2006

Transient force atomic force microscopy: systems approaches to emerging applications

Deepak Ranjan Sahoo
Iowa State University

Transient Force Atomic Force Microscopy:

Systems approaches to emerging applications

by

Deepak Ranjan Sahoo

A dissertation submitted to the graduate faculty

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Electrical Engineering

Program of Study Committee:
Murti V. Salapaka, Major Professor
Wolfgang Kliemann
Degang Chen
Nicola Elia
Zhengdao Wang

Iowa State University

Ames, Iowa

2006

Graduate College
Iowa State University

This is to certify that the doctoral dissertation of

Deepak Ranjan Sahoo

has met the dissertation requirements of Iowa State University

Signature was redacted for privacy.

Major Professor

Signature was redacted for privacy.

For the Major Program

DEDICATION

To my parents Lambodar Sahoo and Nalini Prava Dei, sister Dipti and brother Dibya.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1 TRANSIENT FORCE ATOMIC FORCE MICROSCOPY

In the existing dynamic mode operation of Atomic Force Microscopes (AFMs) steady-state signals like amplitude and phase are used for detection and imaging of material. Due to the high quality factor of the cantilever probe the corresponding methods are inherently slow. In this dissertation, a novel methodology for fast interrogation of material that exploits the transient part of the cantilever motion is developed. This method effectively addresses the perceived fundamental limitation on bandwidth due to high quality factors. It is particularly suited for the detection of small time scale tip-sample interactions. Analysis and experiments show that the method results in significant increase in bandwidth and resolution as compared to the steady-state-based methods. This article demonstrates the effectiveness of a systems perspective to the field of imaging at the nano-scale and for the first time reports *real-time* experimental results and scanning applications of the transient method.

1.1 Introduction

Desirable properties of manufactured products arise from the manner in which atoms are arranged in its material. Until recently, ways of manipulating and interrogating matter were limited to aggregate methods where the control and investigation of matter was achieved at scales much larger than atomic scales. The investigation (placement) of material to decipher (build) complex structures atom by atom was not possible, and thus optimal desired specificity of the material properties was not achievable.

Recent demonstrations of nanoscience provide ample evidence indicating the feasibility of rational *control, manipulation and interrogation* of matter at the atomic scale. There is a promise that material can be tailored at the atomic scale. The atomic force microscope (AFM)

[1] is an instance of the impact of nanotechnology where a cantilever is utilized to image and manipulate sample properties (see Figure 2.1). As stated in the National Nanotechnology Initiative Plan [2], *"These instruments, including scanning tunneling microscopes, atomic force microscopes, and near-field microscopes, provide the eyes and fingers required for nanostructure measurement and manipulation."*

In a typical AFM (see Figure 2.1), the deflection of the cantilever due to the sample is registered by the laser incident on the cantilever tip, which reflects into a split photo-diode. The piezoelectric scanner is used to position the sample. In static mode the cantilever deflection is solely due to the tip-sample interaction. The piezoelectric scanner is rastered in the lateral direction and the deflection of the tip is used to interpret sample properties. In the dynamic mode, the cantilever support is forced sinusoidally using a dither piezo. The changes in the oscillations introduced due to the sample are interpreted to infer its properties.

Cantilevers have been utilized in biological sciences to perform remarkable feats such as cutting DNA strands [3] and investigating the activity of RNA polymerase (a protein complex) [4]. On a similar note, there are impressive proposals on using cantilever-based nanoprobes to interrogate cell dynamics with significant impact on human health. Another intriguing application of the cantilever is in the detection of single electron spin based on magnetic detection principles [5, 6, 7, 8]. Such research has significant ramifications for quantum computing technology. Yet another putative application of cantilever-based sensing is its use in detecting toxic chemical and biological agents with very high sensitivity, with obvious significance to security.

In spite of the underlying promise of cantilever-based technology, considerable challenges need to be overcome. Pivotal to harnessing its vast potential is ultra-fast *interrogation.* This is apparent as the interrogation, for example, of atoms or spins of electrons needs to be accomplished for material that has macroscopic dimensions. Most of the current techniques rely on the steady-state data of the cantilever tip position. This results in very slow interrogation speeds due to the typical high quality factor of the cantilever that lead to high settling times. It is to be noted that high quality factors are a must for high resolution. Thus, in existing steady state methods a tradeoff has to be achieved between bandwidth and resolution.

In this dissertation, the hitherto unexplored idea of using the transient signals for increasing bandwidth and resolution is presented. A preliminary version of this work has appeared as a letter in Reference [9]. Sensing based on transient signal has the potential to detect tip-sample interaction changes at extremely high rates. This method has the promise of detecting one event in 4 periods of cantilever oscillation, which translates to 25 million events per second for a 100 MHz cantilever. Such a technology has tremendous implications in numerous applications, for example, in cantilever-based retrieval of high-density data and detection of chemical and/or biological molecules on a surface. Researchers at IBM [10] have demonstrated areal densities of upto $3Tb/in^2$ (for data storage), however, the reading is performed in static mode (contact mode) which results in signal deterioration due to wear and, also needs to be corrected for thermal drift during extended periods of operation. Furthermore, the wear becomes more severe as the data rates are increased. Transient signal detection being a dynamic method is gentle and has virtually no wear. Moreover, as will be shown later the data rates depend only on the cantilever frequency and to a large extent independent of the quality factor Q. Thus, the high resolution needs can be effectively decoupled from the high bandwidth needs.

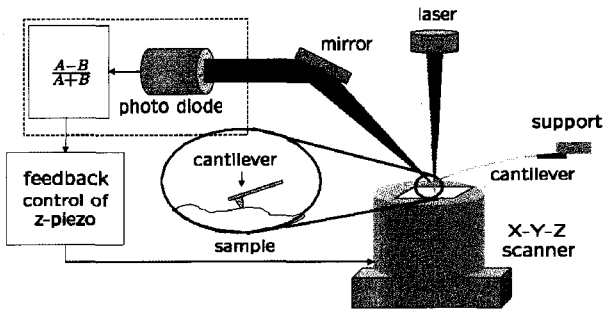


Figure 1.1 In a typical AFM, the deflection of the cantilever due to the sample is registered by the laser incident on the cantilever tip, which reflects into a split photo-diode. The piezoelectric scanner is used to position the sample. In static mode the cantilever deflection is solely due to the tip-sample interaction. The piezoelectric scanner is rastered in the lateral direction and the deflection of the tip is used to interpret sample properties. In the dynamic mode, the cantilever support is forced sinusoidally using a dither piezo. The changes in the oscillations introduced due to the sample are interpreted to infer its properties.

