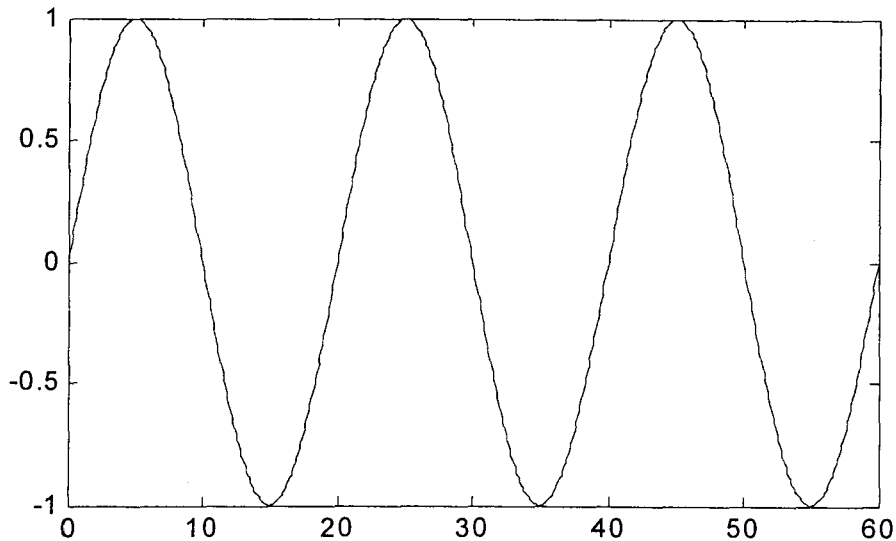
Figure 3.2. Single sinusoidal trace of wavelength 20

where r is chosen to be 20 to simulate the dominant wavelength of real silicon surfaces, and ε is 0.01 to represent a 1% in-plane strain.

Correlation is conducted on x=0, 1, ..., 20 for one period of the original signal. Computations are executed in Mathematica using the analytical form of the Newton-Raphson method. Results are shown in Figure 3.3. Iterations stop when the result differs from the previous iteration by less than the specified tolerance, which is chosen as 10^{-5} for this case. Thus, results with errors within this bound are considered as being accurate - errors can be further reduced by lowering the tolerance at the expense of computation time.
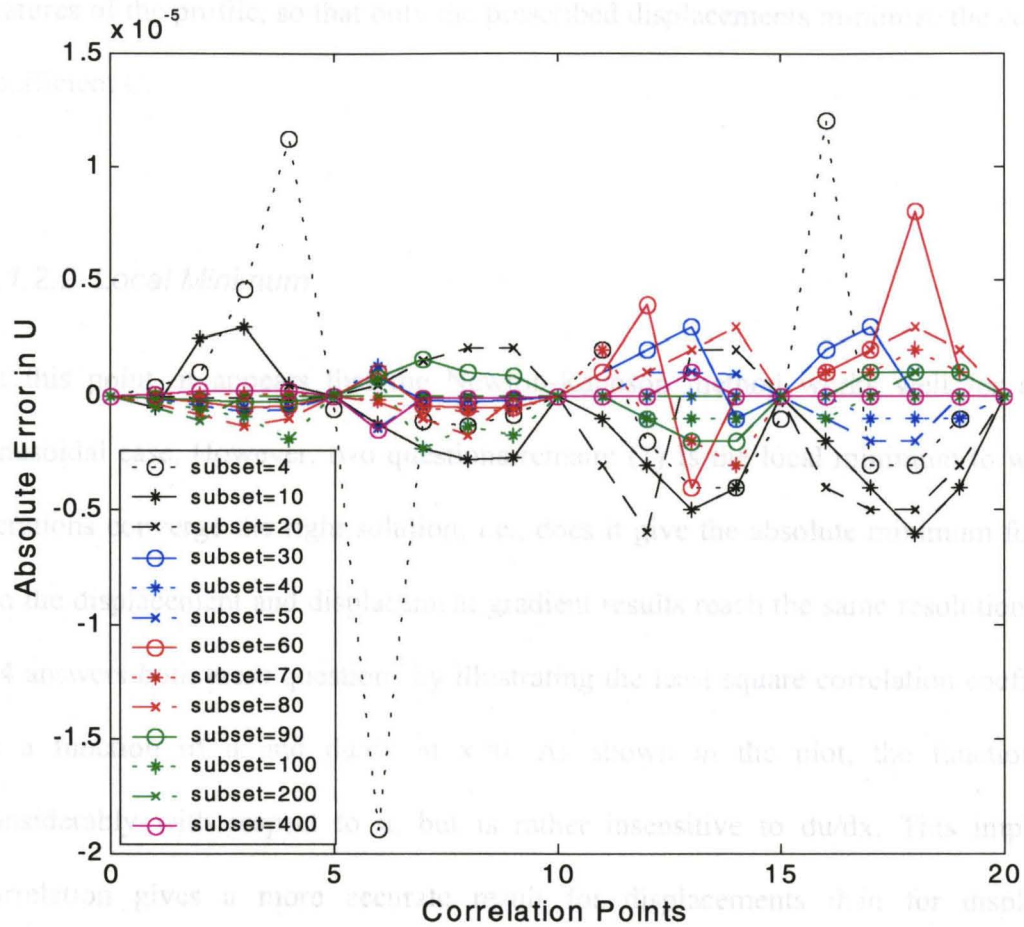
Figure 3.3. Correlation error using different subset sizes
$\varepsilon_x=0.01$, w=0, tolerance=10^{-5}

Figure 3.3 shows that any subset size greater than or equal to 10 (corresponding to half the wavelength of the signal) gives accurate displacement results. This can be explained intuitively. Half of the wavelength always includes a peak feature of the sinusoidal signal, which assures that the undeformed and deformed traces can be correlated. Also, in this case of linear deformation, larger subsets give better results, because the linearity assumption within the subset always holds, no matter how big the subset is. Figure 3.3 also shows that correlation on peak points and midpoints between peaks renders error-free results regardless of subset size. This is because these points represent distinct

features of the profile, so that only the prescribed displacements minimize the correlation coefficient C.

3.1.2.2 Local Minimum

At this point, it appears that the Newton-Raphson method works well for a simple sinusoidal case. However, two questions remain: (1) Is the local minimum to which the iterations converge the right solution, i.e., does it give the absolute minimum for C? (2) Do the displacement and displacement gradient results reach the same resolution? Figure 3.4 answers both these questions by illustrating the least square correlation coefficient C as a function of u and du/dx at x=0. As shown in the plot, the function varies considerably with respect to u, but is rather insensitive to du/dx. This implies that correlation gives a more accurate result for displacements than for displacement gradients, which agrees with previous observations[6] that strains calculated from u are better than du/dx extracted directly from the correlation.

Figure 3.4 also indicates that there is more than one local minimum in the correlation coefficient C. In order to find the absolute minimum, it is necessary to start the iteration at an initial value "sufficiently close" to the final results. That is the reason why results from current correlation point are used instead of constant values as initial guess for the next point.
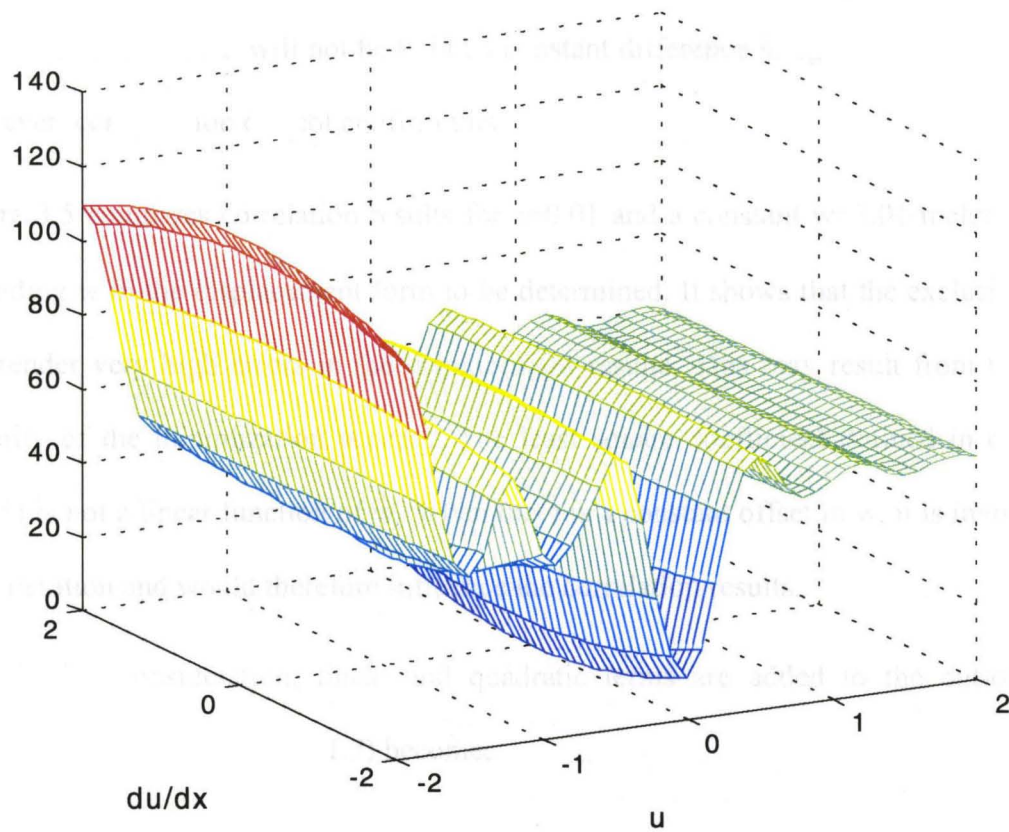
Figure 3.4. Least square correlation coefficient at x=0

3.1.2.3 Out-of-Plane Deformation

Scans obtained from the scanning tunneling microscope and the atomic force microscope before and after deformation are not guaranteed to remain in the same plane, so that locally (linear) out-of-plane deformation can exist. That prompts the question of whether out-of-plane deformation terms need to be included in the optimization process. Previous studies by Vendroux and Knauss[19] found that the addition of w in equation (3.1.5) helped with convergence and precision. Intuitively, if there is a constant out-of-plane

translation, correlation without any out-of-plane term should give the same results, except that the minimum of C will not be 0, but a constant difference integrated over the subset. However, computation did not confirm this.

Figure 3.5 compares correlation results for $\varepsilon=0.01$ and a constant $w=0.01$ including and excluding w in the displacement form to be determined. It shows that the exclusion of w can render very high errors in the displacement results. This may result from the non-linearity of the minimization process. The least square coefficient defined in equation (3.1.5) is not a linear function of w. When there is a constant offset in w, it is involved in each iteration and would therefore influence the correlation results.

In a further consideration, linear and quadratic terms are added to the out-of-plane displacement and equation (3.1.7) becomes

$$(3.1.12) \quad C(x_0,P) = \frac{\int \{f(x)+P_3(x_0)+P_4(x_0)*(x-x_0)+0.5P_5(x_0)*(x-x_0)^2 - g[x+P_1(x_0)+P_2(x_0)*(x-x_0)]\}^2 dx}{\int f(x)^2 dx}$$

where

$$(3.1.13) \qquad\qquad P = \{u, \frac{du}{dx}, w, \frac{dw}{dx}, \frac{d^2w}{d^2x}\}.$$

Correlation results are shown in Figure 3.6 and Figure 3.7. In these two cases, the errors for dw/dx and d^2w/dx^2 are found to be consistently 1% and 2% respectively.
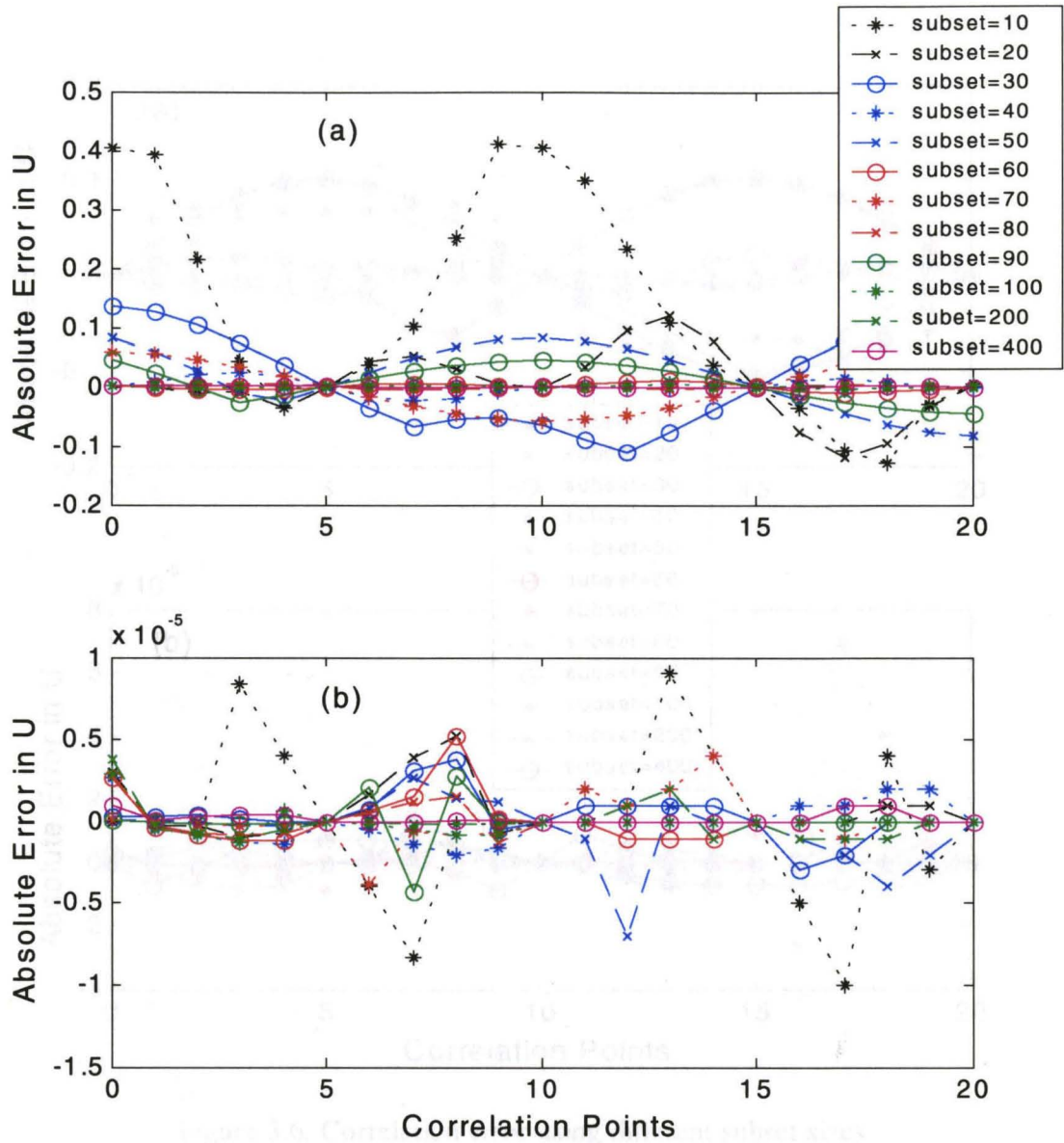
Figure 3.5. Correlation error using different subset sizes
$\varepsilon_x=0.01$, w=0.1, tolerance=10^{-5}
(a) Using {u, du/dx} as parameters for minimization
(b) Using {u, du/dx, w} as parameters for minimization
Note: different scales on the vertical axis for the two plots
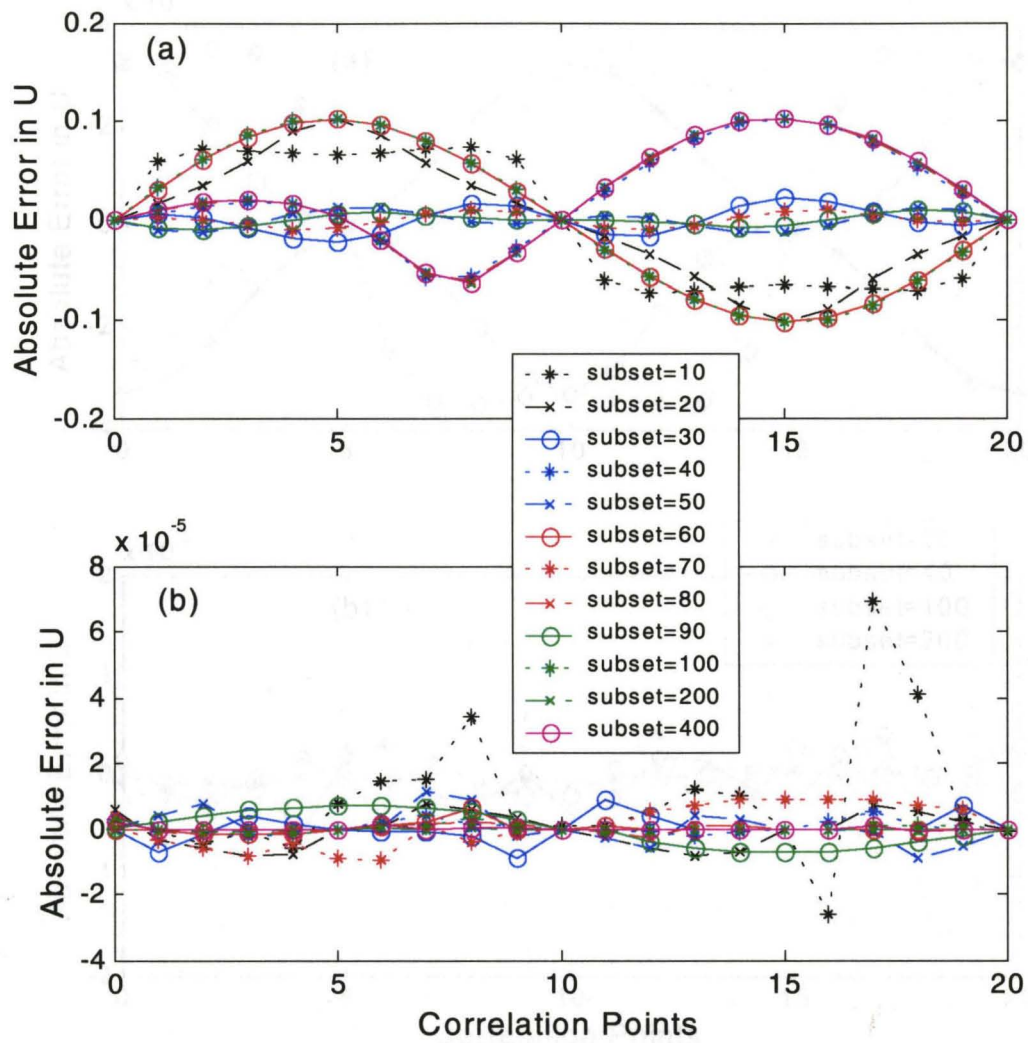
Figure 3.6. Correlation error using different subset sizes
$\varepsilon_x=0.01$, w=0.1+0.005x, tolerance=10^{-5}
(a) Using {u, du/dx, w} as parameters for minimization
(b) Using {u, du/dx, w, dw/dx} as parameters for minimization
Note: different scales on the vertical axis for the two plots
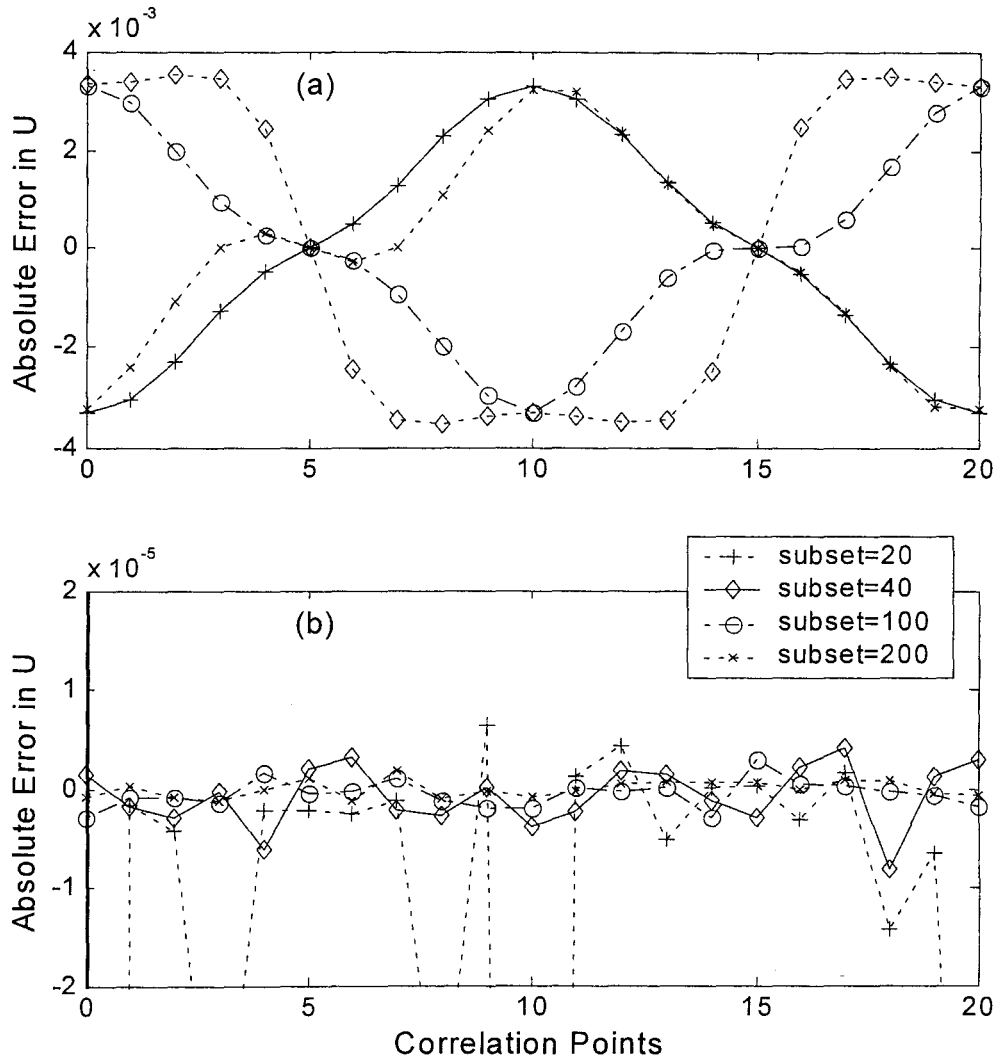
Figure 3.7. Correlation error using different subset sizes[*]
ε_x=0.01, w=0.1+0.005x+0.000025x^2, tolerance=10^{-5}
(a) Using {u, du/dx, w, dw/dx } as parameters for minimization
(b) Using {u, du/dx, w, dw/dx, d^{2}w/dx^2} as parameters for minimization
Note: different scales on the vertical axis for the two plots