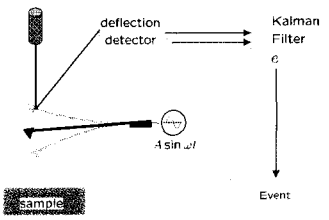


Figure 1.2 In the new architecture an observer (Kalman filter) is designed based on a model of the cantilever that estimates the deflection of the cantilever from the photo-diode signal and the dither excitation signal. The error in estimation e arising due to the sample force is utilized to detect the tip-sample interactions as events.

A systems viewpoint of the AFM dynamics and observer-based approach provide the basic analytical tools to investigate the transient signals. In Section 1.2, the state-space description of the cantilever dynamics is introduced. Section 1.3 describes the observer architecture that tracks the transient signals. Further assumptions on the character of tip-sample interactions can be used to analyze the resolution and bandwidth of the proposed scheme. Moreover, an efficient detection scheme is developed using tools from statistical signal processing. These results are presented in Section 1.4. Experimental results presented in Section 1.7 confirm the effectiveness of the new methodology.

In this article apart from providing a detailed exposition of the transient method, we present the significant step of realizing the transient scheme in the realtime unlike the offline analysis of data in Reference [9]. Furthermore, for the first time scanning results utilizing transients are provided.

1.2 Cantilever Model

For many applications the cantilever is well modeled as a flexible structure. A multi-mode model accurately captures the cantilever dynamics (see Reference [11]). Typically a first or second mode approximation is enough to describe the dynamics. The state space representation

of the cantilever dynamics is given by,

$$\begin{aligned} \dot{\overline{x}} &= A\overline{x} + B(\eta + w), \\ y &= C\overline{x} + v, \end{aligned} \tag{1.1}$$

where $\overline{x}$ is the state vector, A and B are matrices which are functions of the cantilever parameters, η is the thermal (process) noise component, w describes all other external forces acting on the cantilever, y is the photo-diode output that measures the deflection of the free end of the cantilever and v is the photo-diode (measurement) noise. A first mode approximation of the cantilever dynamics, given by,

$$\ddot{p} + 2\xi\omega_0\dot{p} + \omega_0^2 p = \eta + w, \tag{1.2}$$

can be recast as

$$\begin{aligned} \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} &= \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -2\xi\omega_0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} (\eta + w), \\ y &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + v, \end{aligned} \tag{1.3}$$

where state x_1 denotes the cantilever-tip position (p), state x_2 denotes the cantilever-tip velocity ($\dot{p}$), ω_0 and ξ denote the first resonant frequency of the cantilever and the damping factor in the operating medium, respectively. Note that $\xi = \frac{1}{2Q}$ where Q is the quality factor of the cantilever. We will denote the equivalent dither forcing by g and the tip-sample interaction force by $\phi(p)$ that depends on the tip position p. ϕ typically has characteristics of long range attractive force and short range strong repulsive force [12]. In the above framework, we have $w = g + \phi(p)$ (see Figure 1.3). The cantilever model described above can be identified precisely using the thermal-noise response (see Ref. [11]).

1.3 Observer Architecture

In the systems viewpoint the cantilever dynamics is separated as an independent system from the sample subsystem that affects the cantilever in a feedback manner. Such a perspective was first introduced into the Atomic Force Microscopy literature in [13, 14] where it was utilized to identify the sample interaction potentials. This systems perspective of the AFM (see Figure 1.3) facilitates the design of an observer (see Figure 1.4) that provides an estimate of the state x.

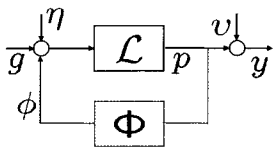


Figure 1.3 In the systems perspective, the AFM dynamics is viewed as an interconnection of a linear cantilever system with the non-linear tip-sample interaction forces in feedback. $\mathcal{L}$ models the cantilever dynamics that is well represented by a linear time invariant system, Φ models the tip-sample interaction force that depends on the tip deflection p. g is the dither forcing, η is the thermal noise, υ is the photo-diode (measurement) noise and y is the measured photo-diode signal.

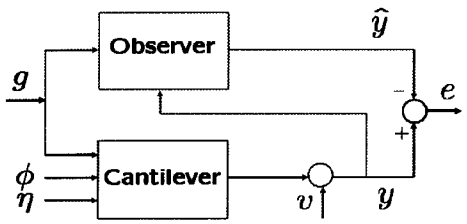


Figure 1.4 The observer estimates the states to be $\hat{\overline{x}}$. The actual state is $\overline{x}$. The error process is denoted by e which is the actual output minus the estimated output. The error process is a zero mean stationary stochastic process in the absence of the tip-sample interaction force ϕ. When the cantilever is subjected to the sample force the cantilever dynamics is altered whereas the observer dynamics remains the same. This is registered as a dynamic profile in the error process e.

The observer dynamics is given by,

$$\overbrace{\begin{aligned}\dot{\hat{\overline{x}}} &= A\hat{\overline{x}} + Bw + L(y - \hat{y}); \hat{\overline{x}}(0) = \hat{\overline{x}}_0, \\ \hat{y} &= C\hat{\overline{x}},\end{aligned}}^{Observer}$$

where $\hat{\overline{x}}$ is the estimate of the state $\overline{x}$. The error in the estimate is given by $\tilde{\overline{x}} = \overline{x} - \hat{\overline{x}}$.

$$\overbrace{\begin{aligned}\dot{\tilde{\overline{x}}} &= A\overline{x} + B(w + \eta) - A\hat{\overline{x}} - Bw - L(y - \hat{y}), \\ &= (A - LC)\tilde{\overline{x}} + B\eta - L\upsilon, \\ \tilde{\overline{x}}(0) &= \overline{x}(0) - \hat{\overline{x}}(0).\end{aligned}}^{State\ Error\ Dynamics}$$

The error in the estimate of the output y is given by,

$$e = y - \hat{y} = C\tilde{\overline{x}} + \upsilon. \tag{1.4}$$

The error process is a good measure of the transients due to changes in tip-sample interactions. Note that only the cantilever-tip position is available as a measured quantity, not its velocity. The error between the observed state and the actual state of the cantilever, when no noise terms are present ($\eta = \upsilon = 0$) is only due to the mismatch in the initial conditions of the observer and the cantilever-tip. It is evident that if the observer gain L is chosen so that the eigenvalues of the matrix $(A - LC)$ are in the strict left half complex plane, the state error $\tilde{x}$ due to the initial condition mismatch $\tilde{x}(0)$ goes to zero with time. The system is observable and therefore the eigenvalues of $(A - LC)$ can be placed anywhere [15]. It can be shown that under the presence of the noise sources η and υ, the error process e approaches a zero mean wide sense stationary stochastic process after the observer has tracked the state of the cantilever.

We now provide the intuition behind the new methodology. When the observer has tracked the cantilever state and the cantilever is subjected to a force due to the sample, the cantilever dynamics get altered. The observer sees the effects of the sample only through the measured cantilever position y which is the output of the photo-diode (see Figure 1.4). This introduces

a lag (that is dependent on the gain L) between the time when the cantilever sees the sample's influence and when the observer realizes the sample's influence. This is registered as a change in the signal e. Thus the error signal e is a good measure of the transient signal. The magnitude of this error signal can be quite large even though the oscillation might not have changed significantly.

When the change in the tip-sample potential persists, the observer by utilizing the input y may eventually track the altered behavior. Thus the error signal e shows the signature at the initial part of the change. But this deviation from stationarity might not persist significantly when the change in the tip-sample behavior persists. This is in contrast to the steady state methods where the information is available not in the initial part but after the cantilever has come to a steady state (for example in tapping-mode scheme this would mean an eventual lowered amplitude value when a step is encountered that persists). Thus the transient signal based scheme is a good *edge detector.*

There is considerable freedom on how fast the observer tracks the cantilever dynamics. Note that high quality factors are detrimental to high bandwidth in steady state methods; however required for high resolution. *By utilizing the observer based architecture presented in this article a method for effectively isolating the high bandwidth needs from the high resolution needs is obtained.*

1.4 Tip-Sample Impact Model

The error profile due to a tip-sample interaction change can be better characterized if a model of the effect of the tip-sample interaction change on the cantilever-motion is available. We assume that the sample's influence on the cantilever tip is approximated by an impact condition where the position and velocity of the cantilever tip instantaneously assume a new value (equivalent to resetting to a different initial condition). This is satisfied in most typical operations because in the dynamic mode, the time spent by the tip under the sample's influence is negligible compared to the time it spends outside the sample's influence [16]. The assumption is also corroborated by experimental results provided later.

1.5 Bandwidth and Noise Analysis

For bandwidth and noise analysis the first mode model for the cantilever, given by Equation (1.3), is assumed. The dynamics of the signal e which is the difference between the photo-diode signal y and its estimated value ($\hat{y}$) is given by,

$$e(s) = \frac{\eta(s) + (s^2 + \frac{\omega_0}{Q}s + \omega_0^2)\upsilon(s) + (s + \frac{\omega_0}{Q})\nu_1 + \nu_2}{s^2 + (\frac{\omega_0}{Q} + l_1)s + (\omega_0^2 + l_2 + \frac{\omega_0}{Q}l_1)}. \tag{1.5}$$

where $[\nu_1,\ \nu_2]^T$ is the initial condition reset due to change in tip-sample interaction and $L = [l_1\ l_2]^T$ is the gain of the observer that must satisfy the stability criterion: $(\frac{\omega_0}{Q} + l_1) > 0$ and $(\omega_0^2 + \frac{\omega_0}{Q}l_1 + l_2) > 0$. The signal e contains the dynamic profile due to tip-sample interaction change $[\nu_1, \nu_2]^T$ which is corrupted by both thermal noise η and photo-diode noise υ.

From Equation (1.5) it can be seen that the transfer function from the unknown state jump $[\nu_1, \nu_2]^T$ to the signal e is a second order transfer function. Hence the tracking bandwidth is characterized by,

$$B \propto \frac{\omega_0}{Q} + l_1. \tag{1.6}$$

Since the choice of the gain term l_1 is independent of the quality factor Q, the tracking bandwidth of the observer is effectively decoupled from Q.

The mean square contribution to e of the instantaneous state jump is given by,

$$\begin{aligned} < p_\nu^2 > &= \frac{1}{\pi}\int_0^\infty \frac{(\frac{\omega_0}{Q}\nu_1 + \nu_2)^2 + \omega^2\nu_1^2}{(\omega^2 - \omega_0^2 - l_2 - \frac{\omega_0 l_1}{Q})^2 + \omega^2(\frac{\omega_0}{Q} + l_1)^2} d\omega, \\ &= \frac{(\omega_0^2 + \frac{\omega_0 l_1}{Q} + l_2)\nu_1^2 + (\frac{\omega_0}{Q}\nu_1 + \nu_2)^2}{2(\omega_0^2 + l_2 + \frac{\omega_0 l_1}{Q})(\frac{\omega_0}{Q} + l_1)}. \end{aligned} \tag{1.7}$$

$< p_\nu^2 >$ characterizes the signal power.