---

[*]Fewer results are displayed because the computation takes much longer for the prescribed deformation.

From these plots, we deduce that it is important to include the out-of-plane deformation terms to the right order in the correlation. One also observes that as higher order terms are introduced into the displacement representation, the smallest subset that gives accurate results needs to be enlarged. For example, when w=0.01 and {u, du/dx, w} is used, the subset has to be greater than 4 to render acceptable results. This number increases to 20 for the case when prescribed out-of-plane displacement w=0.01+0.005x+0.000025x^2 and when {u, du/dx, w, dw/dx, d^{2}w/dx^2} is used in displacement representation.

3.1.2.4 Multiple Frequencies

Now that an understanding has been achieved of correlating traces of a single frequency, attention is turned to signals with multiple frequencies, as real signals can always be decomposed into multiple sinusoidal traces. Computations from here on are executed in Matlab, using built-in numerical integral functions in Matlab. First, sinusoidal signals with two frequencies are considered:

$$f(x) = \sin(2\pi \frac{x}{r}) + \sin(4\pi \frac{x}{r})$$

$$g(x) = f(\frac{x}{1 + \varepsilon}) + w(x)$$

where r=20, ε=0.01, and w(x)=0.01+0.0005x.

The correlation results are shown in Figure 3.8. Different dominance by each frequency component (while keeping the overall amplitude constant) has also been investigated.

Depending on which frequency component is dominant, the subset size has to equal at least half of the longer wavelength.
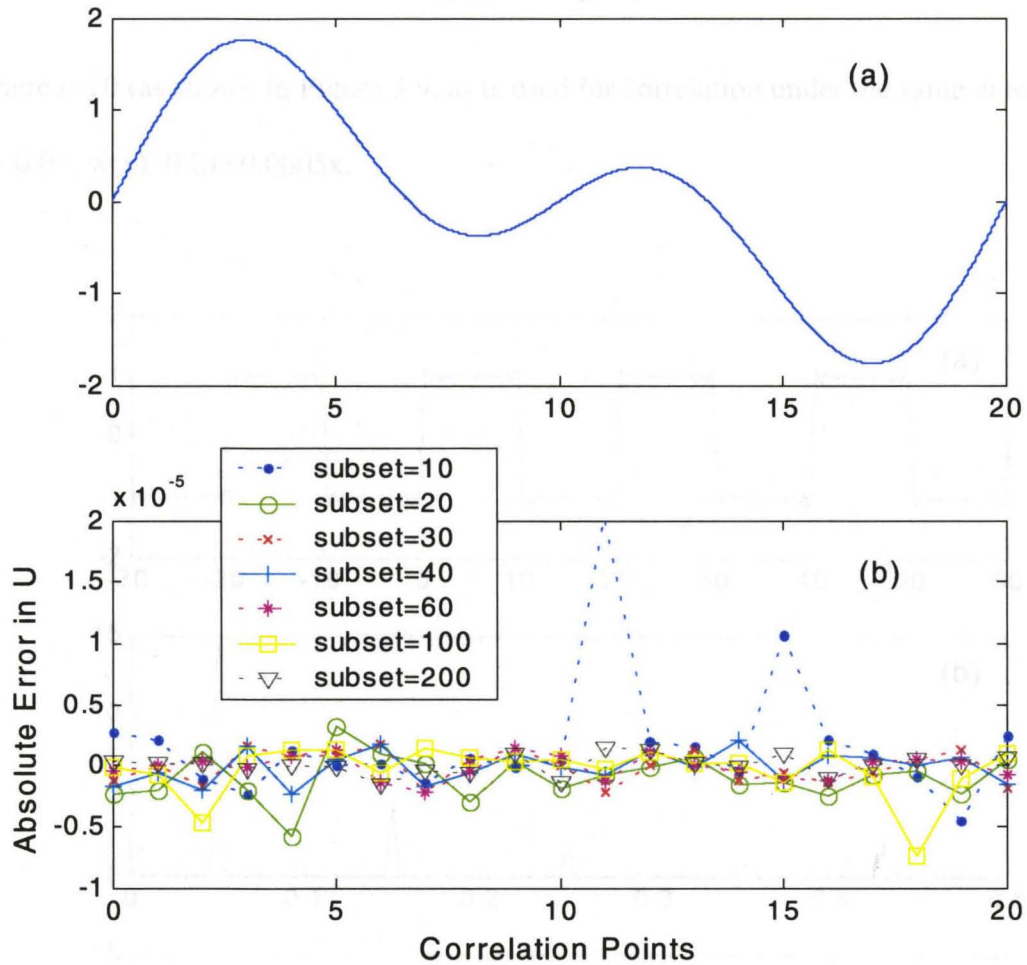
Figure 3.8. Correlation for a signal with two frequencies
(a) Undeformed signal
(b) Correlation error for w=0.01+0.0005x, ε_x=0.01, tolerance=10^{-5}, using {u, du/dx, w, dw/dx} as parameters for minimization

To verify this argument, a square wave with 11 terms

$$z = \frac{4}{\pi} \sum_{i=0}^{10} \frac{\sin((2i-1)\frac{x}{r}2\pi)}{2i-1}$$

where r=20 (as shown in Figure 3.9, a) is used for correlation under the same deformation

ε_x=0.01, w(x)=0.01+0.0005x.
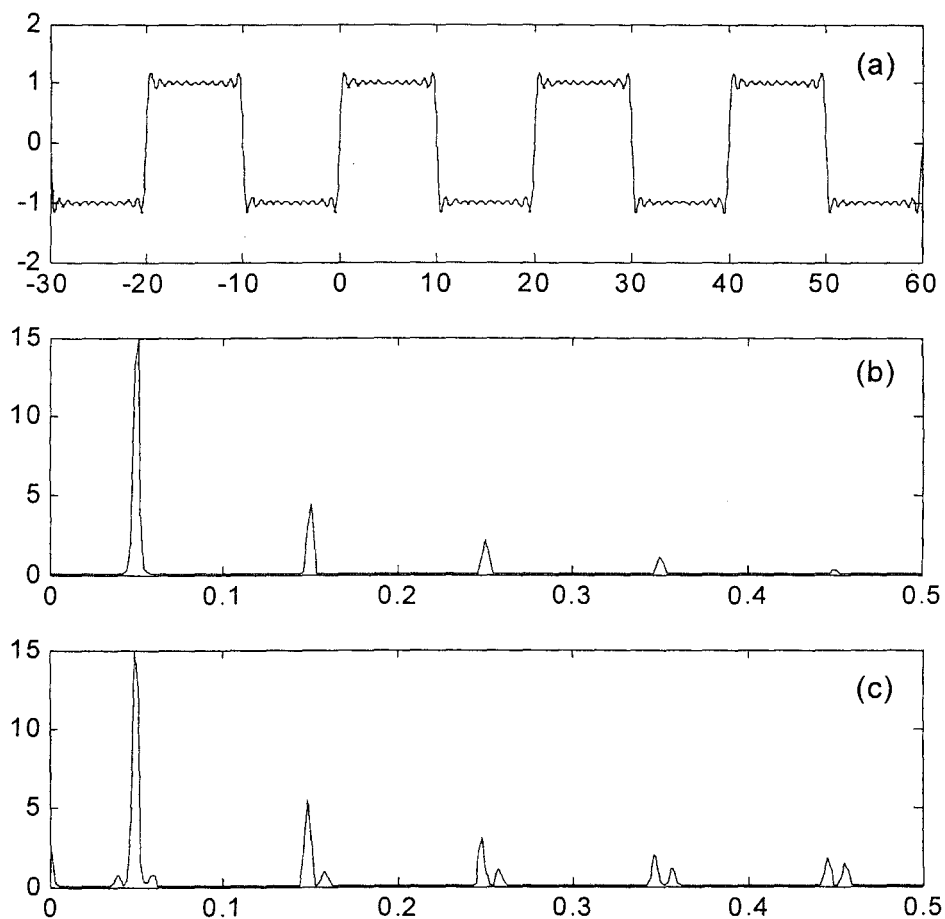
Figure 3.9. Square wave and power spectrum
(a) Undeformed signal
(b) Power spectrum of undeformed signal
(c) Power spectrum of deformed signal

The square wave is similar to the surface of a real specimen typically used for scanning tunneling and atomic force microscope calibration. Its power spectrum indicates the dominance of a few spatial frequencies. The correlation results with the subset size as a parameter are shown in Figure 3.10. As for two frequencies, one finds again that except for a few non-convergent points, the subset has to equal at least half of the longest wavelength, which is 10 in this case.
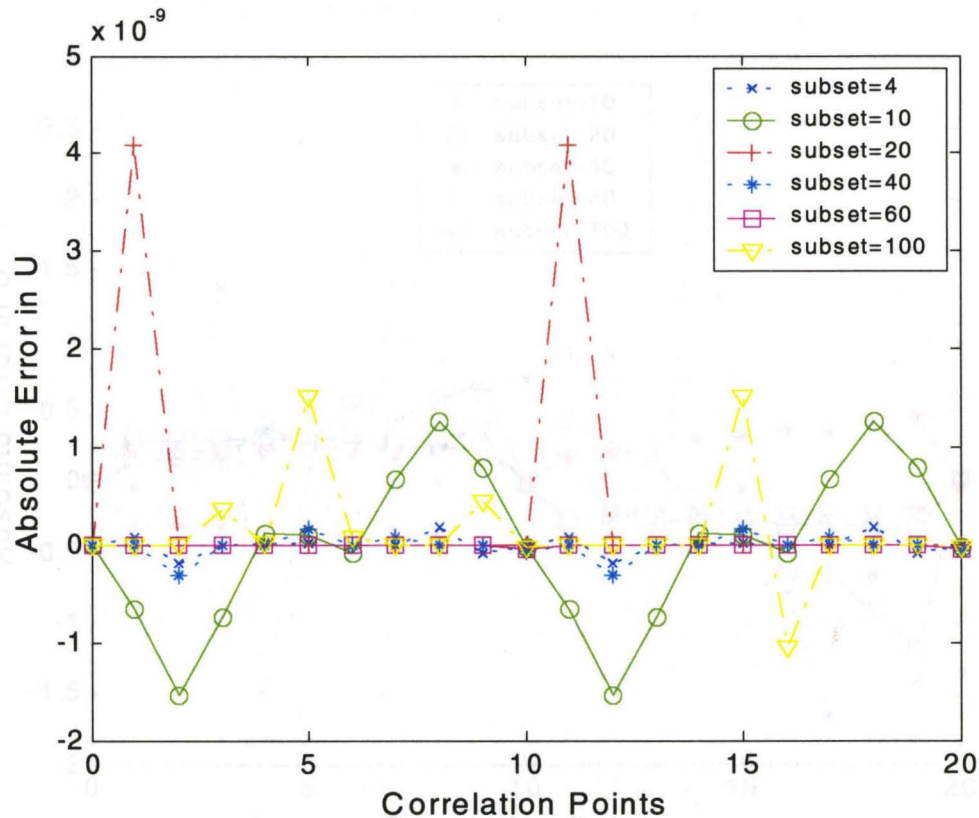
Figure 3.10. Correlation error for a square wave
$\varepsilon_x=0.01$, w=0.01+0.0005x, tolerance=10^{-5}

3.1.2.5 Uncertainty Associated with High Frequency

We consider next the question whether adding random unwanted signal to the deformed square wave will leave the error at the same general level. To answer this question, a high

frequency sinusoidal signal is superposed on the deformed square wave signal with the wavelength being one tenth of the fundamental one (20) and 1% of its amplitude. Figure 3.11 shows that the error increases to a 10^{-3} level. This indicates that high frequency uncertainty with even small amplitudes influences the results rather severely. It appears therefore that filtering the unwanted signal from any signal should improve the deformation extraction of the process.
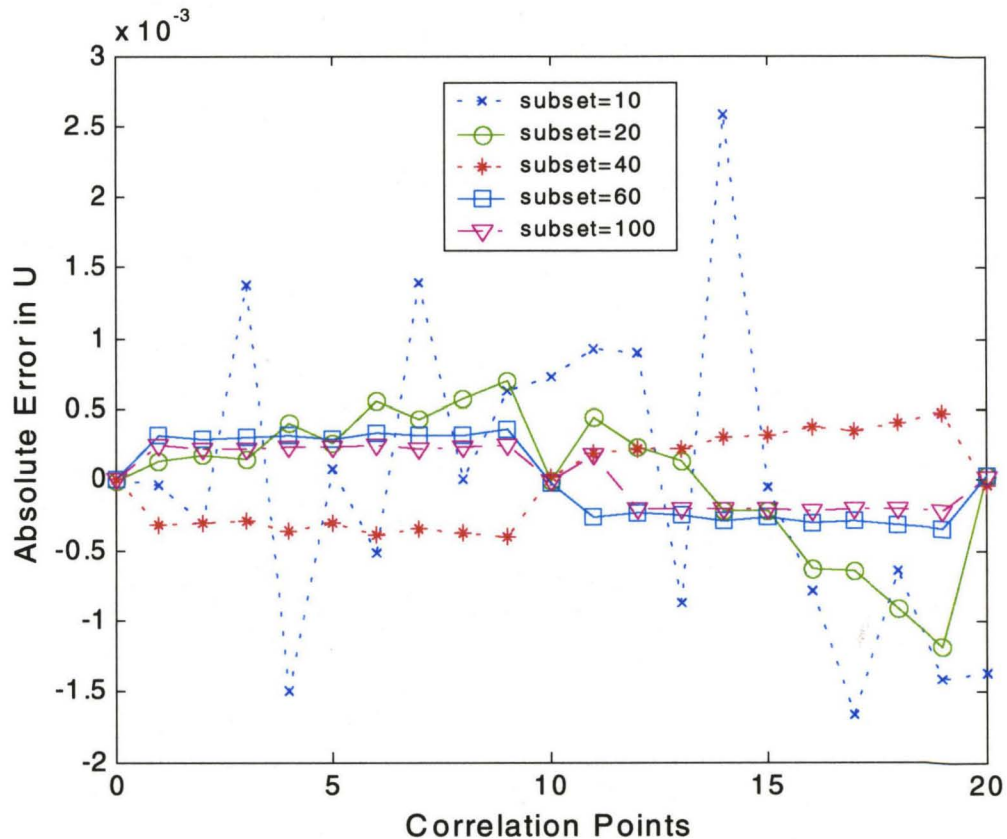
Figure 3.11. Correlation error for a square wave with uncertainty
$\varepsilon_x=0.01$, w=0.01+0.005x, tolerance=10^{-5}

3.1.3 Real Scan Signals

Real images recorded by means of a scanning tunneling microscope or an atomic force microscope are composed of discrete data. To achieve sub-pixel resolution for image

correlation, these images need to be interpolated so that height information of points between adjacent pixel points can be obtained. Errors associated with this interpolation may be introduced. Therefore, different interpolation schemes are briefly compared. In addition to the studies on continuous signals, correlation is applied to sinusoidal signals subject to non-linear in-plane deformation. Results show that more in-plane displacement gradients need to be included. Moreover, digital image correlation can be simplified by replacing the exact Hessian matrix with an approximate Hessian matrix[18]. This simplification improves computation time without sacrificing precision in the results.