The thermal noise acting on the cantilever is white and has mean equal to zero. Assuming the thermal noise power is equal to P, the mean square contribution to e of the thermal noise

is given by,

$$\begin{aligned} < p_\eta^2 > &= \frac{P}{\pi}\int_0^\infty \frac{1}{(\omega^2 - \omega_0^2 - l_2 - \frac{\omega_0 l_1}{Q})^2 + \omega^2(\frac{\omega_0}{Q} + l_1)^2} d\omega, \\ &= \frac{P}{2(\omega_0^2 + l_2 + \frac{\omega_0 l_1}{Q})(\frac{\omega_0}{Q} + l_1)}. \end{aligned} \tag{1.8}$$

Thus the signal to noise ratio due to thermal noise per unity noise power is given by,

$$SNR_\eta = \sqrt{\frac{< p_v^2 >}{< p_\eta^2 >}} = \sqrt{(\omega_0^2 + \frac{\omega_0 l_1}{Q} + l_2)\nu_1^2 + (\frac{\omega_0}{Q}\nu_1 + \nu_2)^2}. \tag{1.9}$$

It can be seen from Equation(1.6) and (1.9) that with increasing values of l_1 and l_2, SNR_η and bandwidth B increase.

The measurement noise in the photo-diode is assumed to be white and has mean equal to zero. Assuming the photo-diode noise power is equal to R, the power spectral density (psd) of the photo-diode noise v in signal e is given by,

$$P_{vv}(\omega) = R\left[1 - \frac{(2\omega_0^2 + l_2 + \frac{\omega_0 l_1}{Q})(\frac{\omega_0 l_1}{Q} + l_2) + \omega^2(l_1^2 - l_2)}{(\omega^2 - \omega_0^2 - l_2 - \frac{\omega_0 l_1}{Q})^2 + \omega^2(\frac{\omega_0}{Q} + l_1)^2}\right] = R[1 - P_{vv}^{lp}(\omega)].$$

It can be assumed that the signal power due to state jump $[\nu_1, \nu_2]^T$ in e is contained within a bandwidth $B = K(\frac{\omega_0}{Q} + l_1)$ for $K \ggg 0$. Note that the thermal noise contribution in e is also contained within bandwidth B . The signal e can be filtered out for $\omega > B$ and the signal-to-noise-ratio in e (SNR_v) due to measurement noise can be analyzed. Note that the noise power of the low pass component $P_{vv}^{lp}(\omega)$ of psd of photo-diode noise $P_{vv}(\omega)$ is contained within bandwidth B. The mean square contribution of the photo-diode noise to e (where e is

filtered and effectively contains the signal of interest) is approximately given by,

$$\begin{aligned} < p_v^2 > &\approx \frac{1}{\pi}\int_0^B P_{vv}(\omega)d\omega, \\ &\approx R\left[K(\frac{\omega_0}{Q}+l_1) - \frac{1}{\pi}\int_0^\infty \frac{(2\omega_0^2 + l_2 + \frac{\omega_0 l_1}{Q})(\frac{\omega_0 l_1}{Q}+l_2) + \omega^2(l_1^2 - l_2)}{(\omega^2 - \omega_0^2 - l_2 - \frac{\omega_0 l_1}{Q})^2 + \omega^2(\frac{\omega_0}{Q}+l_1)^2} d\omega\right], \\ &= R\left[K(\frac{\omega_0}{Q}+l_1) - \frac{(\omega_0^2 + l_2 + \frac{\omega_0 l_1}{Q})(\omega_0^2 + \frac{\omega_0 l_1}{Q} + l_1^2) - \omega_0^4}{2(\omega_0^2 + l_2 + \frac{\omega_0 l_1}{Q})(\frac{\omega_0}{Q}+l_1)}\right]. \end{aligned} \tag{1.10}$$

The signal to noise ratio due to measurement noise per unity noise power is given by,

$$SNR_v = \sqrt{\frac{< p_\nu^2 >}{< p_v^2 >}} \approx \sqrt{\frac{(\omega_0^2 + l_2 + \frac{\omega_0 l_1}{Q})\nu_1^2 + (\frac{\omega_0}{Q}\nu_1 + \nu_2)^2}{(\omega_0^2 + l_2 + \frac{\omega_0 l_1}{Q})\left((4K-1)\frac{\omega_0 l_1}{Q} + (2K-1)l_1^2 + 2K(\frac{\omega_0}{Q})^2 - \omega_0^2\right) + \omega_0^4}}. \tag{1.11}$$

It can be seen that SNR_v decreases with increasing values of l_1 and l_2. Therefore the bandwidth constraint in the scheme to detection the signal due to state jump $[\nu_1, \nu_2]^T$ is mainly imposed by the measurement noise. It is evident that a desired tradeoff between signal to noise ratio and bandwidth can be obtained by an appropriate choice of l_1 and l_2 that is independent of Q. This provides considerable flexibility when compared to existing steady state methods where Q determines the bandwidth. Note that due to the small measurement noise, the observer gain l_1 can be chosen large enough so that the cantilever state is tracked within a couple of cycles of the dither forcing. Therefore the optimal bandwidth is primarily dictated by the resonant frequency ω_0 of the cantilever.

1.6 Detection Algorithm

With the added assumption of the impact condition, the sample detection problem is formulated by considering a discretized model of the cantilever dynamics given by (1.1).

$$\begin{aligned} x_{k+1} &= Fx_k + G(g_k + \eta_k) + \delta_{\theta,k+1}\nu, \\ y_k &= Hx_k + v_k,\ k \geq 0, \end{aligned} \tag{1.12}$$

where matrices F, G, and H are obtained from matrices A, B and C by discretizing the continuous time model of the cantilever described by Equation(1.1) and $\delta_{i,j}$ denotes the dirac delta function defined as $\delta_{i,j} = 1$ if $i = j$ and $\delta_{i,j} = 0$ if $i \neq j$. θ denotes the time instant when the impact occurs and ν signifies the magnitude of the impact. Essentially the impact is modeled as an instantaneous change in the state by ν at time instant θ. In this setting the time of impact and the resulting change in the state are unknown quantities. The profile of the change in the mean of the error signal due to the sample can be pre-calculated and one can then employ detection and estimation methods to search for the presence of such a profile in the error sequence to not only detect the samples presence but also estimate the sample parameters.

We assume that the input noise and the output noise are white and uncorrelated. Note that the input and output noise power P and R can be measured experimentally. Given the following noise characteristics

$$E\left\{\begin{bmatrix}\eta_i\\ v_i\\ x_0\end{bmatrix}\begin{bmatrix}\eta_j\\ v_j\\ x_0\\ 1\end{bmatrix}^T\right\} = \begin{bmatrix}P\delta_{ij} & 0 & 0 & 0\\ 0 & R\delta_{ij} & 0 & 0\\ 0 & 0 & \Pi_0 & 0\end{bmatrix}$$

the optimal transient observer is a Kalman Filter [17]. Let the steady state Kalman observer gain be given by $L = L_K$. The error process is known as the innovation sequence when the optimal transient observer is employed. Moreover when the tip-sample interaction is absent, the innovation process asymptotically approaches a zero mean white process.

With a Kalman observer and an impact model when the sample is encountered (given by the model in (1.12)), the innovation sequence e_k can be written as [18],

$$e_k = y_k - \hat{y}_k \ = \Upsilon_{k;\theta}\ \nu + \gamma_k \ , \tag{1.13}$$

where $\{\Upsilon_{k;\theta}\ \nu\}$ is a known dynamic state profile with unknown arrival time θ defined by

$$\Upsilon_{k;\theta} = H(F - L_K H)^{k-\theta} \tag{1.14}$$

and $\{\gamma_k\}$ is a zero mean white noise sequence which is the measurement residual had the jump not occurred. The statistics of γ is given by,

$$E\{\gamma_j \gamma_k^T\} = V\delta_{ij}$$

where $V = HP_{\tilde{x}}H^T + R$ and $P_{\tilde{x}}$ is the steady state error covariance obtained from the Kalman filter. V is a function of P and R.

Thus determining when the cantilever is "hitting" the sample and when it is not, is equivalent to deciding whether the dynamic profile is present or not. A number of detection and estimation methods exist in statistical signal processing literature which address problems of similar nature. This problem can be formulated in the framework of binary hypothesis testing and a suboptimal version of the generalized likelihood ratio test [18] can be used to make the decision whether the profile is present or not. The innovation process is windowed into M samples for the analysis. The selection of the window size M is primarily determined by the length of the dynamic profile. This problem formulated in the framework of binary hypothesis testing is given by,

$$\begin{aligned} H_0 &: e_k = \gamma_k,\ k = 1, 2, \ldots, M \\ versus & \\ H_1 &: e_k = \Upsilon_{k;\theta}\ \nu + \gamma_k,\ k = 1, 2, \ldots, M \end{aligned} \tag{1.15}$$

where γ_k is a zero mean white gaussian process, $e_k = \gamma_k$ is the observed innovation and $\{\Upsilon_{k;\theta}\ \nu\}$ for $k = 1, 2, \ldots, M$ is a known *dynamic profile* with unknown arrival time θ and unknown magnitude of the state jump ν. Note that if hypothesis H_0 is true, then the cantilever has not interacted with the sample within that window of M data points. Selection of hypothesis

H_1 implies that the decision made is that the oscillating cantilever interacted with the sample producing the dynamic profile. To simplify the analysis, it is assumed that the impact or state jump occurs at the first sample of each time window of M samples (i.e. $\theta = 1$). This implies,

$$\overline{e} = \Upsilon\nu + \overline{\gamma} \tag{1.16}$$

where, $\overline{e} = [e_1, e_2, \cdots, e_M]^T$, $\Upsilon = \left[H, H(F - L_K H), H(F - L_K H)^2, \cdots, H(F - L_K H)^{M-1}\right]^T$ and $\overline{\gamma} = [\gamma_1, \gamma_2, \cdots, \gamma_M]^T$. Note that when the observer gain L is not equal to the Kalman observer gain L_K, the dynamic profile assumes a similar expression as in Equation(1.16) with observer gain L replacing L_K. However in that case the noise $\overline{\gamma}$ will not necessarily be a white noise. When $\overline{\gamma}$ is not white and is a known colored noise, a whitening filter may be introduced and the corresponding dynamic profile $\Upsilon^w\nu$ output from the whitening filter may be detected in corresponding white noise. When $\overline{\gamma}$ is white, the maximum likelihood estimate of the state jump, $\hat{\nu}$ is given by,

$$\hat{\nu} = (\frac{\Upsilon^T\Upsilon}{V})^{-1}\frac{\Upsilon^T\overline{e}}{V}. \tag{1.17}$$

The dynamic profile if present in the selected M samples is given by

$$\overline{s} = \Upsilon\hat{\nu} = \Upsilon(\frac{\Upsilon^T\Upsilon}{V})^{-1}\frac{\Upsilon^T\overline{e}}{V}.$$

To select between the two hypotheses, the likelihood ratio is computed which is given by,

$$l(M) = \frac{\overline{e}^T\overline{s}^T}{V} = \frac{\overline{e}^T\Upsilon}{V}(\frac{\Upsilon^T\Upsilon}{V})^{-1}\frac{\Upsilon^T\overline{e}}{V}. \tag{1.18}$$