For this part of the study, a scanning image of a silicon surface shown in Figure 3.12 is used. For the one-dimensional study, a line profile at y=250 in Figure 3.13 is used.
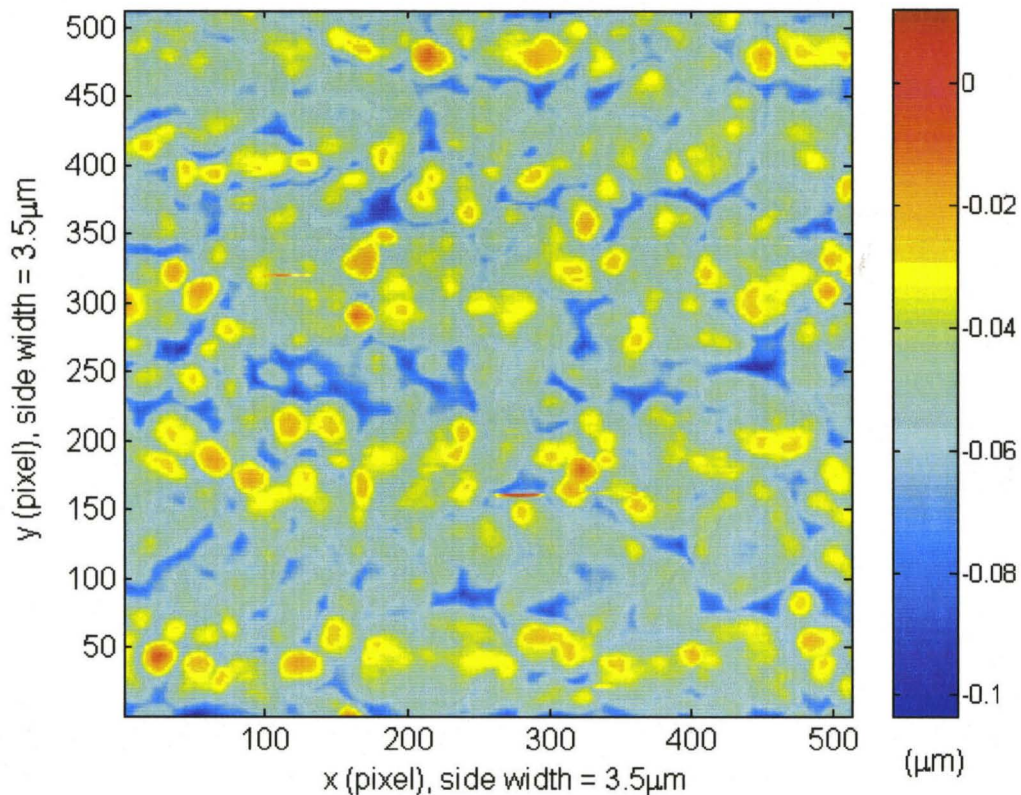
Figure 3.12. Silicon surface scan
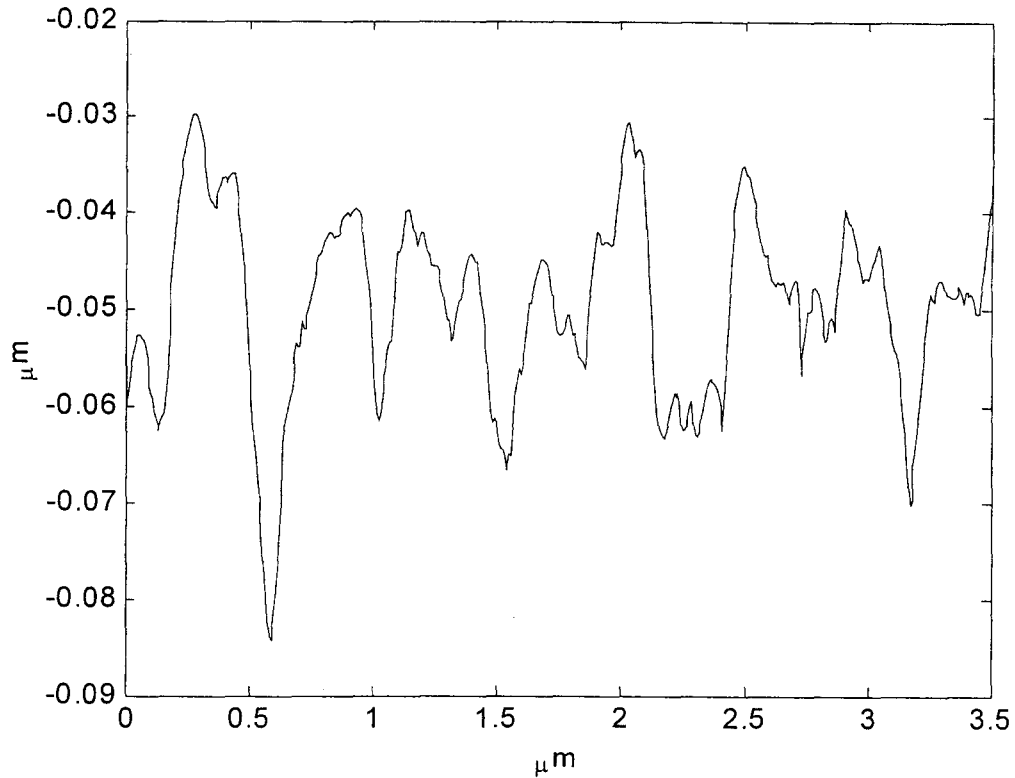
Figure 3.13. Line profile taken from silicon surface

3.1.3.1 Interpolation

To apply digital image correlation to discrete data, interpolation is required since discrete points do not move in units of pixel spacing. In this way, sub-pixel displacements may be accessed. Previous studies by Peters[13], Sutton[14] and Bruck[16] showed that cubic spline interpolation gives better results than linear and quadratic interpolations. This finding is substantiated by this work and cubic spline interpolation has been adopted.

To examine the effects of the parameters studied by means of smooth and analytical signals on interpolated data, correlation on discrete data subject to uniform in-plane translation and strains are repeated. The results agree with those for continuous signals.

This suggests that errors introduced by cubic spline interpolations are negligible compared to the errors originating in the algorithm.

3.1.3.2 Inhomogeneous Deformation

Digital image correlation is generally expected to be applied to inhomogeneous deformations such as, in the extreme case, those near a crack tip. It is, therefore, important to understand how well inhomogeneous deformations can be extracted by means of image correlation. For this purpose, the following displacement field with comparable linear and non-linear components is applied to the line profile in Figure 3.13.

$$u = 0.01x + 0.001x^2$$
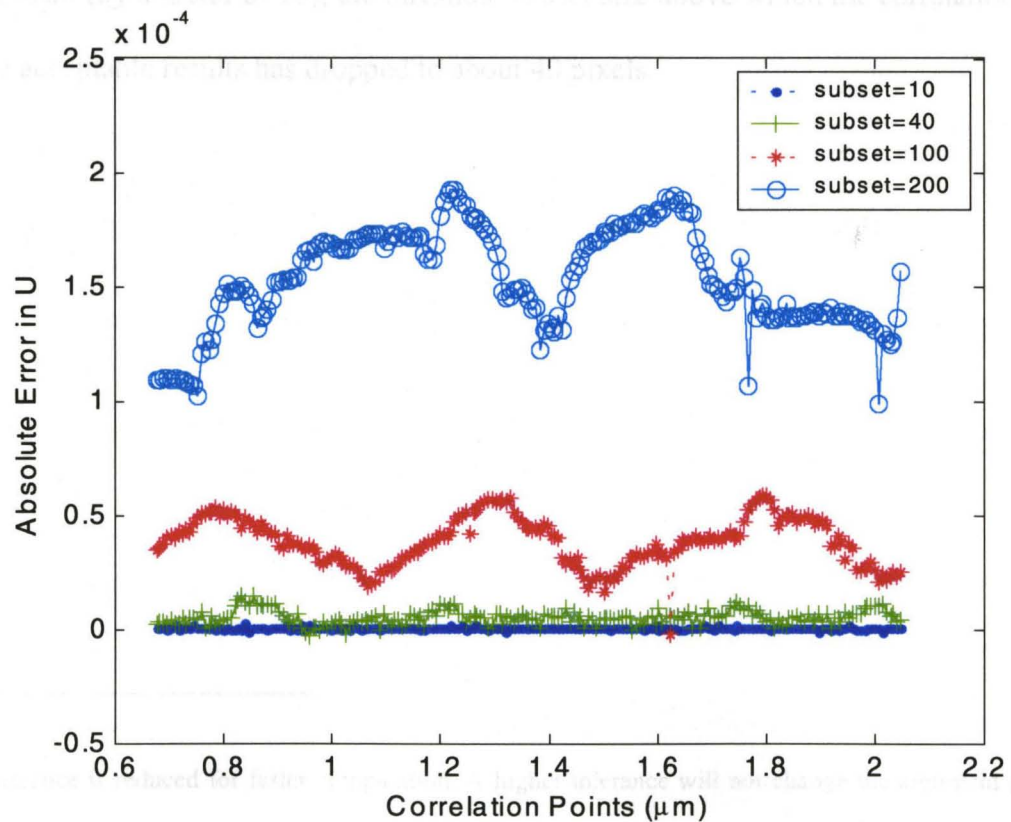
Figure 3.14. Correlation error for u=0.01x+0.001x^2, w=0
{u, du/dx} as parameters for minimization, tolerance=10^{-4}[*]

The quality of the correlation process using $\{u, \dfrac{du}{dx}\}$ to represent displacements is then

measured by errors in the displacement u as illustrated in Figure 3.14. One notes that the

error increases for larger subset sizes. The reason is that for non-linear deformations the

assumption of linear deformation within subsets only holds for small subset sizes. Once

the subset exceeds a threshold value (between 100 and 200 pixels in this case), the error

in the results is above the specified tolerance. This observation is verified by increasing

the amplitude of the non-linear term. For the displacement field $u = 0.01x + 0.01x^2$, the

correlation results are plotted in Figure 3.15. Because the non-linear displacement is more

dominant (by a factor of 10), the threshold subset size above which the correlation fails to

give acceptable results has dropped to about 40 pixels.

---

[*] Tolerance is reduced for faster computation. A higher tolerance will not change the argument presented

here.
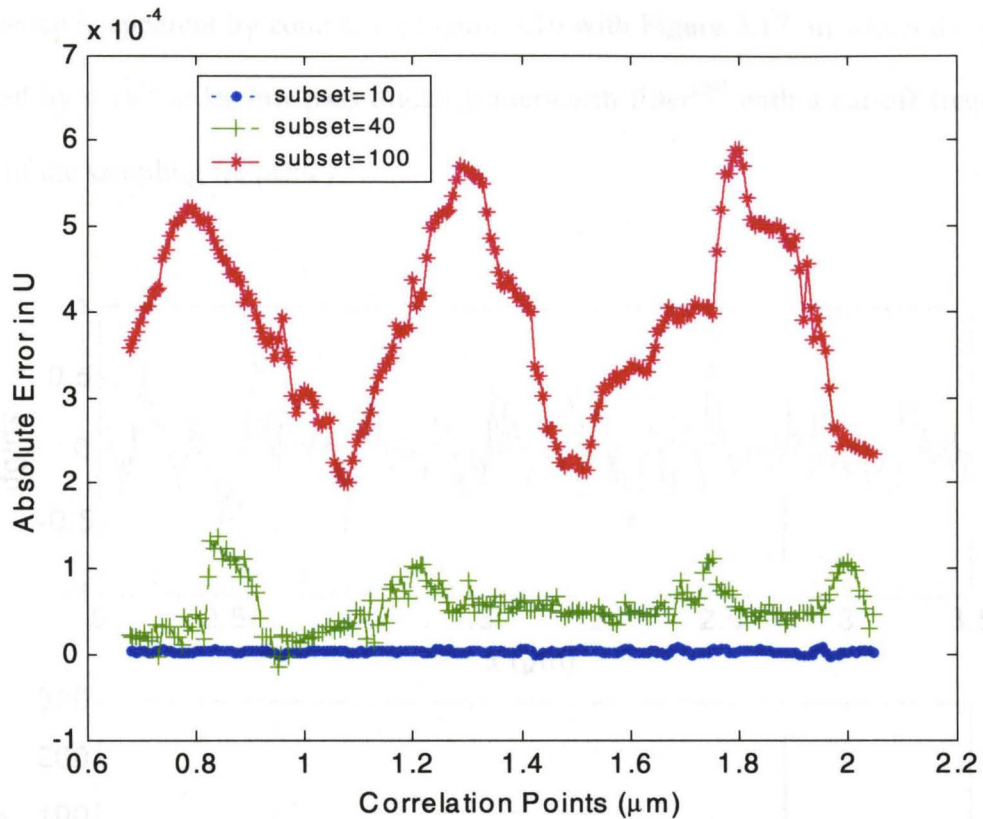
Figure 3.15. Correlation error for u=0.01x+0.01x², w=0
{u, du/dx} as parameters for minimization, tolerance=10⁻⁴

Next we examine whether the addition of a second order displacement gradient improves

the correlation results for large subset sizes. Then $\{u, \dfrac{du}{dx}, \dfrac{d^2u}{dx^2}\}$ are used as parameters

for the displacement field $u = 0.01x + 0.001x^2$, w=0. We find that for subset sizes of 40

and 100, the iterations fail to converge at many points. This problem is traced back to the

spiky nature of the second order gradient of the height profile used in the Hessian matrix

as shown in Figure 3.16. In numerical computations, a small error in the Hessian matrix

can invoke a big difference in the solution of the linear equations. The problem can be

solved by filtering out the high frequency component from the original profile. The

difference is apparent by comparing Figure 3.16 with Figure 3.17, in which the profile is filtered by a 10th order low pass digital Butterworth filter[20] with a cut-off frequency of 15% of the sampling frequency.
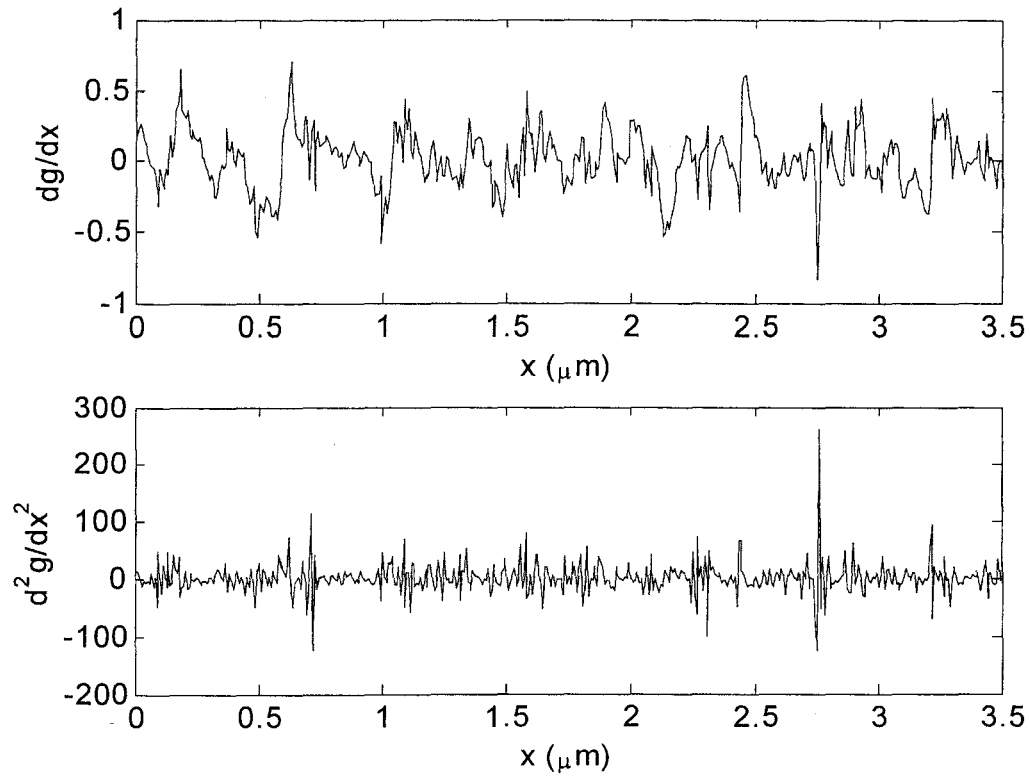
Figure 3.16. Differential terms in the Hessian matrix
Note: scale difference on the vertical axis for the two plots
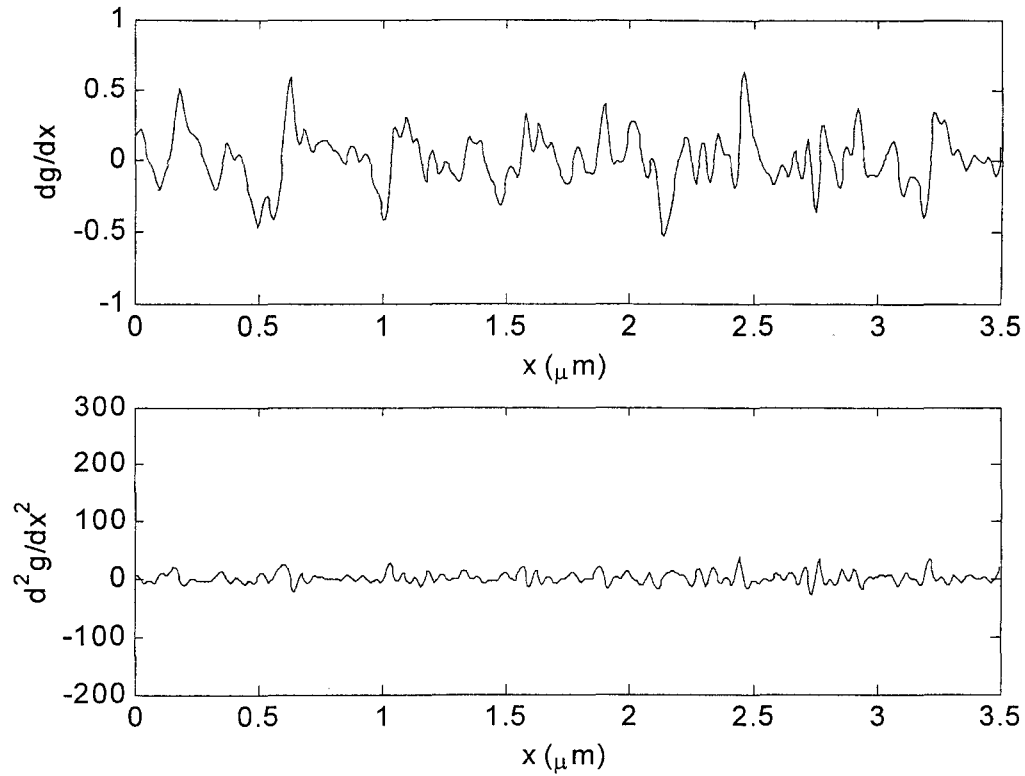
Figure 3.17. Differential terms in the Hessian matrix for a filtered signal
Note: scale difference on the vertical axis for the two plots

Figure 3.18 shows the correlation results of non-linear deformation applied to the filtered profile. The correlation error is well within the tolerance 10^{-4} regardless of the subset size. This result is consistent with the study on the out-of-plane deformations in the sense that correlation provides virtually exact results when sufficient displacement representation is included in the correlation.
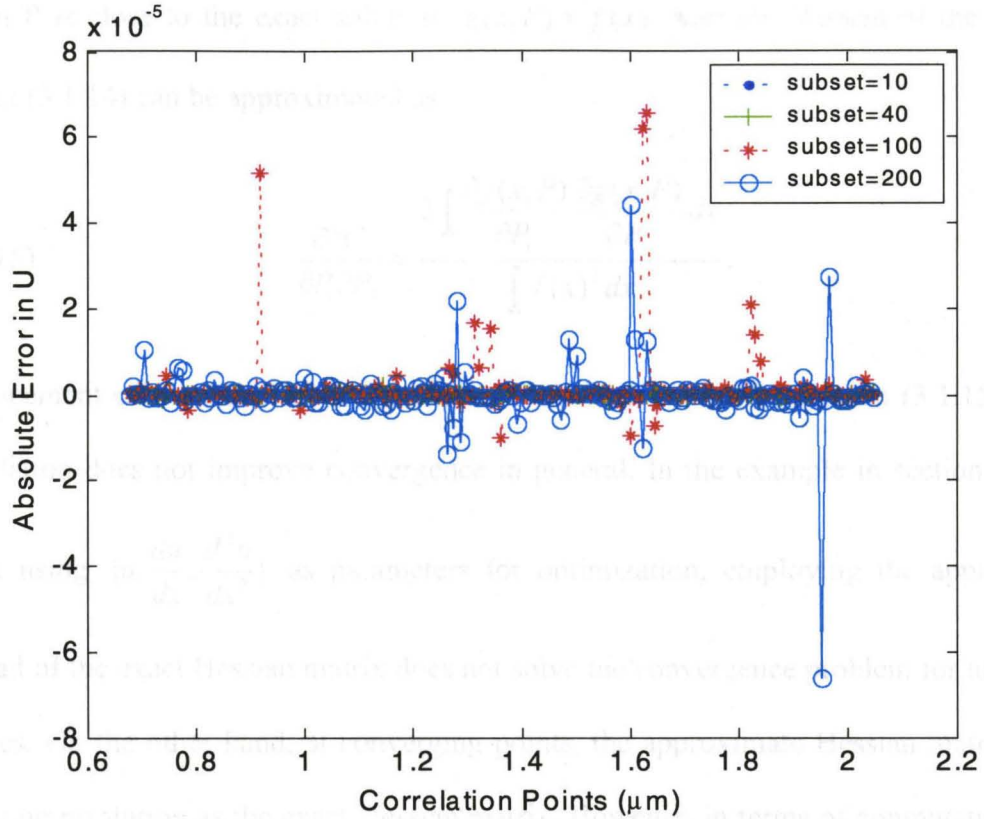
Figure 3.18. Correlation error for u=0.01x+0.001x², w=0
{u, du/dx, d²u/dx²} as parameters for minimization, tolerance=10⁻⁴

3.1.3.3 Exact Hessian Matrix vs. Approximate Hessian Matrix

Vendroux and Knauss[19] showed that approximating the exact Hessian matrix by omitting the second order gradient term does not affect the correlation results. To verify this argument, we approximate the Hessian matrix (3.1.11) as follows.

Plugging C (3.1.7) into an element (i=1,2, j=1,2) of the Hessian matrix renders

$$(3.1.14) \quad \frac{\partial^2 C}{\partial P_i \partial P_j} = \frac{\int_s 2\{[f(x) + P_3 - g(x,P)]\frac{\partial^2 g(x,P)}{\partial P_i \partial P_j} + \frac{\partial g(x,P)}{\partial P_i}\frac{\partial g(x,P)}{\partial P_j}\}dx}{\int_s f(x)^2 dx}.$$

When P is close to the exact solution, $g(x,P) \approx f(x)$, and the element of the Hessian matrix (3.1.14) can be approximated as

(3.1.15)
$$\frac{\partial^2 C}{\partial P_i \partial P_j} \approx \frac{2 \int_s \frac{\partial g(x,P)}{\partial P_i} \frac{\partial g(x,P)}{\partial P_j} dx}{\int_s f(x)^2 dx}.$$

Replacement of the exact Hessian matrix (3.1.14) by the approximation (3.1.15) in the correlation does not improve convergence in general. In the example in section 3.1.3.2, when using $\{u, \frac{du}{dx}, \frac{d^2 u}{dx^2}\}$ as parameters for optimization, employing the approximate instead of the exact Hessian matrix does not solve the convergence problem for unfiltered images. On the other hand, at converging points, the approximate Hessian matrix gives the same resolution as the exact Hessian matrix. However, in terms of computation time, the approximate Hessian matrix is much faster due to the lack of "spiky" second order gradient terms (as shown in Figure 3.16).

3.1.4 Conclusion

To simplify the study of parametric variation in the digital image correlation method, one-dimensional deformation fields have been considered, so far, to investigate the precision and limitation of this method. This work systematically identifies and evaluates the key parameters in this technique.

In general, the correlation method and Newton-Raphson algorithm work well, and provide more accurate measures of the displacements than those of their derivatives. For

homogeneous in-plane deformations, any subset size larger than half of the local wavelength of the signal, or the least size that includes some feature of the line profile, gives accurate correlation results. When out-of-plane deformations exist, a sufficient number of out-of-plane terms (sufficient to describe the deformation) need to be included in the correlation minimization: this is true even for a constant offset. Low amplitude, high frequency uncertainty between the undeformed and deformed signals can introduce significant errors into the correlation. Therefore, it is necessary to filter out such unwanted signals before image correlation is performed. In this work, a 10th order low-pass digital Butterworth filter with a cut-off frequency of 15% of the sampling rate has proven effective.

When discrete data are used instead of continuous signals, interpolation is needed for image correlation to reach sub-pixel resolution. It has been confirmed that cubic spline interpolation performs better than linear or quadratic interpolation.

- When inhomogeneous in-plane deformations exist, the assumption of linear deformation within a subset is still useful for sufficiently small subsets. However, if sufficient non-linear terms are included in the displacement representation, correlation renders acceptable results regardless of subset sizes.

- High frequency components in the signals may create convergence problems because of the "spiky" nature of the second order gradients of the height profile, especially when higher order displacement terms are included.

- In the Newton-Raphson algorithm, the Hessian matrix can be approximated by neglecting the second order terms. The approximate Hessian matrix yields the

same result as the exact one, but is significantly more economical in terms of computation time.

3.2 Two-Dimensional Digital Image Correlation

Although the study of one-dimensional digital image correlation provides considerable insight into the effect of key parameters involved, a comprehensive study of two-dimensional digital image correlation is needed to complete the study before applying this technique to two-dimensional topographical images obtained with probe microscopes.

The following two-dimensional analysis starts with the theoretical derivation. First, the simple case of a continuous, doubly sinusoidal surface subject to rigid body translations is explored, and the influence of the spatial sampling rate on the correlation process is examined. Then linear in-plane deformation is applied, and we study the influence of the size of the subset over which the minimization algorithm is conducted. In addition to in-plane deformation, out-of-plane deformation is superposed to confirm the conclusions drawn in the one-dimensional study. Next, we examine surfaces with multiple frequency content. A box surface with 11 frequencies is used for correlation and the results are compared with those for the single frequencies. Moreover, high frequency, low amplitude noise is again added to the well-defined surface to examine errors induced by noise. Furthermore, real scans with a complex frequency spectrum are also used. Inhomogeneous in-plane deformations are applied and correlation results are analyzed.

Finally a summary of the important parameters is given and a calibration method to work with noisy experimental data is proposed.

3.2.1 Theoretical Background

In parallel to the one-dimensional situation, let F be a point of coordinates (x, y) in an undeformed configuration and f(x, y) the height of the profile at F. After deformation, the point is mapped into $G = g(\tilde{x}, \tilde{y})$, where

$$(3.2.1) \qquad \begin{aligned} \tilde{x} &= x + u(x,y) \\ \tilde{y} &= y + v(x,y) \end{aligned}$$

$$(3.2.2) \qquad g(\tilde{x}, \tilde{y}) = f(x,y) + w(x,y)$$

with u, v the in-plane displacements and w the out-of-plane displacement of F.

Let F_0 of coordinates (x₀, y₀) be mapped to G_0 having coordinates $(\tilde{x}_0, \tilde{y}_0)$, and S be a subset around point F_0. Assuming that S is sufficiently small, the deformation of a point within S can be written as

$$(3.2.3) \qquad \begin{aligned} \tilde{x} &= x + u(x_0,y_0) + \left.\frac{\partial u}{\partial x}\right|_{(x_0,y_0)} (x - x_0) + \left.\frac{\partial u}{\partial y}\right|_{(x_0,y_0)} (y - y_0) \\ \tilde{y} &= y + v(x_0,y_0) + \left.\frac{\partial v}{\partial x}\right|_{(x_0,y_0)} (x - x_0) + \left.\frac{\partial v}{\partial y}\right|_{(x_0,y_0)} (y - y_0) \end{aligned}$$

Define a least square coefficient C on the subset S such as

$$(3.2.4) \qquad C = \frac{\int_S [f(x,y) - g(\tilde{x}, \tilde{y})]^2 \, dS}{\int_S f^2(x,y) \, dS} \, .$$

It is again obvious that C is zero when the mapping is exact. Based on the assumption of linear deformation within the subset and upon consideration of out-of-plane displacement*, the least square coefficient at point F_0 can be rewritten as

$$(3.2.5) \qquad C(F_0) = \frac{\int_S [f(F) + w(F_0) - g(x + \begin{pmatrix} u & \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ v & \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{pmatrix}_{F_0} \begin{pmatrix} 1 \\ x - x_0 \\ y - y_0 \end{pmatrix})]^2 \, dS}{\int_S f(F)^2 \, dS} .$$

Define a seven-dimensional vector

$$(3.2.6) \qquad P(F) = \{u(F), \frac{\partial u}{\partial x}(F), \frac{\partial u}{\partial y}(F), v(F), \frac{\partial v}{\partial x}(F), \frac{\partial v}{\partial y}(F), w(F)\} ,$$

so that C becomes

$$(3.2.7) \qquad C(F_0, P) = \frac{\int_S [f(F) + P_7(F_0) - g(x + \begin{pmatrix} P_1 & P_2 & P_3 \\ P_4 & P_5 & P_6 \end{pmatrix}_{F_0} \begin{pmatrix} 1 \\ x - x_0 \\ y - y_0 \end{pmatrix})]^2 \, dS}{\int_S f(F)^2 \, dS} .$$

Expanding C(P) again as a truncated Taylor series around P_0 leads to

$$(3.2.8) \qquad C(P) = C(P_0) + \nabla C(P_0)^T (P - P_0) + \tfrac{1}{2}(P - P_0)^T \nabla\nabla C(P_0)(P - P_0) .$$

Differentiating equation (3.2.8) with respect to P on both sides yields

$$(3.2.9) \qquad \nabla C(P) = \nabla C(P_0) + \nabla\nabla C(P_0)(P - P_0) .$$

When C reaches a minimum at P, $\nabla C(P) = 0$. Equation (3.2.9) becomes

---

* Initially only the zero order term is included. Addition of higher order terms will be discussed later.

$$(3.2.10) \qquad \nabla\nabla C(P_0)(P - P_0) = -\nabla C(P_0) .$$

Solving for P iteratively from equation (3.2.10) will converge to the solution of the minimization problem, provided that an initial guess is chosen that allows convergence. The double gradient $\nabla\nabla C(P)$ is again the Hessian matrix where

$$(3.2.11) \qquad \nabla\nabla C(P) = \left(\frac{\partial^2 C}{\partial P_i \partial P_j} \right)_{i=1,..,7; j=1,..,7} .$$

Because C may have multiple minima, results might converge to a local minimum instead of the absolute minimum. Therefore, the initial guess needs to be "sufficiently close" to the solution. As used in the one-dimensional study, a coarse correlation is typically deployed either by hand or by a simple program before correlation. Results from the coarse correlation are used as initial values for the first point to be correlated. Along each line of grid points, correlation results of the current point are used as initial values for the next point. When correlation moves to the next line of grid points, correlation results from the first point of the previous line are used as initial values for this point, and so on.

3.2.2   Continuous Surfaces With Well Defined Frequency Content

Unlike the one-dimensional case, the two-dimensional digital image correlation involves at least seven parameters, which makes correlating continuous surfaces by using the analytical form in any mathematical software more complex. Therefore, even the first-step study in two dimensions has to be conducted with interpolation, and errors associated with this interpolation may enter. However, careful choice of the interpolation scheme can minimize this error. As mentioned in the one-dimensional study, cubic spline

interpolation provides better results than linear or quadratic interpolation. Therefore, bicubic spline interpolation is used for the two-dimensional case. First a single sinusoidal surface in both directions (as shown in Figure 3.19) is deformed in various ways and the correlation algorithm is applied to extract the deformation. A digital image correlation program written in C++ is used to perform the two-dimensional correlation.

Figure 3.19. A sinusoidal surface with wavelength 14 in x and 19 in y (not to scale)

3.2.2.1 Sampling Frequency

As a first demonstration of the two-dimensional digital image correlation program, rigid body translations

$$u= 0.12563489012, \quad v= 0.964832030485^*$$

are applied to the sinusoidal surface shown in Figure 3.19 and correlation is conducted on the center square [20, 30] x [20, 30]. The very first factor that appears to have a substantial influence on the correlation results is the sampling frequency, or the number of sampling points within one length unit. Table 3.1 compares the correlation results for the sinusoidal surface sampled at different spacing (length between two adjacent sampling points). During the correlation, iterations stop when the result differs from the previous iteration by less than the specified tolerance, which is chosen as 10^{-6} in this case. Thus, results with errors within this bound are considered as being accurate; errors can be further reduced by lowering the tolerance at the expense of computation time.