The likelihood ratio is compared with a threshold value as $l(M) \gtrless_{H_0}^{H_1} \epsilon$ to arrive at a decision whether the dynamic profile is present or not (equivalent to deciding whether a tip-sample interaction has occurred or not). The threshold ϵ is chosen to provide a suitable tradeoff between the rate of false detection and the rate of missed detection [19]. The false alarm P_F and detection probability P_D are calculated as, $P_F = P_0(\Upsilon) = \int_\epsilon^\infty p(l = L|H_0)dL$ and $P_D(\nu) = P_1(\Upsilon) = \int_\epsilon^\infty p(l = L|H_1, \nu)dL$ respectively. $P_0(\Upsilon)$ and $P_0(\Upsilon)$ are the probability of detecting

the dynamic profile Υ under hypothesis H_0 and H_1 respectively. $p(l = L|H_0)$ and $p(l = L|H_1, \nu)$ are the probability density functions of likelihood ratio l under hypothesis H_0 and H_1 with state jump ν respectively. Since $l = -\frac{1}{2}\sum_{i=1}^{M} e_i^T V^{-1} e_i$ under hypothesis H_0 and e_i are independent identically distributed (iid) gaussian random variables, $p(l = L|H_0)$ is Chi-squared (χ^2) density with M degrees of freedom. Similarly $l = -\frac{1}{2}\sum_{i=1}^{M} e_i^T V^{-1} e_i - \frac{1}{2}\nu \sum_{i=1}^{M} \Upsilon_i^T V^{-1} e_i$ under hypothesis H_1, $p(l = L|H_1)$ is a non-central χ^2 density with noncentrality parameter $\nu^T(\frac{\Upsilon^T \Upsilon}{V})\nu$. Therefore P_D is dependent upon values of ν. For specified P_F or P_D, the threshold value ϵ can be computed from the tables in [20]. Also given ϵ, the values P_F or $P_D(\nu)$ can be computed similarly. To compute P_D we use ν as the minimum jump that is required to be detected.

Since the dynamic profile Υ is a function of observer gain L, there is a possibility of further optimizing Equation(1.18) by making it a function of observer gain L.

1.7 Experimental Results

Experiments were performed to ascertain the efficacy of the new transient signal based approach.

1.7.1 High Speed Detection of Sample

A *Digital Instruments* multi-mode AFM was used in the experiments. The piezo dynamics was used to provide *data bits* like peaks in sample profile. The frequency response of the piezo was obtained using an HP 3563A control system analyzer and a model was fit to the response. The model response is compared with that obtained experimentally in Figure 1.5. Figure 1.6 shows the experimental response of the piezo to a voltage pulse of amplitude 0.5V, time period 1000 μs and duty ratio 50 %. The piezo dynamics results in the occurrence of 4 peaks separated by approximately 100 μs during the on time of the applied voltage pulse. The maximum width and height of each peak is approximately 35 μs and 20 nm, respectively.

Frequency response of cantilever was obtained using HP 3563A control system analyzer. A 2^{nd} order (1^{st} mode) model of the cantilever was obtained from the frequency response data.

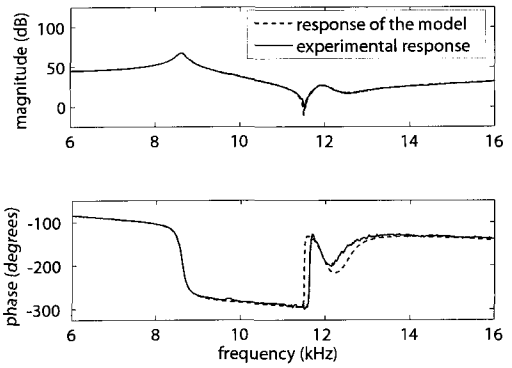


Figure 1.5 Experimental frequency response of the z-piezo and simulated frequency response of a 11^{th} order z-piezo model are plotted.

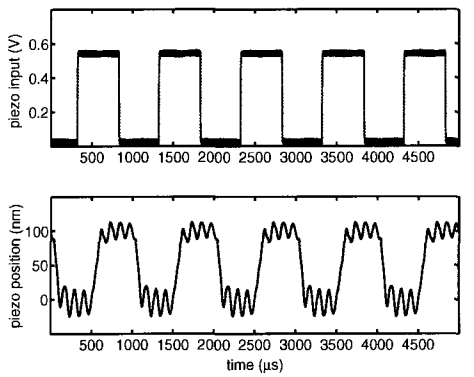


Figure 1.6 An input voltage pulse of amplitude 0.5V, period 1000 μs and on time 500 μs results in the above piezo response. The four peaks during the on time are separated by approximately 100 μs.

Experimental and simulated frequency response of cantilever are plotted in Figure 1.7. The resonant frequency $f_0 = 60.25$ kHz, quality factor $Q = 104$, dc gain $K = 1.41$ and phase offset $\phi_{off} = -147.7$ degrees were obtained from frequency response data. The right half plane zero in the transfer function that fits the frequency response data can be attributed to inertia of other mechanical components in AFM. This is the first time experimentally obtained transfer function of AFM is used for real-time application.

A Kalman filter was designed from the 2^{nd} order model of the cantilever and experimentally obtained thermal noise and measurement noise power. The Kalman filter is implemented as an

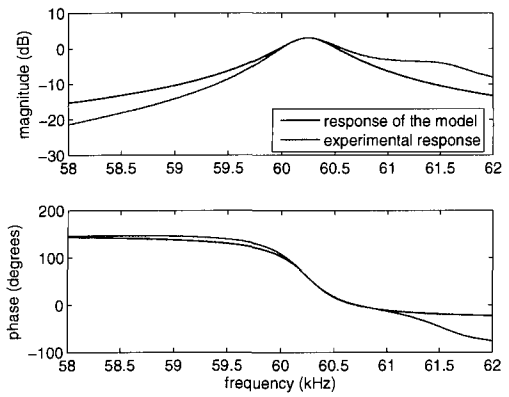


Figure 1.7 Experimental frequency response of the cantilever and simulated frequency response of a 1^{st} mode (2^{nd} order) model of cantilever are plotted.

Analog filter circuit as shown in Figure 1.8. The cantilever deflection signal y and the dither excitation signal $g = w$ are inputs to the circuit and the estimated cantilever deflection signal $\hat{y}$ is output from the circuit. The cantilever deflection y, forcing w and estimated deflection $\hat{y}$ signals are measured simultaneously at a sampling frequency of 5 MHz using a digital signal processing board (Conejo from Innovative Integration).