Sampling Spacing	0.2	1	2
Error in u	Mean: 1.099e-7 STD: 0	Mean: 2.517e-5 STD: 7.78e-7	Mean: 4.691e-4 STD: 5.61e-5
Error in v	Mean: -3.05e-8 STD: 0	Mean: -2.69e-6 STD: 4.073e-6	Mean: 6.7e-5 STD: 1.692e-4

Table 3.1. Correlation error for different sampling frequencies

The larger error at smaller sampling frequency (larger sampling spacing) is readily explained in terms of the interpolation. Interpolation is introduced when information between adjacent pixels is needed. As demonstrated in Figure 3.20, the interpolated surface represents the real surface better with more sampling points, thereby reducing the error incurred by interpolation. The interpolation is performed in the x direction first,

---

* A large number of digits are included intentionally in the prescribed displacements to obtain the correlation error precisely.

then in the y direction. Thus the error from the interpolation along x is carried on to the interpolation along y, which suggests larger error in the y direction. This is verified in (b) and (c) of Figure 3.20. For large sampling spacing, the interpolated image appears smoother in the x direction than in the y direction. The difference between these two directions gradually disappears as the sampling spacing decreases. One deduces from Table 3.1 and Figure 3.20 that the correlation error is in direct proportion to the interpolation error corresponding to different sampling frequencies.
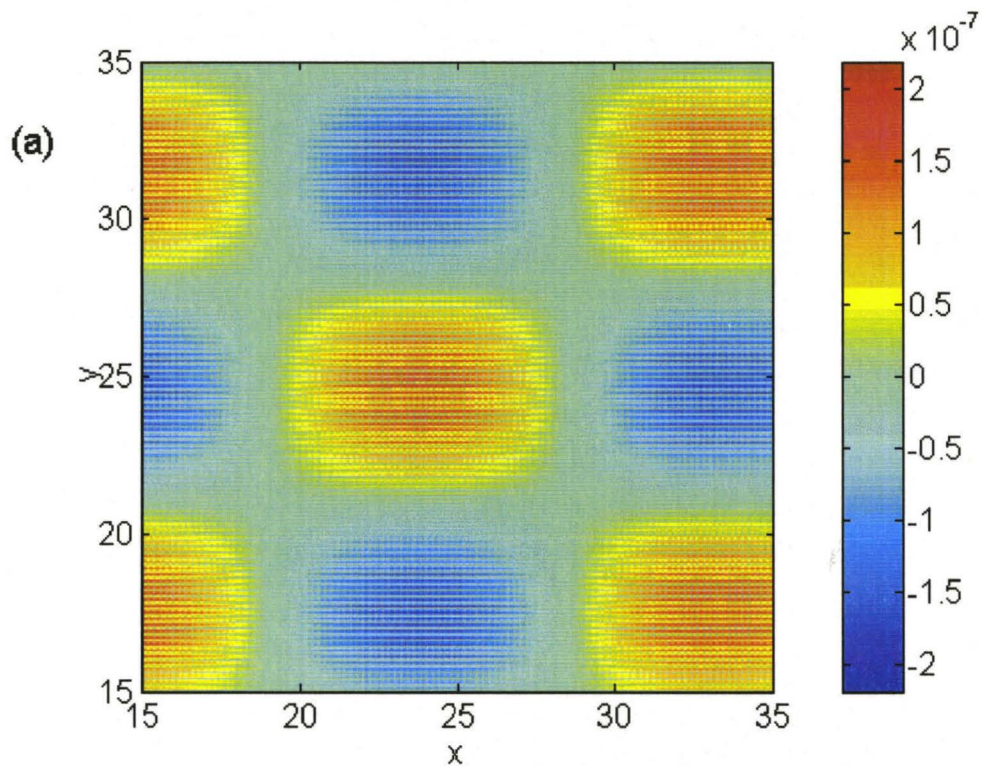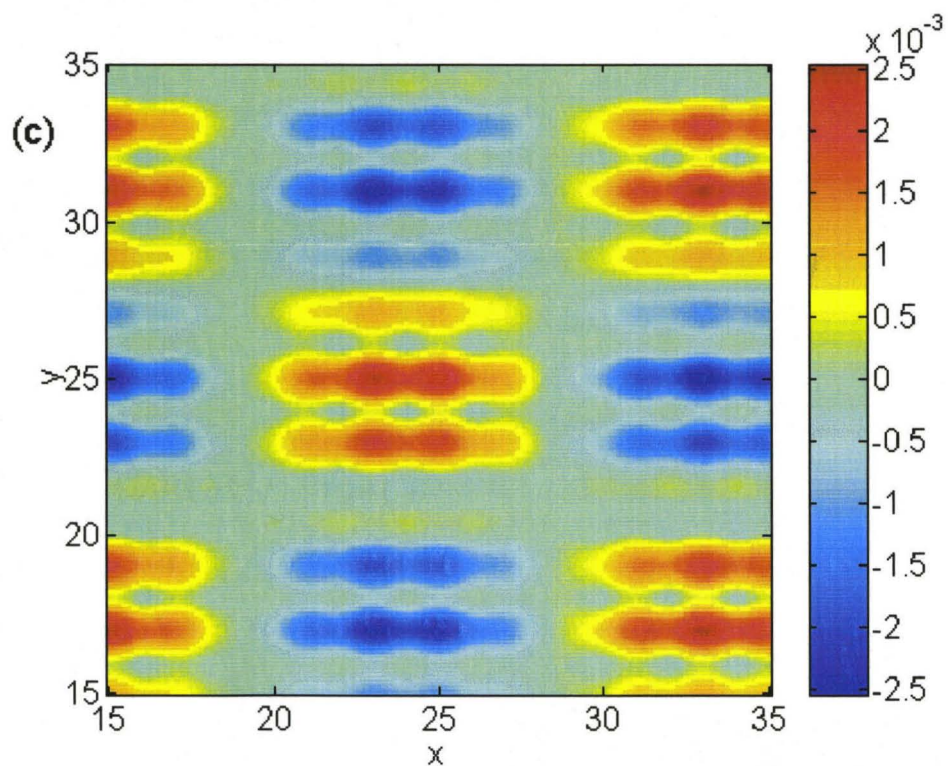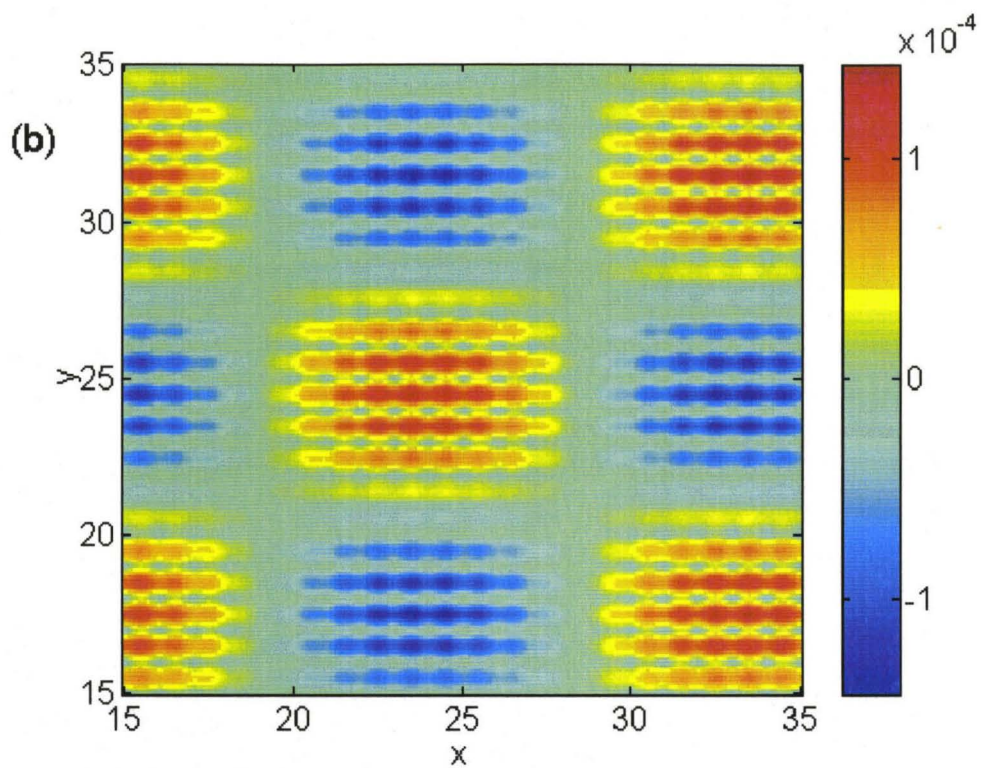
Figure 3.20. Interpolation error for different sampling frequencies
(a) sampling spacing 0.2, (b) sampling spacing 1, (c) sampling spacing 2

Correlations are also performed for rigid body translation in the out-of-plane direction as well as in-plane and out-of-plane rigid body rotations. Errors in the results are consistent in magnitude with the results in Table 3.1. When there are enough sampling points to represent the surface (with sampling spacing ~0.2 in the above case), correlation will give results within the specified error tolerance (10^{-6} in this case).

3.2.2.2 Subset Size

Due to instrumentation limits, there is an upper bound of the possible number of sampling points that can be achieved in scanning images (with either the scanning tunneling microscope or the atomic force microscope). Therefore, the digital image correlation may fail to give results within the desirable error bound specified. That prompts the question: given the limit of the sampling frequency, what can be done to improve the correlation results? The answer follows from the experience with the one-dimensional case. The first important factor to investigate is the subset size[*].

Subset size	2	4	10	20	38
Error in u: Mean	-3.5884e-6	-5.1975e-6	-4.5752e-6	2.5702e-7	3.2810e-7
STD	3.9963e-5	5.7880e-5	3.9625e-5	1.0818e-5	5.5446e-6
Error in v: Mean	1.2661e-5	1.9353e-5	6.2595e-6	9.4041e-6	1.6998e-5
STD	5.5906e-5	8.2725e-5	3.6267e-5	1.8103e-5	4.8169e-5

---

[*] In two-dimensional digital image correlation, the subset size is defined as the size of each side of the subset, similar to the one-dimensional version.

Error in $\partial u/\partial x$: Mean	6.3229e-5	6.2352e-5	4.0537e-7	-1.8363e-6	-1.8366e-6
STD	3.6303e-5	4.3321e-5	1.3785e-5	3.6738e-6	7.4409e-7
Error in $\partial v/\partial y$: Mean	7.3356e-5	8.5455e-5	3.0782e-5	6.4141e-7	1.7476e-6
STD	4.059e-5	4.5477e-5	4.8834e-5	6.6137e-6	7.5507e-6

Table 3.2. Correlation error for different subset sizes

Correlation is performed for the homogeneous deformation ε_{xx}=0.02, ε_{yy}=0.005 on the sinusoidal surface (shown in Figure 3.19) sampled at a spacing of 1. The results are shown in Table 3.2, with a specified tolerance 10^{-6}. The increase of the subset does provide marginally better correlation results. However, the difference is not as pronounced as for one-dimensional continuous signals. Part of the reason is that the addition of another dimension offers more geometrically distinct features, so that it is "easier" to locate the mapping even for small subsets with only nearest neighboring pixels. Another feature distinguishing two-dimensional correlation results from the one-dimensional version is that displacement results are not necessarily significantly better than the results for displacement gradients.

As in one-dimensional correlation, the Newton-Raphson algorithm works well in two dimensions. The only difference in the implementation of the correlation process is that at the first point on each line to be correlated, the correlation results from the first point of the previous line are used as the initial guess to start the iteration, instead of that for the last correlation point.

3.2.2.3 Out-Of-Plane Deformation

The next factor considered is the out-of-plane deformation (Poisson Effect) that is commonly expected in the use of probe microscopes for solid mechanics deformation studies. The present investigation on one-dimensional correlation as well as previous work by Vendroux and Knauss[19] found that the addition of w improved convergence of the minimization process. In the two-dimensional digital image correlation program, w is included by default. The gradients of w for up to 2nd order as well as the in-plane displacement gradients can be included when needed. When these terms are included, $P_7(F_0)$ in equation (3.2.7) is replaced with

$$P_7(F_0) + P_8(x - x_0) + P_9(y - y_0) + \frac{1}{2}P_{10}(x - x_0)^2 + P_{11}(x - x_0)(y - y_0) + \frac{1}{2}P_{12}(y - y_0)^2$$

and (3.2.6) becomes

$$(3.2.12)\ P(F) = \left\{u, \frac{\partial u}{\partial x}, \frac{\partial u}{\partial y}, v, \frac{\partial v}{\partial x}, \frac{\partial v}{\partial y}, w, \frac{\partial w}{\partial x}, \frac{\partial w}{\partial y}, \frac{\partial^2 w}{\partial x^2}, \frac{\partial^2 w}{\partial x \partial y}, \frac{\partial^2 w}{\partial y^2}\right\}\Bigg|_F, \text{ a 12-element vector.}$$

We first apply a constant out-of-plane displacement w=0.2 to the sinusoidal surface (shown in Figure 3.19) sampled at a spacing of 0.2 and conduct correlation on the center square [15, 35] x [15, 35] with a tolerance of 10^{-5} and a subset size of 10.

Parameters included in the z-direction	w	w, $\partial w/\partial x$, $\partial w/\partial y$
Error in u: Mean	-4.4293e-17	-4.0224e-16
STD	8.1958e-16	6.7952e-16
Error in v: Mean	-2.4833e-16	-4.5686e-16
STD	3.5369e-16	3.9750e-16
Error in w: Mean	-4.4409e-16	-4.4409e-16
STD	4.4594e-16	4.4594e-16

Table 3.3. Correlation error for rigid out-of-plane displacement w=0.2

Table 3.3 shows that correlation results are well within the specified error bound. In a further consideration, linear terms are added to the out-of-plane displacement

$$w=0.2+0.01x+0.001y$$

with results shown in Table 3.4. Note that correlation error with only w in the displacement representation is fairly large. Further tests with out-of-plane inclinations of up to 20-50% are conducted. Correlation results for these cases, in which at least linear terms are included in the displacement representation, are consistent with those shown in Table 3.4. Correlation errors with only w in the displacement representation increase with larger out-of-plane inclination.

Parameters included in the z-direction	w	w, $\partial w/\partial x$, $\partial w/\partial y$	w,$\partial w/\partial x$,$\partial w/\partial y$, $\partial^2 w/\partial x^2$,$\partial^2 w/\partial x\partial y$,$\partial^2 w/\partial y^2$
Error in u: Mean	-0.0021	4.5319e-17	-3.8849e-17
STD	0.0451	1.1036e-15	8.4492e-16
Error in v: Mean	-1.4732e-4	-1.5939e-16	-8.6764e-17
STD	0.0565	5.5127e-16	8.5374e-16
Error in w: Mean	-2.2678e-5	-1.4337e-17	-1.4337e-17
STD	0.0021	3.2844e-17	3.2844e-17

Table 3.4. Correlation error for out-of-plane deformation w=0.2+0.01x+0.001y
Note: significant improvement from first column to second and third column

Furthermore, quadratic terms are added in the specified displacement, so that the prescribed out-of-plane displacement is

$$w=0.2+0.01x+0.01y+0.0015x^2+0.0005xy+0.001y^2.$$

Table 3.5 compares correlation results with and without quadratic terms included in the displacement representation. The same conclusion is reached: the more dominant quadratic term contribution in the prescribed out-of-plane displacement, the higher correlation errors without quadratic terms in the representation.

Parameters included in z direction	$w, \partial w/\partial x, \partial w/\partial y$	$w, \partial w/\partial x, \partial w/\partial y,$ $\partial^2 w/\partial x^2, \partial^2 w/\partial x \partial y, \partial^2 w/\partial y^2$
Error in u: Mean	3.7297e-4	-2.4541e-16
STD	0.0205	2.0936e-15
Error in v: Mean	4.0023e-4	5.7388e-16
STD	0.0257	2.5507e-15
Error in w: Mean	0.0124	-4.3434e-16
STD	0.0064	4.6618e-16

Table 3.5. Correlation error for out-of-plane deformation
$w=0.2+0.01x+0.01y+0.0015x^2+0.0005xy+0.001y^2$

From these results, we deduce that it is important to include a sufficient number of out-of-plane deformation terms in the correlation expression to represent the prescribed displacement field, so that the algorithm converges to the absolute minimum.

3.2.2.4 Multiple Frequencies

After the study on correlating sinusoidal surfaces with a single frequency, we examine next surfaces with multiple frequency components. A grid surface composed of square waves with 11 terms

$$z = \frac{1}{4}\left(\sum_{i=0}^{10} \frac{\sin((2i-1)x\frac{2\pi}{w_x})}{2i-1} + 1\right)\left(\sum_{j=0}^{10} \frac{\sin((2j-1)y\frac{2\pi}{w_y})}{2j-1} + 1\right)$$

sampled at a spacing of 0.2 (as shown in Figure 3.21) is used for this purpose. The surface is first subject to the deformation

$$\varepsilon_{xx}=0.01, \; \varepsilon_{yy}=0.001.$$

The box surface (composed of box shaped depressions) represents a real specimen that is typically used for scanning tunneling and atomic force microscope calibrations. A trace taken out at x=25 is shown in Figure 3.22. Its power spectrum indicates the dominance of a few spatial frequencies. Correlation results with different subset sizes are shown in Table 3.6.
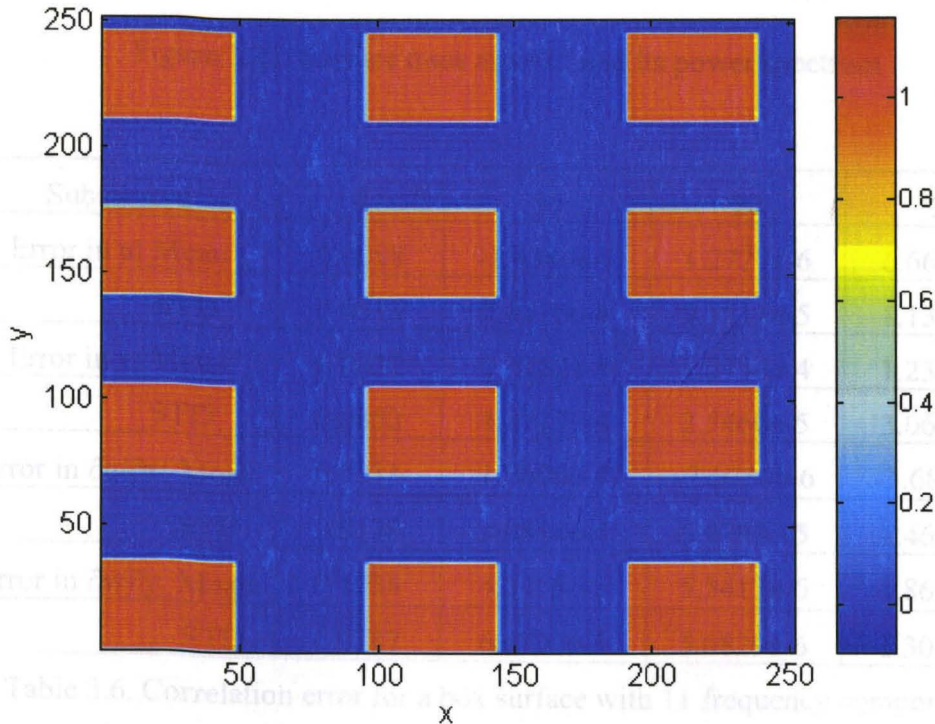
Figure 3.21. Computer generated box surface
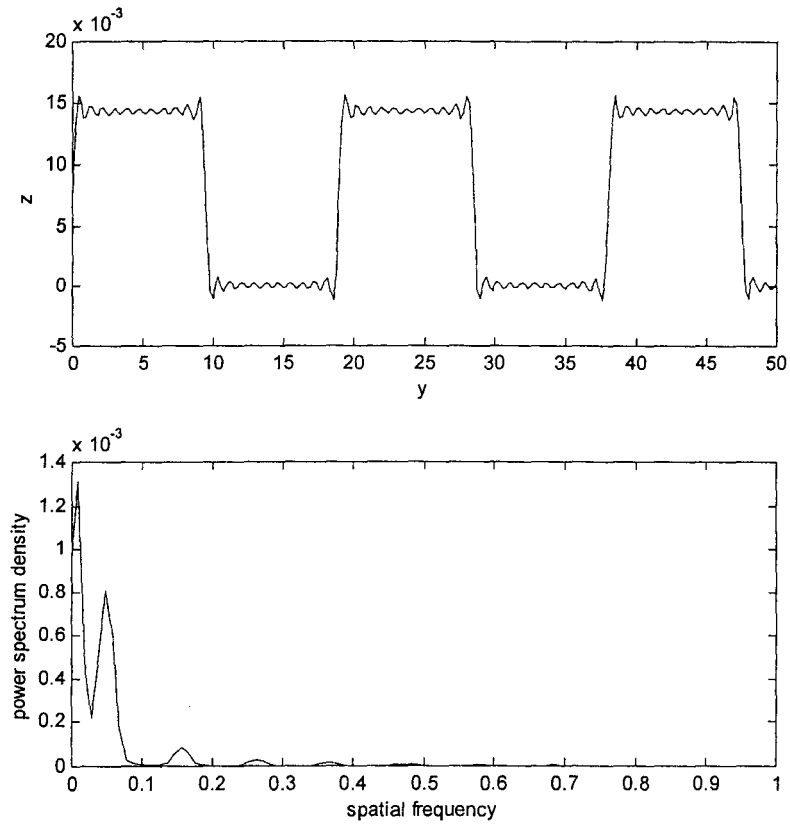
Figure 3.22. Surface trace at x=25 and its power spectrum