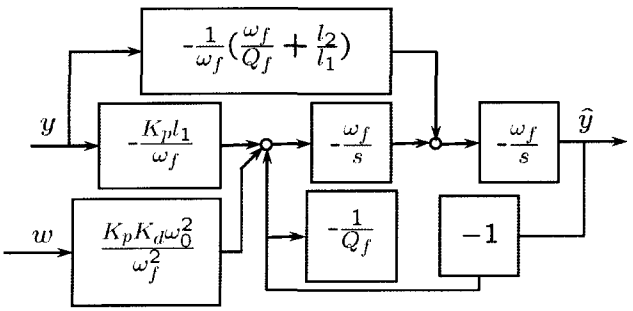


Figure 1.8 The Kalman filter is implemented as shown in the block diagram. It is designed based on the first mode model of the cantilever. y, f and $\hat{y}$ are the cantilever deflection, dither forcing and estimated deflection signals. K_d is the amplifier gain at the dither-piezo, K_p is the photo-diode sensitivity. ω_f, Q_f and l_1, l_2 are the natural frequency and quality factor and gains of Kalman filter.

A graphite (HOPG) sample was placed over the xyz-piezo scanner and voltage pulse of 0.5 V

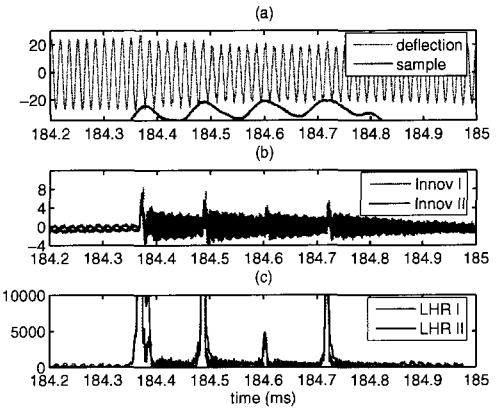


Figure 1.9 (a) Cantilever position given by deflection signal is plotted above sample position. (b) Innovation I from analog Kalman filter and innovation II from 4^{th} order Kalman filer is plotted. Every time cantilever interacted with the sample the innovation signals lost their zero mean nature. (c) The likelihood ratios LHR I and II calculated from innovation sequences I and II are plotted. Likelihood ratio shot up when the cantilever hit the sample. It fell to very low values even when the second mode oscillations persisted in the innovation signal.

amplitude and 1 kHz frequency was applied to z-scanner. Then a cantilever freely oscillating in air was brought close to the sample while monitoring innovation signal. The innovation signal e was obtained by subtracting deflection signal y from estimated deflection signal $\hat{y}$ using an analog adder circuit implemented along with analog Kalman filter. It was observed that when the hits between cantilever and sample occured, innovation signal lost its zero mean nature. Dither excitation, deflection and innovation signal is collected at 5 MHz sampling. A second mode model (4^{th} order) of the cantilever was obtained experimentally. Another 4^{th} order Kalman filter based on this model was designed and innovation sequence was obtained from dither excitation and deflection signal. Innovation signal from analog Kalman filter (Innov I) and innovation sequence from 4^{th} order Kalman filter (Innov II) are plotted in Figure 1.9(b). Cantilever position given by deflection signal is plotted above sample position in Figure 1.9(a). It can be observed that when the cantilever hits the peaks in the sample profile, the innovation signal loses its zero mean nature. Innovation signal from 4^{th} order Kalman filter captures the

dynamic profile due to the first mode response of the cantilever very well. The analog Kalman filter does not have 2^{nd} mode model information of the cantilever. Therefore 2^{nd} mode response of the cantilever shows up in the innovation signal from analog Kalman filter. Likelihood ration test was carried out on the innovation signals based on the dynamic profiles of the 1^{st} model of the cantilever. The likelihood ratios LHR I and II corresponding to innovation sequences I and II are plotted in Figure 1.9(c). Innovation sequence I contains the 1^{st} model dynamic profiles and 2^{nd} mode response of the cantilever. Innovation sequence II contains the 1^{st} and 2^{nd} mode dynamic profiles. However, the likelihood ratios based on 1^{st} mode dynamic profiles only match well to one another in terms of lateral width and vertical height. Therefore likelihood ratio test based on 1^{st} mode dynamic profiles is sufficient to detect sample.

In this experiment the cantilever encountered peaks in the sample profile time separated by 100 μseconds. This translates to a spatial bandwidth of 10 kHz. In other words the experiment demonstrates a sample detection bandwidth of 10,000 bits per second. With suitable choice of observer and likelihood ratio test, the sample can be detected within 4 cycles of the cantilever oscillation which implies an imperial bandwidth of $\frac{f_0}{4}$. Therefore it is possible detect at 16.25 kHz using a 70 kHz cantilever.

1.7.2 Fast Imaging of DNA

In another experiment TF-AFM based imaging was demonstrated by imaging DNA sample. Lambda DNA (Catalog♯ D1501 from Promega Corporation) of concentration 500μg/ml was diluted to 50μg/ml using Tris/HCl/EDTA buffer (10mM Tris/HCl, 1mM EDTA, pH 6.6-6.8). The DNA solution was further diluted to 1μg/ml using $NiCl_2$ buffer (40mM HEPES, 5mM $NiCl_2$, pH 6.6-6.8). 20μl of DNA solution was incubated on a freshly cleaved mica surface for 10minutes. Mica surface was rinsed with pure water and blown dry in nitrogen. For this experiment Kalman observer was implemented on a FPGA (from Xilinx) and innovation signal from FPGA board was fed to a rms-to-dc circuit board (AD8361 evaluation board from Analog Devices). Output of rms-to-dc circuit was connected to auxiliary input of atomic force microscope (Multimode AFM from Veeco Instruments) for imaging in AFM graphic user

interface (Nanoscope v.5).

TF-AFM imaging operates in parallel with AM-AFM (amplitude modulated) imaging (see Figure 1.10). It requires no existing hardware modification except for an electronic add-on.