Subset size	4	10	20	28
Error in u: Mean	0.0058	-1.4380e-5	1.2727e-6	6.6612e-6
STD	0.0515	2.8376e-4	9.1511e-5	1.1332e-4
Error in v: Mean	0.0021	2.3162e-4	1.2794e-4	1.2305e-4
STD	0.0852	1.5717e-4	2.5460e-5	3.0604e-5
Error in $\partial u/\partial x$: Mean	0.0011	1.5606e-4	-7.6030e-6	-3.6886e-6
STD	0.0128	3.0814e-4	5.4798e-5	1.4650e-5
Error in $\partial v/\partial y$: Mean	0.1034	4.3455e-5	5.5413e-6	3.8696e-6
STD	0.3737	6.6770e-5	5.0829e-6	1.3046e-6

Table 3.6. Correlation error for a box surface with 11 frequency components

ε_{xx}=0.01, ε_{yy}=0.001, tolerance=10^{-6}

Comparing this to the results for the sinusoidal surface with a single frequency, one finds that for the same subset size, even at higher sampling frequencies, correlation results for multiple frequency surfaces are distinctly not as good as those for a single frequency. The reason is that although the dominant frequency is the same, the box surface is composed of much higher frequencies of smaller amplitude. These higher frequency components require a higher sampling rate to achieve accurate correlation results as demonstrated in section 3.2.2.1. Therefore, for the same or slightly higher sampling frequency than for a single frequency surface, correlation renders significantly larger errors. This argument is verified by the correlation results shown in Table 3.7 for box surfaces with only 2 terms

$$z = \frac{1}{4}(\sum_{i=0}^{1}\frac{\sin((2i-1)x\frac{2\pi}{w_x})}{2i-1}+1)(\sum_{j=0}^{1}\frac{\sin((2j-1)y\frac{2\pi}{w_y})}{2j-1}+1) .$$

Subset size	4	10	20	29.6
Error in u: Mean	5.2066e-7	-3.8843e-7	-5.0465e-18	9.0909e-8
STD	2.9527e-6	9.4315e-7	1.6025e-17	2.8868e-7
Error in v: Mean	-1.2645e-7	3.0579e-7	1.2066e-7	6.1157e-8
STD	6.6773e-6	1.9975e-6	1.8435e-7	1.7906e-7
Error in $\partial u/\partial x$: Mean	1.2071e-6	1.1041e-7	9.9174e-10	7.2727e-9
STD	3.3308e-6	4.1586e-7	1.1481e-7	6.7577e-8
Error in $\partial v/\partial y$: Mean	1.0732e-6	1.3312e-7	2.8992e-8	2.2810e-9
STD	9.2138e-6	2.5759e-7	9.8413e-8	1.7992e-8

Table 3.7. Correlation error for box surface with 2 terms

ε_{xx}=0.01, ε_{yy}=0.001, tolerance=10^{-6}

3.2.2.5 Uncertainty Associated with High Frequency

We consider next the question of whether adding random unwanted signal to the deformed box surface will leave the error at the same general level. To answer this question, a high frequency sinusoidal signal in both the x and the y directions is superposed on the deformed box surface with the wavelength being one-tenth of the fundamental one (14 in x and 19 in y) and 1% of its amplitude. Table 3.8 shows that the correlation error is increased by more than 3 orders of magnitude.

Subset size	4	10	20	28
Error in u: Mean	-0.0021	-1.4405E-04	2.8509E-05	3.5505E-05
STD	0.0238	0.0023	2.3485E-04	2.4048E-04
Error in v: Mean	-6.0569E-04	-2.6735E-04	5.8072E-04	2.3899E-04
STD	0.0148	0.0039	9.3240E-04	6.9673E-04
Error in $\partial u/\partial x$: Mean	0.0057	-6.8666E-04	7.8892E-05	5.8794E-05
STD	0.015	8.1891E-04	1.2206E-04	5.9988E-05
Error in $\partial v/\partial y$: Mean	-0.0026	-2.3344E-04	-2.0027E-05	-1.9859E-06
STD	0.0135	0.0017	2.0367E-04	8.2358E-05

Table 3.8. Correlation error for box surface with 2 terms with uncertainty

$\varepsilon_{xx}=0.01$, $\varepsilon_{yy}=0.001$, tolerance=10^{-6}

This indicates that regular high frequency uncertainty[*] with even small amplitude influences the results rather severely. As shown in the previous section, this could be due to an insufficient number of sampling points. Alternatively, it could also be traced to the

---

[*] This high frequency uncertainty is called regular because it has a well-defined frequency content. The influence of more random uncertainty on the correlation process will be discussed in the subsequent chapter.

fact that the superposition of the unwanted signal causes the image correlation algorithm to fail to reach the minimum at the prescribed displacements, as illustrated in section 3.1.2.5 for one-dimensional study. In order to find out which one is the main cause, one could increase the number of sampling points to see whether it improves the correlation. If it does, one then deduces that the correlation error increase associated with the introduction of uncertainty is due to an insufficient sampling rate.

However, due to the limit of the computation facility, it takes too much computation time to increase the number of sampling points by a factor of 10. So instead, we correlate a simpler surface with less frequency content, which needs fewer sampling points to achieve the specified correlation precision. Table 3.9 shows the correlation results on the single sinusoidal surface (shown in Figure 3.19) sampled at a spacing of 0.5.

Subset size	4	10	20	28
Error in u: Mean	1.3223E-08	-1.9008E-08	-1.8182E-08	-7.2727E-08
STD	3.3351E-06	2.3340E-06	7.3768E-07	5.2456E-07
Error in v: Mean	1.2314E-08	4.4653E-07	3.1653E-08	-9.1653E-08
STD	3.8121E-06	1.9431E-06	7.8264E-07	6.1459E-07
Error in $\partial u/\partial x$: Mean	3.5881E-06	4.6031E-07	-5.7455E-08	-7.9471E-08
STD	2.1899E-06	7.8691E-07	2.1845E-07	1.0946E-07
Error in $\partial v/\partial y$: Mean	4.5069E-06	1.5165E-06	1.3023E-07	-6.5545E-08
STD	2.3886E-06	3.0882E-06	4.4038E-07	3.025E-07

Table 3.9. Correlation error for doubly sinusoidal surface at a sampling spacing of 0.5
$\varepsilon_{xx}=0.01$, $\varepsilon_{yy}=0.001$, tolerance=10^{-5}

After adding unwanted sinusoidal signal to the surface in both the x and the y directions with the wavelength being one-tenth of that of the fundamental surface, and possessing 1% of its amplitude, the correlation is conducted for the same deformation. The results in

Table 3.10 show that this low amplitude, high frequency uncertainty has severely increased the correlation error.

Subset size	4	10	20	28
Error in u: Mean	4.0415E-04	2.3396E-04	-7.4382E-05	-1.5731E-05
STD	0.0101	0.0012	3.1814E-04	4.0028E-04
Error in v: Mean	0.0106	-8.6994E-04	5.7198E-04	3.6785E-04
STD	0.0184	0.0014	4.1713E-04	6.2405E-04
Error in $\partial u/\partial x$: Mean	-3.1323E-04	-1.971E-05	7.6646E-06	3.2994E-06
STD	0.0085	3.3156E-04	1.3868E-04	4.6389E-05
Error in $\partial v/\partial y$: Mean	0.0013	8.1732E-07	-3.2954E-05	9.8643E-07
STD	0.0155	0.0013	2.0713E-04	1.3262E-05

Table 3.10. Correlation error for doubly sinusoidal surface with uncertainty at a sampling spacing of 0.5, ε_{xx}=0.01, ε_{yy}=0.001, tolerance=10^{-5}

Next, the surface with uncertainty is re-sampled at a 10-time higher frequency. Correlation results for the same deformation are shown in Table 3.11.

Subset size	4	10	20	28
Error in u: Mean	0.002	7.1821E-05	-6.7131E-05	-4.1799E-05
STD	0.0109	8.8974E-04	2.7394E-04	7.6559E-05
Error in v: Mean	0.0132	-0.0014	6.4990E-04	2.2866E-04
STD	0.0182	0.0012	6.4294E-04	4.5285E-04
Error in $\partial u/\partial x$: Mean	3.7068E-04	1.7642E-04	5.5764E-07	7.0739E-06
STD	0.0071	6.2308E-04	1.2088E-04	8.2619E-05
Error in $\partial v/\partial y$: Mean	-7.6285E-04	-5.875E-05	-1.1918E-05	-5.6204E-06
STD	0.0145	0.0011	8.9993E-05	1.0307E-04

Table 3.11. Correlation error for sinusoidal surface with uncertainty at a sampling spacing of 0.05, ε_{xx}=0.01, ε_{yy}=0.001, tolerance=10^{-5}

Comparison between Table 3.10 and Table 3.11 reveals that although an increase of sampling points gives marginally better results at large subset sizes, on average, the improvement on the correlation results is barely noticeable. Therefore, we rule out the sampling rate as the key factor in increasing the correlation error when high frequency uncertainty is introduced, and deduce that the introduction of low amplitude, high frequency uncertainty adds substantial error into the correlation results, because the uncertainty causes the correlation to fail to locate the right mapping between undeformed and deformed images. This argument was also shown to hold for the one-dimensional case. It appears then that filtering out high frequency unwanted signals from the topographical images may be crucial in improving the deformation extraction of the process.

3.2.3   Real Scan Images

Because well-defined, continuous surfaces have to be interpolated before correlation, there is no additional parameter to consider for correlation on images taken from real scans, except that real scan images contain more complex frequency spectra. In the following study on two-dimensional digital image correlation of real images, a scanned image of a silicon surface, shown in Figure 3.12, is used.

3.2.3.1 Linear In-plane and Out-of-plane Deformation

The first objective is to verify that the correlation algorithm performs acceptably for real images. A prescribed displacement field with linear in-plane and out-of-plane deformations is applied numerically:

$$u=0.01+0.01(x-x_0), \quad v=0.02+0.001(y-y_0), \quad w=0.1+0.02(x-x_0)+0.05(y-y_0),$$

where (x_0,y_0) is the point at the upper left corner. Correlation is performed on the center 80x80 square and results are listed in Table 3.12.

Subset size (pixel)	10	20	40	100
Error in u: Mean	2.5290E-06	2.5290E-06	1.9472E-06	1.5933E-06
STD	2.5499E-05	6.1881E-06	1.8857E-06	2.7077E-07
Error in v: Mean	2.7470E-06	2.8327E-05	6.2673E-06	1.0671E-06
STD	9.9514E-04	1.5456E-04	2.8447E-05	4.1170E-06
Error in $\partial u/\partial x$: Mean	1.2761E-06	1.2761E-06	1.5330E-06	9.9238E-07
STD	7.8315E-06	2.0537E-06	7.7103E-07	1.1895E-06
Error in $\partial v/\partial y$: Mean	4.5051E-05	6.1415E-06	-1.4117E-06	-4.7852E-06
STD	3.5848E-04	5.7841E-05	1.1507E-05	9.6550E-06
Error in w: Mean	1.7405E-07	1.7405E-07	5.0309E-08	-1.8578E-08
STD	3.5525E-06	7.7975E-07	1.8150E-07	7.9078E-08
Error in $\partial w/\partial x$: Mean	-6.7968E-06	-5.1939E-07	-1.6598E-06	-1.8386E-06
STD	3.5525E-06	2.2395E-05	3.2562E-06	3.7430E-07
Error in $\partial w/\partial y$: Mean	5.6990E-06	-3.2625E-06	-4.2601E-06	-4.7698E-06
STD	1.0208E-04	1.4056E-05	1.9367E-06	6.4772E-07

Table 3.12. Correlation error for a real scan image

$\varepsilon_{xx}=0.01$, $\varepsilon_{yy}=0.001$, $w=0.1+0.02x+0.05y$, tolerance$=10^{-6}$

The correlation algorithm functions relatively well, although the errors are slightly larger than for the well-defined Fourier surfaces. The level of error remains the same even when high frequency components of the image are filtered out. This suggests that the errors most likely come from the interpolation process.

3.2.3.2 Inhomogeneous Deformation

Digital image correlation is generally expected to be applied to inhomogeneous deformations such as, e.g., those near a crack tip. It is, therefore, important to understand how well inhomogeneous deformations can be extracted. For this purpose, the following displacement field with comparable linear and non-linear components is applied to the silicon surface shown in Figure 3.12.

$$u = 0.01 + 0.01(x - x_0) + 0.001(x - x_0)^2 \text{ *}$$
$$v = 0.05 + 0.01(y - y_0) + 0.01(y - y_0)^2$$

Subset size (pixel)	10	20	40	100
Error in u: Mean	1.5687E-06	3.8416E-06	4.5794E-06	2.2899E-05
STD	5.5068E-05	1.1015E-05	6.5794E-06	1.1853E-05
Error in $\partial u/\partial x$: Mean	-6.3099E-05	2.1811E-05	6.5803E-06	-1.6799E-05
STD	1.4000E-03	2.7349E-04	8.1752E-05	6.0065E-05
Error in v: Mean	4.9364E-06	1.0764E-05	3.4601E-05	2.1424E-04

* A more dominant second order term is imposed on v than u to examine its influence on the correlation when insufficient displacement representation is used for correlation.

STD	6.1546E-05	1.8314E-05	1.1085E-05	3.1601E-05
Error in $\partial v/\partial y$: Mean	3.4048E-04	7.5458E-05	2.5262E-05	9.0002E-05
STD	2.9000E-03	5.5184E-04	1.9754E-04	2.2343E-04

Table 3.13. Correlation error using $\{u, \partial u/\partial x, \partial u/\partial y, v, \partial v/\partial x, \partial v/\partial y, w\}$, tolerance=$10^{-6}$

The quality of the correlation process using (3.2.6)

$$P(F) = \{u(F), \frac{\partial u}{\partial x}(F), \frac{\partial u}{\partial y}(F), v(F), \frac{\partial v}{\partial x}(F), \frac{\partial v}{\partial y}(F), w(F)\}$$

to represent displacements is then measured by errors illustrated in Table 3.13. One notes that the error decreases initially as the subset size increases from 10 pixels to 40, but then starts to increase for larger subset sizes. The reason is that for small subsets, the quadratic term in the displacements is not dominant, so that the correlation results follow the same trend as for linear deformations. However, once the subset gets large enough so that the quadratic term is comparable to the linear term, the assumption of linear deformation within subsets does not hold any more. Therefore, the error begins to increase for larger subsets. This observation is verified by comparing the results for u and v. Because the non-linear component is more dominant in v, the threshold subset size over which the correlation error starts to increase for larger subsets is smaller. In addition, for the same subset size of 100 pixels, the error in v is larger by one magnitude.

Subset size (pixel)	10	20	40	100
Error in u: Mean	5.9027E-06	2.1717E-06	1.9422E-06	1.3354E-06
STD	1.0870E-04	2.4921E-05	7.9567E-06	3.4580E-06
Error in $\partial u/\partial x$: Mean	4.8001E-05	2.8564E-05	1.1972E-05	1.9307E-06
STD	2.5000E-03	3.6053E-04	5.7230E-05	7.1462E-06
Error in v: Mean	-6.4945E-06	3.1079E-06	3.9865E-06	1.2340E-06

STD	1.2120E-04	3.2478E-05	1.3765E-05	7.1102E-06
Error in $\partial v/\partial y$: Mean	8.0229E-04	1.0048E-04	2.4250E-05	-1.2927E-05
STD	4.7000E-03	6.7578E-04	1.4307E-04	2.9799E-05

Table 3.14. Correlation error using $\{u,\partial u/\partial x,\partial u/\partial y,\partial^2 u/\partial x^2,\partial^2 u/\partial x\partial y,\partial^2 u/\partial y^2, v,\partial v/\partial x,\partial v/\partial y,\partial^2 v/\partial x^2,\partial^2 v/\partial x\partial y,\partial^2 v/\partial y^2,w\}$, tolerance=$10^{-6}$

Next we examine whether the addition of the second order displacement gradients improves the correlation results for large subset size.

$$(3.2.13) \qquad P(F) = \{u,\frac{\partial u}{\partial x},\frac{\partial u}{\partial y},\frac{\partial^2 u}{\partial x^2},\frac{\partial^2 u}{\partial x\partial y},\frac{\partial^2 u}{\partial y^2},v,\frac{\partial v}{\partial x},\frac{\partial v}{\partial y},\frac{\partial^2 v}{\partial x^2},\frac{\partial^2 v}{\partial x\partial y},\frac{\partial^2 v}{\partial y^2},w\}\Big|_F$$

is used to represent the displacement field[*]. We find that for small subset sizes less than a threshold value (40 for u, 20 for v), the results are comparable to those without the second order terms. But for large subset sizes, the improvement in the results is significant, especially for v, in which the quadratic term is more dominant.

This observation is consistent with the study on the out-of-plane deformation in section 3.2.2.3 in the sense that correlation provides results within the specified error bound when sufficient displacement terms are included to represent the prescribed displacement field.

---

[*] Higher order terms for w are not included, because there is no out-of-plane displacement.

3.2.3.3 The Effect of Uncertainty

As demonstrated in section 2.2.2.5, high frequency unwanted signal of even very small amplitudes can introduce disproportionately large errors into correlation results. However, in real scans via probe microscopy, noise is intrinsic in the instrumentation. Thus the challenge is to filter out the unwanted signal and extract the real deformation from the undeformed and deformed images. It is here proposed to address this endeavor by first performing repetitive scans before the deformation is applied. By comparing the repetitive scans, a frequency spectrum of the uncertainty can be obtained. This serves as the criterion to determine the threshold of the high frequency to be filtered out. If the scans are sampled at a sufficient fine spacing and the uncertainty is small compared to the primary signal, this method should be able to extract the real topography of the surface, to then determine the deformation field from the undeformed and deformed images. A more complete study on the noise will follow in the next chapter.

3.2.4 Conclusion

This study on two-dimensional digital image correlation systematically identifies and evaluates the important parameters in the process. As in one-dimensional correlation, it has been confirmed that Newton-Raphson algorithm functions well in the two-dimensional correlation. One difficulty in the two-dimensional correlation lies in the impossibility to work with the analytical form of any well-defined surfaces, because the number of parameters involved becomes too large. Therefore, interpolation between adjacent pixels is needed even for an initial exploration of two-dimensional correlation.

It has been verified quantitatively that cubic spline interpolation performs better than linear and quadratic interpolations. Because of interpolation, the spatial sampling rate becomes an important factor in the correlation process. Given the tolerance 10^{-6}, a rough number of 50-75 sampling points per wavelength are needed to achieve accurate results, or to obtain correlation error within the tolerance. However, given a fixed sampling rate due to the limitation of experimental apparatus, larger subsets do improve correlation results for homogeneous in-plane deformations. When more than one spatial frequency exist in the surface, the sampling rate is the determining factor for good correlation results. In contrast to one-dimensional digital image correlation, two-dimensional correlation provides comparable accuracy for the displacements and their derivatives. When out-of-plane deformation exists, a sufficient number of out-of-plane terms (enough to represent the displacement field to the appropriate degree of a Taylor expansion) need to be included. As in one-dimensional correlation, high frequency uncertainty with low amplitude between the undeformed and deformed images can introduce large errors into the correlation results. Therefore, it is necessary to filter out high frequency unwanted signals before correlation is initiated.

For inhomogeneous deformations, the assumption of linear deformation within a subset is still useful if sufficiently small subsets can be used. The more dominant the non-linear contribution is in the deformation, the larger the correlation error. However, if sufficient non-linear displacement gradient terms are included to represent the prescribed displacement field, correlation derives acceptable results regardless of the subset size. Because noise is intrinsic in experimental data, a calibration method to determine how to

filter out high frequency noise is proposed. This will then be tested by means of real scanned images in the next chapter.

4 Noise Analysis

The study of both one-dimensional and two-dimensional digital image correlation reveals that uncertainty between the undeformed and deformed signals has a significant influence on the correlation results. Noise is intrinsic in experimental measurements. Therefore, it is important to understand the nature and characteristics of the noise before one can fully evaluate its influence on the digital image correlation. In this chapter, the noise originated from a commercial atomic force microscope (AFM) used to perform surface measurements on silicon and other materials is examined. First, the temporal noise when the AFM is not scanning is collected and its frequency content is analyzed. A low-pass filter is applied and the comparison between the filtered and unfiltered signals is discussed. When the AFM is in the scanning mode, this temporal noise is translated to an inconsistency or uncertainty among repetitive scans. Digital image correlation is performed on the line profiles from consecutive scans and the result is discussed. Finally, two-dimensional AFM scan images are presented. Problems with these images are examined and a possible improvement to compensate for the hysteresis of the AFM piezo actuator is proposed.

4.1 Temporal Noise

As a first step analysis on the noise originated from the atomic force microscope, the probe is kept positioned over a fixed point on the surface. The height reading is recorded

as a function of time shown in Figure 4.1. Four independent readings are considered. Except for the irregularities at the beginning of each data set, the temporal noise ranges from 1 to 3nm.
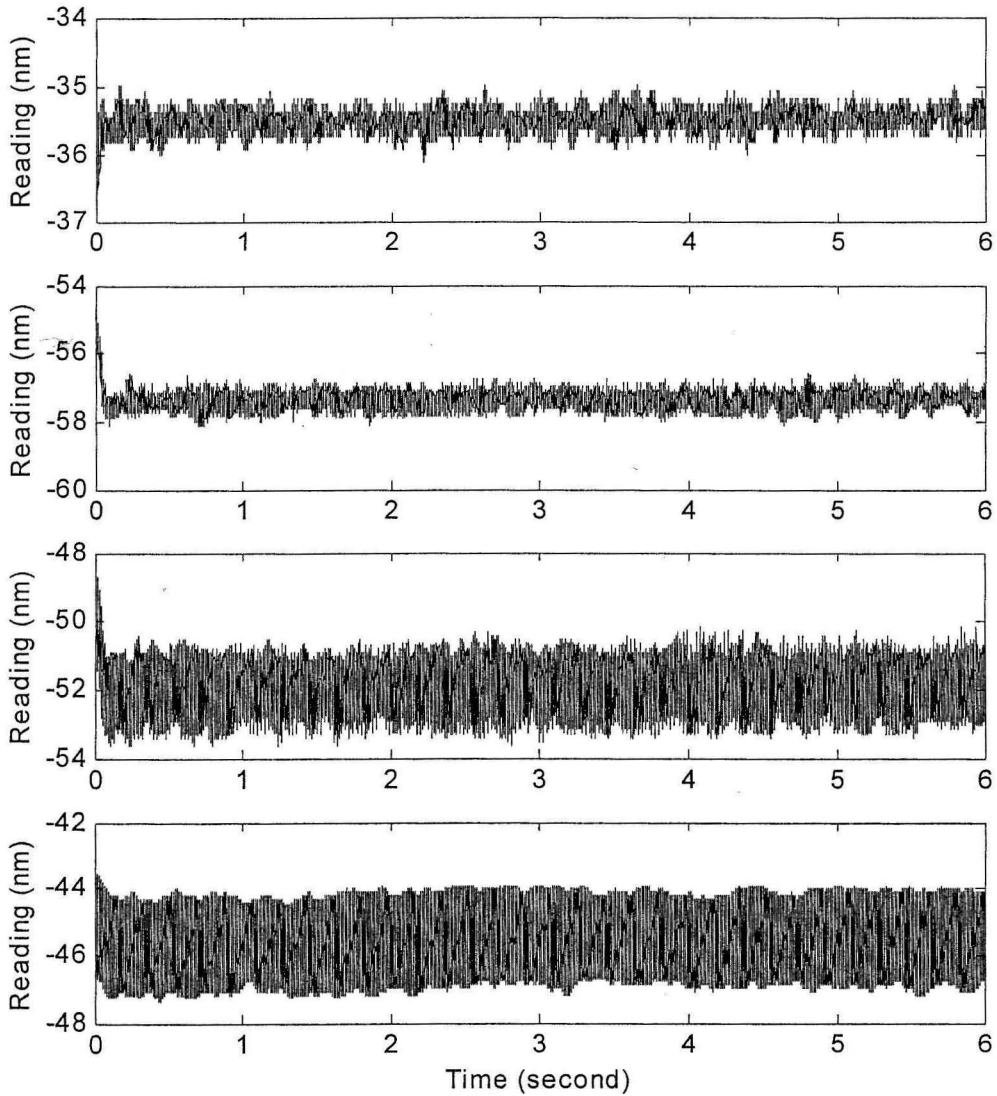

Figure 4.1. Noise as a function of time

To analyze the frequency content of the temporal noise, a Fourier transform is applied to the first record in Figure 4.1. The initial regularity in the record is not considered for the frequency analysis, because it might have an influence on the analysis. The part of the signal used is shown in Figure 4.2 and its frequency spectrum is displayed in Figure 4.3. In the frequency spectrum, peaks are found at frequencies of approximately 90Hz, 170Hz and 250Hz. A low-pass filter at a cut-off frequency at around 80Hz is applied to the signal. The filtered signal and its frequency spectrum are displayed in the same figures for comparison. Figure 4.2 shows that the amplitude of the signal has been reduced from about 1nm to 0.3nm. The remaining noise is mostly white noise, with a uniform frequency distribution as shown in Figure 4.3 (b).
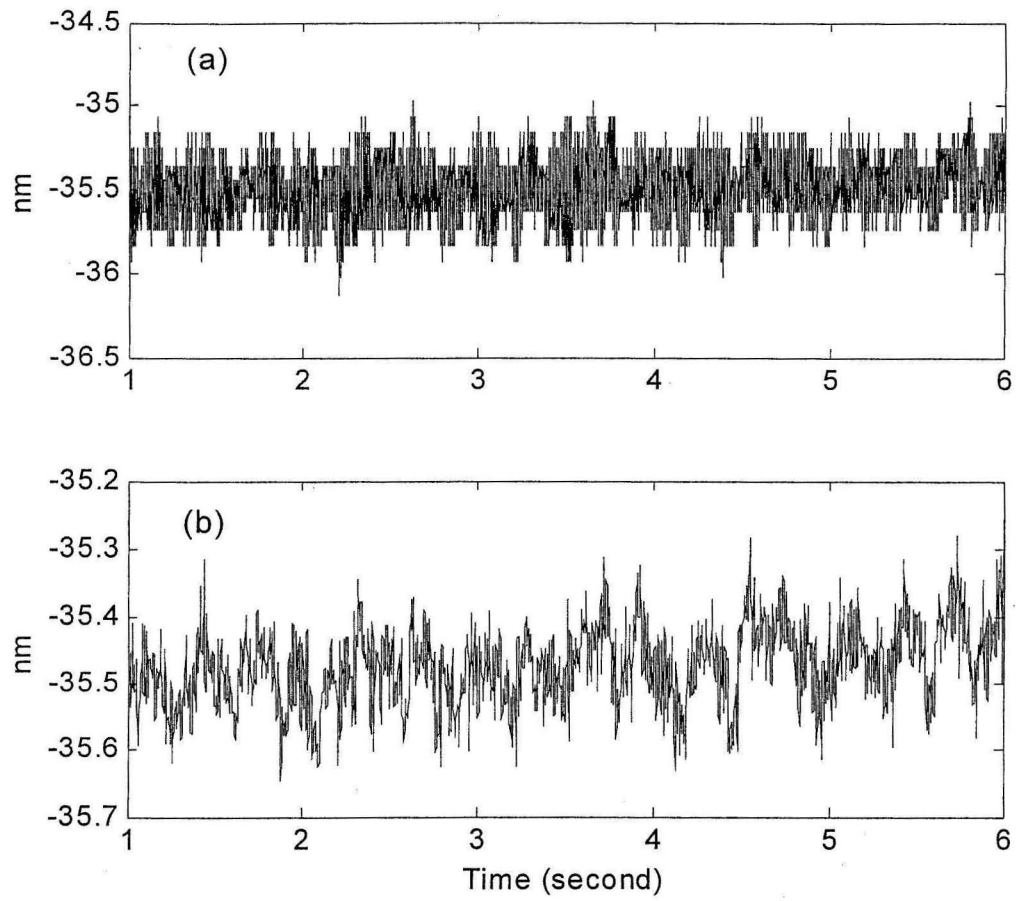
Figure 4.2. Temporal noise
(a) Original signal, (b) Filtered signal at a cut-off frequency ~80Hz

Figure 4.3. Frequency spectrum
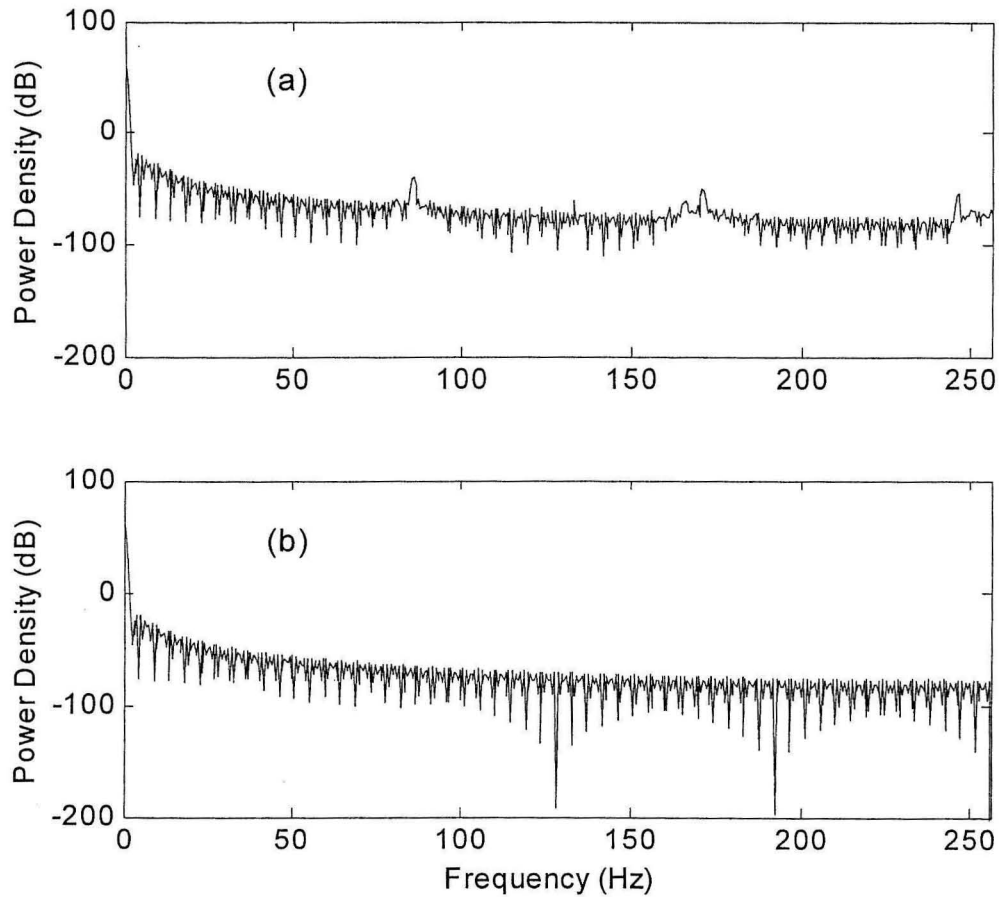(a) Original signal, (b) Filtered signal at a cut-off frequency ~80Hz

During scanning, this temporal noise will be translated to an inconsistency among repetitive scans, or a spatial uncertainty. This spatial uncertainty, in turn, will generate an "artificial" displacement, when digital image correlation is performed on repetitive scans. This will be examined in detail in the next section.

4.2 Spatial Uncertainty

Next, repetitive line scans are performed on a silicon surface. The line profiles look consistent at a first glance at Figure 4.4. However, a closer comparison between the first line taken from the top two images in Figure 4.4 demonstrates that the repeatability holds to within a certain range (up to 3nm) that is consistent with the amplitude of the temporal noise. Figure 4.5 shows that this range of the difference remains the same, even after the "high spatial frequency" noise (translated from temporal noise) is filtered out.

Figure 4.4. Repetitive line scans

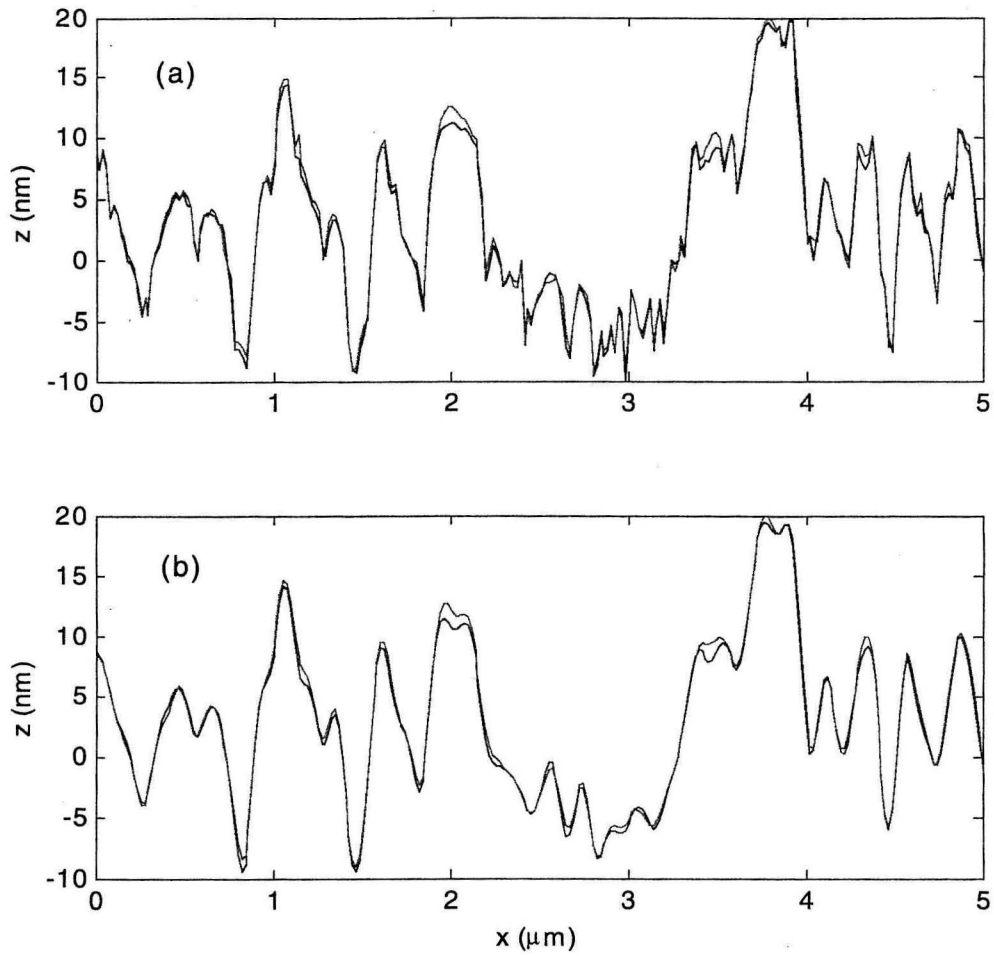Figure 4.5. Line profiles taken from two consecutive scans
(a) Original signal, (b) Filtered signal

Although amplitude is small, the uncertainty between these two line profiles is not negligible when the digital image correlation is performed. Figure 4.6 shows that "artificial" displacements for up to half a pixel (corresponding to about 5nm) have been generated from this inconsistency of repetitive scans.
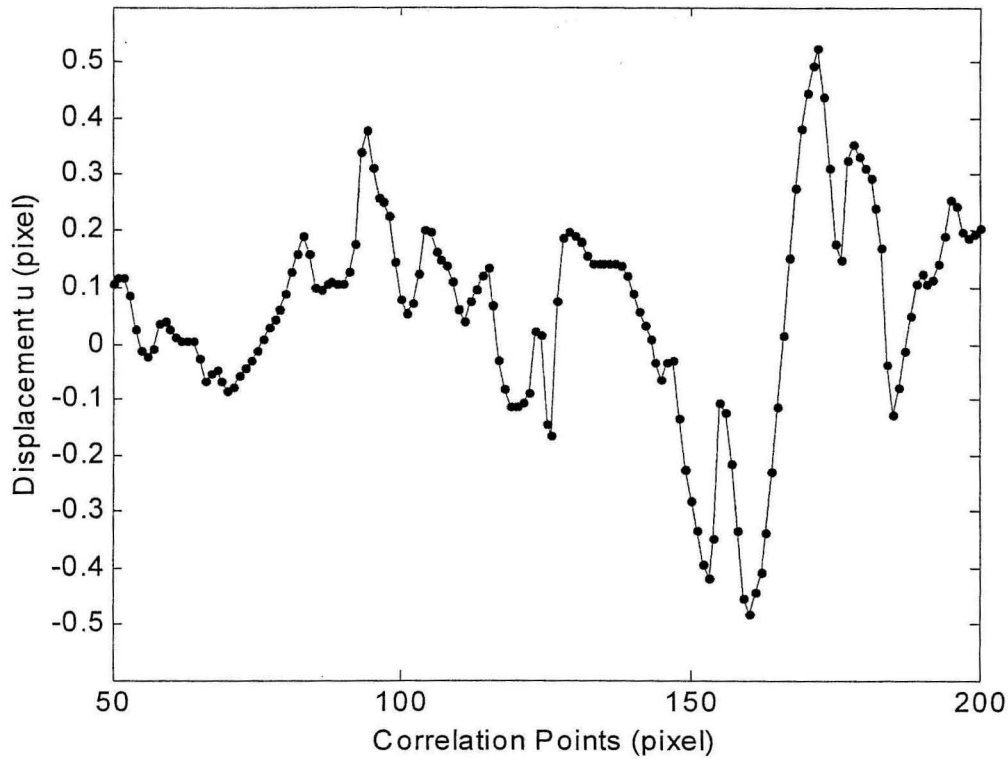
Figure 4.6. "Artificial" displacement from repetitive line scans

This "artificial" displacement will cloud the calculation of real displacements when image correlation is performed for a deformed specimen. To find out how this affects the image correlation quantitatively, a uniform tension in the x direction $\varepsilon_x=0.01$ with fixed end at x=0 pixel is applied to one of the two filtered line profiles in Figure 4.5, (b). Correlation is then performed, with a subset size of 20 pixels and a tolerance of 10^{-5}. The comparison between the calculated displacements and the prescribed displacements is shown in Figure 4.7. The error is consistent with the "artificial" displacement generated by the repetitive lines scans shown in Figure 4.6.
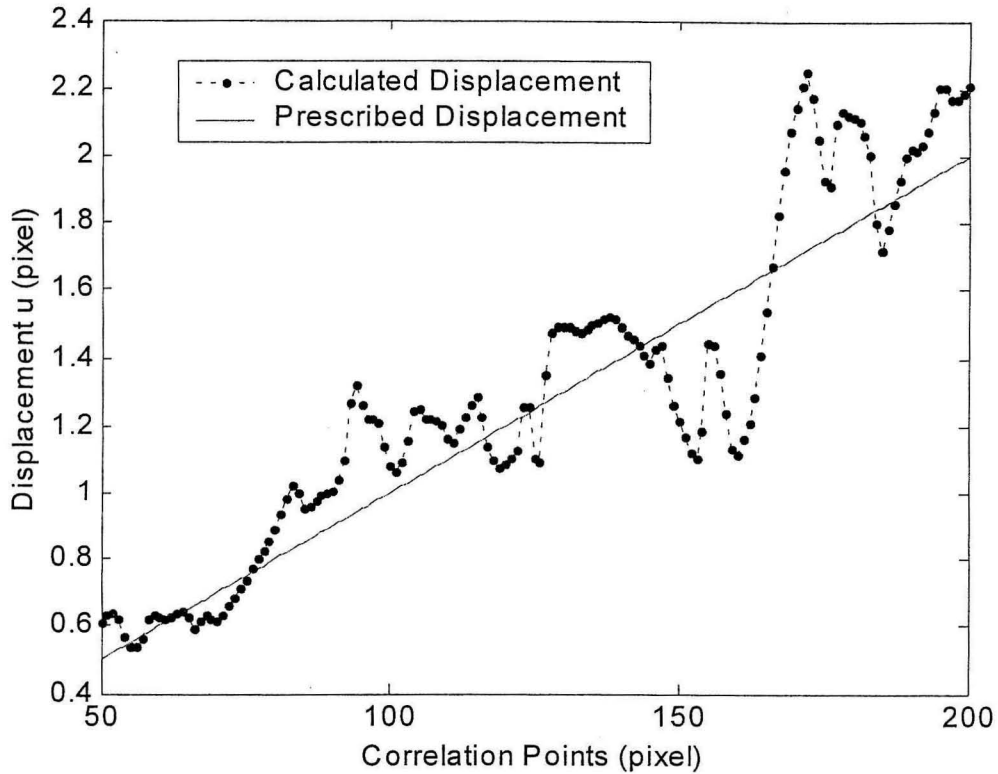
Figure 4.7. Calculated displacement for homogeneous deformation ε_x=0.01, subset=20pixel, tolerance=10^{-5}

For this particular case, the error in the displacement u from the image correlation is comparable to the magnitude of the prescribed displacement. However, this error remains constant regardless of the deformation applied. For example, when a uniform strain ε_x=0.1 is applied, the error will appear much smaller compared to the prescribed displacement, simply because the prescribed displacement is 10 times larger. Moreover, even for a deformation ε_x=0.01, if the points to be correlated are far away from the fixed end so that the displacements at these points are larger, the error will again appear smaller in comparison.

Because the uncertainty between repetitive scans originates from random temporal noise (shown in Figure 4.1) that is intrinsic to the instrumentation of the atomic force microscope, it is not possible to predict the pattern of the uncertainty by means of calibrations. One can only deduce that the calculated displacement from the AFM scans may have an error as large as half a pixel, but one cannot eliminate this error. However, by carefully choosing the area to be correlated (that has large displacements within the specimen), one can minimize the relative error and extract the strains more accurately.

4.3   Problems with Two-Dimensional Images

After the study on repetitive line scans, we turn to two-dimensional scan images. As for the line scans, a Si_3N_4 surface is scanned repetitively with the AFM and shown in Figure 4.8. These images show that more serious problems exist for two-dimensional AFM scans. Movements are observed along both in-plane directions. In addition, there exist color "bands" along the x direction. This is because of the rastering motion of the AFM probe: it scans along the y direction first, the same as the scanning sequence shown in Figure 2.15 for the STM. Both of these problems originate from the hysteresis of the piezo-ceramic actuators, as mentioned in section 2.2. Specifically, because of the hysteresis, the probe may not return to the same point after each line scan along y. If the probe is slightly farther away from the surface, the starting point for the next line scan would appear slightly higher in the scanned image; if the probe is slightly closer from the surface, the starting point for the next line scan would appear slightly lower in the scanned image. This explains the obvious inclination of the scan along x direction. When

the change in the probe position after each line scan is not exactly the same, color bands will form, as shown in Figure 4.8. Hysteresis also takes place on the two in-plane directions, which explains the translation between the two images.
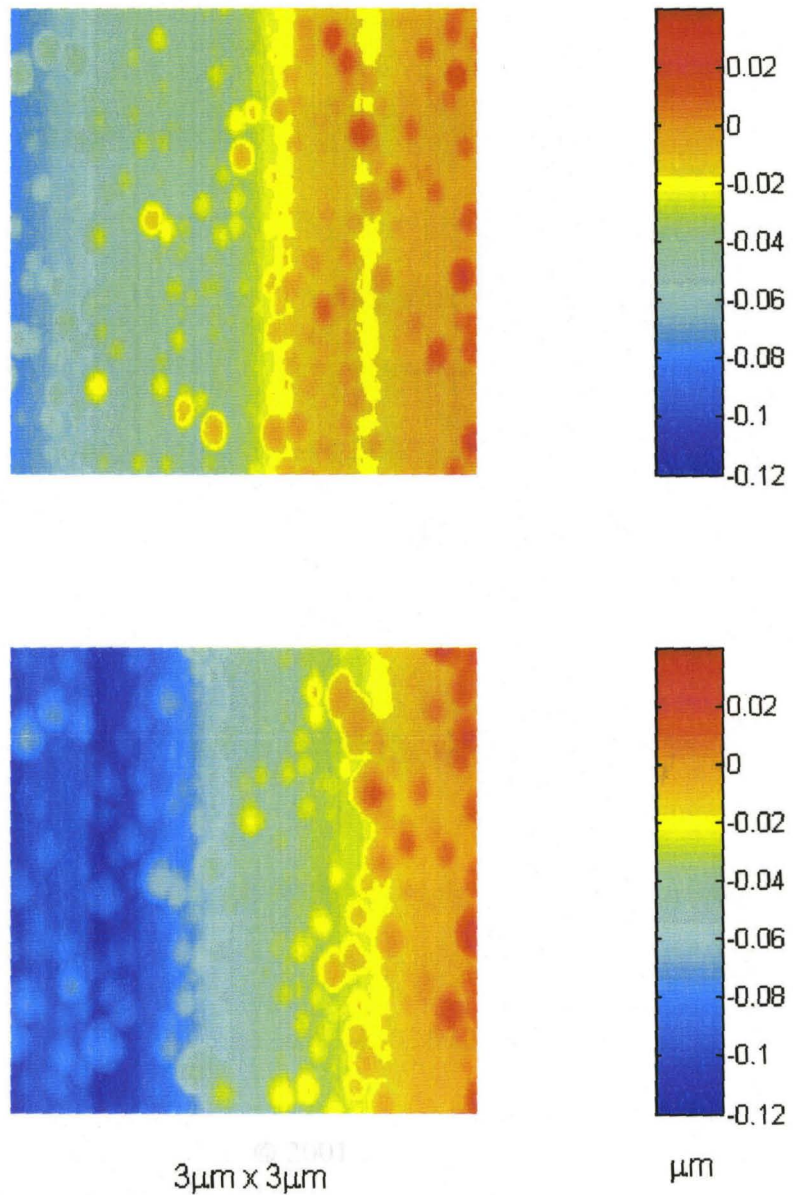

Figure 4.8. Two repetitive scans

The hysteresis problem is intrinsic to the instrumentation of the AFM. By carefully selecting various parameters involved in the scan, one can reduce the influence of the hysteresis. However, it is unlikely that one can eliminate the hysteresis completely. Therefore, it is proposed here for the AFM manufacturer to follow a procedure to minimize the effect of the hysteresis: after each line scan, the position of the tip is compared with that at the starting point. If any difference exists, the probe is brought back to exactly the same position, before performing the next line scan. After the two-dimensional scan is completed, the probe moves back to its starting position, which should be exactly the same position as the original point, so that repetitive scans will always start from the same position. This procedure can be easily implemented in the AFM software.

4.4 Conclusion

Noise analysis on scan images from the atomic force microscope has been performed in this chapter. Data recording when the AFM probe is not scanning reveals that the amplitude of the temporal noise is up to 3nm. In the frequency spectrum, peaks are observed at around 90Hz, 170Hz and 250Hz. After a low-pass filter with a cut-off frequency at about 80Hz is applied, the amplitude of the temporal noise is reduced by two thirds. The remaining noise is white noise, which is impossible to eliminate completely. When the AFM probe is scanning, the temporal noise is translated into a spatial uncertainty or inconsistency for repetitive line scans. The magnitude of the uncertainty is again up to 3nm. Due to this uncertainty, when digital image correlation is performed on

these repetitive line scans, an "artificial" displacement of up to half a pixel (corresponding to 5nm) is recovered. This sets an inaccuracy for displacement measurements by means of the atomic force microscope and the digital image correlation: When the specimen is deformed, the extracted displacement can contain an error of up to 5nm. Although this error is intrinsic to the AFM, careful selection of the points to be correlated within a line segment subject to certain deformation can reduce the relative error. For example, when a uniform tension is applied to a specimen, correlating points farther away from the fixed end will produce less error in the strain calculation, because the prescribed displacements of such points are larger.

For two-dimensional scans, the AFM exhibits more serious problems, because the hysteresis of the piezo-ceramic actuator is not compensated. The hysteresis in the in-plane directions make repetitive scans translated with respect to each other, while the hysteresis in the out-of-plane direction creates inclination and color band problems for the scanned image. A solution is proposed to force the probe to come back to the same position after each line scan before moving to the next line as well as after the whole image scan.

References

[1] Knauss W.G., "Perspectives in Experimental Solid Mechanics," International Journal of Solids and Structures, 37, 251-266 (January 2000)

[2] Howland R., Benatar L., A Practical Guide to Scanning Probe Microscopy, Park Scientific Instruments

[3] Chen Julian, Introduction to Scanning Tunneling Microscopy, Oxford University Press, 1993

[4] J.G. Simmons, "Generalized Formula for the Electric Tunnel Effect between Similar Electrodes Separated by a Thin Insulating Film," J. Appl. Phys. 34, 1973-1803 (1963)

[5] M.D. Pashley, J.B. Pethica, J.H. Coombs, "Scanning Tunneling Microscope Studies," Surface Science, 152, 27-32 (1985)

[6] Vendroux G., "Scanning Tunneling Microscopy in Micromechanics Investigations," Ph.D. thesis, California Institute of Technology (1994)

[7] Vendroux G., Knauss W.G., "Submicron Deformation Field Measurements: Part 1. Developing a Digital Scanning Tunneling Microscope," Experimental Mechanics, 38(1), 18-23 (March 1998)

[8] Vendroux G., Schmidt N., Knauss W.G., "Submicron Deformation Field Measurements: Part 3. Demonstration of Deformation Determinations," Experimental Mechanics, 38 (3) 154-160 (September 1998)

[9] Knauss W.G., "Fundamental Studies related to the Strength of Interfaces and Adhesive Bonds," Final Technical Report on a Research Program ONR Grant N00014-91-J-1427, June 1999

[10] Horowitz P., Hill W., The Art of Electronics, Cambridge University Press, 1989

[11] Stout D.F., Handbook of Operational Amplifier Circuit Design, McGraw–Hill Book Company, 1976

[12] Gene F. Franklin, J. David Powell, Abbas Emami-Naeini, Feedback Control of Dynamic Systems, Addison-Wesley Publishing Company, 1994.

[13] Peters W.H., Ranson W.F., "Digital Image Techniques in Experimental Stress Analysis," Opt. Eng., 21(3), 427-432 (1982)

[14] Sutton M.A., Wolters W.J., Peters W.H., Ranson W.F., McNeil S.R., "Determination of Displacements Using an Improved Digital Image Correlation Method," Image Vision Computing, 1(3), 133-139 (1983)

[15] Sutton M.A., Cheng M., Peters W.H., Chao Y.J., McNeil S.R., "Application of an Optimized Digital Image Correlation Method to Planar Deformation Analysis," Image Vision Computing, 4(3), 143-150 (1986)

[16] Bruck H.A., McNeil S.R., Sutton M.A., Peters W.H., "Digital Image Correlation Using Newton-Raphson Method of Partial Differential Correction," Experimental Mechanics, 29(3), 261-267 (1989)

[17] Luo P.F., Chao Y.J., Sutton M.A., Peters W.H., "Accurate Measurement of Three-dimensional Deformations in Deformable and Rigid Bodies Using Computer Vision," Experimental Mechanics, 50(2), 123-132 (1993)

[18] Vendroux G., "Correlation: A Digital Image Correlation Program for Displacement and Displacement Gradient Measurements," GALCIT Report SM90-19, California Institute of Technology (1990)

[19] Vendroux G., Knauss W.G., "Submicron Deformation Field Measurements: Part 2. Improved Digital Image Correlation," Experimental Mechanics, 38(2), 86-91 (June 1998)

[20] Oppenheim A.V., Schafer R.W., Discrete-Time Signal Processing, Prentice-Hall, 1989

[21] Gisela Engeln-Müllges, Frank Uhlig, Numeric Algorithms with C (with CD-ROM), Springer, 1996

Appendix A. Etching Procedures for Amplifier Circuit Board

The amplifier circuit boards as described in section 2.3.2.2 are etched out in the lab. Here are the procedures to follow:

1. Make the drawing in AutoCAD. Due to the difficulty in controlling the print size in AutoCAD, the drawing can be exported to an .eps file in MS Word where the size can be adjusted easily.

2. Print the drawing on a special paper that can be later ironed to the copper board. If .eps format is used for printing, the printer needs to be a postscript printer.

3. Use a household iron to iron the print on the copper board. During ironing, turn iron to dry (no steam), temperature medium to high, and push the iron hard on copper board for about 3 minutes.

4. Before the board cools down, soak it into water. Wait till the print comes off by itself. This takes one minute or so. If the print does not come out perfect, an etching pen can be used to make up the parts not showing out well. Use duct tape to cover the other side that should not be etched.

5. Immerse the board in etching solution completely and agitate the solution constantly. In 20 minutes or so, the etching is done. If the etching solution is cold or reused, etching may take longer to finish.

The materials needed for etching can be found in any electronics store.