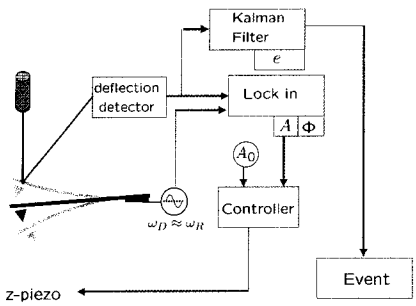


Figure 1.10 The excitation frequency ω_D is set close to the resonant frequency ω_R of the cantilever. The amplitude A of the deflection signal of the cantilever is controlled at a set point amplitude of A_0 by actuating the z-piezo in a feedback manner. The innovation signal from the Kalman filter is analyzed by the transient detection method to infer variations in the sample profile. A_0 is kept close to the free oscillation amplitude so that the tip-sample interaction is mild.

AM-AFM images of DNA are shown in Figure 1.11. The scan size = 2 μm, scan rate = 2 Hz and DNA height = 1.5 nm. Typical scan rate for DNA imaging is 1 Hz. The amplitude and phase image of DNA are shown in Figure 1.12. The height image at 2 Hz of DNA is taken as reference to compare with TF-AFM images.

For TF-AFM imaging of DNA, the scan parameters were optimized so that the cantilever is soft on the sample. Set point amplitude of cantilever was kept close to free oscillation amplitude. The set point amplitude was slowly reduced from its free oscillation value until image appears. Z-piezo feedback gains were optimized for TF-AFM imaging (integral gain $k_i = 0.5$ and proportional gain $k_p = 1$). The scan speed was slowly increased in small steps and TF-AFM and AM-AFM images were collected. DNA height, phase, amplitude and TF-AFM images collected at 12.21 Hz are plotted in Figure 1.13. Due to small closed loop bandwidth of AFM instrument height image of DNA at 12.21 Hz is blurred (see Figure 1.13(a)). Due to high

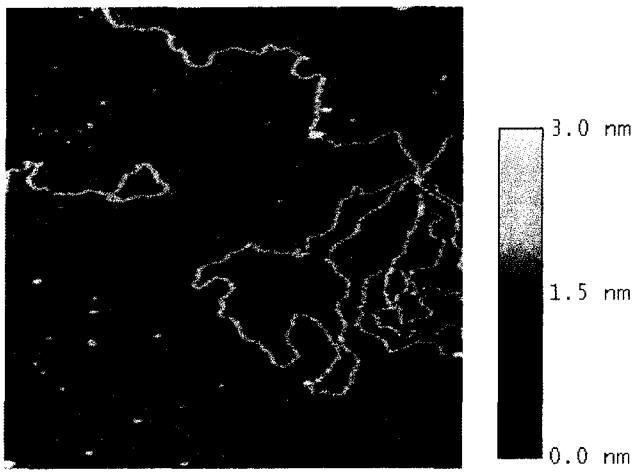
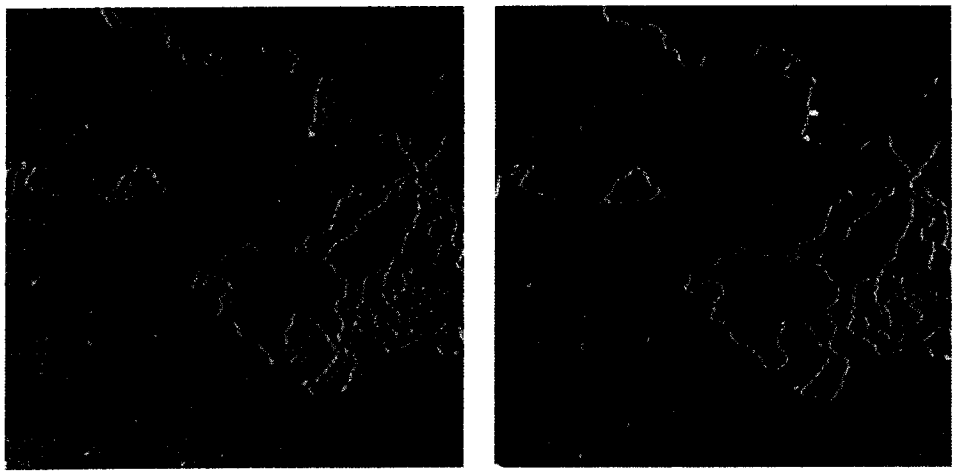


Figure 1.11 AM-AFM height image of Lambda DNA. Scan size = 2μm, scan rate = 2Hz and height = 1.5 nm. This image is taken as reference to compare with TF-AFM images.

quality factor of cantilever phase and amplitude images are also blurred (see Figure 1.13(b) and (c)). However, TF-AFM image provides an accurate image of DNA even at 12.21 Hz (compare Figure 1.11 and Figure 1.13(d)). Even 12.21 Hz, the small specs present on the sample are distinctively captured in TF-AFM image than compared to height, phase and amplitude images. Scan rate was further increased to 20.35 Hz and the corresponding height, phase, amplitude and TF-AFM images are plotted in Figure 1.14. The quality of AM-AFM images i.e. height, phase and amplitude images further detoriates. However, TF-AFM image still provides a good image of DNA (compare Figure 1.11 and Figure 1.14(d)). The vertical streaks in the images in Figure 1.14 are due to oscillations in the xyz-scanner. The xyz-scanner is not tuned (look-up table) for large scan speeds. When the fast-axis scanner takes sharp turn at the edge of the scan area, it rings which is coupled to z-piezo. These vibrations could be eliminated by tuning the AFM system for high speed scanning.



(a) (b)

Figure 1.12 AM-AFM (a) amplitude and (b) phase images of Lambda DNA. Scan size = 2μm, scan rate = 2Hz and height = 1.5 nm.

1.8 Conclusion

Note that all present schemes employ steady-state data. In this article a framework for ultra-fast interrogation of sample in Atomic Force Microscopy is proposed which utilizes the tip-deflection data during the transient state of the cantilever probe. The systems perspective has facilitated the development of this methodology. The cantilever and its interaction with the sample is modeled. A first and second mode approximation model of the cantilever is considered and a Kalman filter is designed to estimate the dynamic states. The tip-sample interaction is modeled as an impulsive force applied to the cantilever in order to detect the presence of sample. The dynamics due to tip-sample interaction is calculated in the innovation sequence and a likelihood ratio test is performed to obtain the decision rule to infer the presence of sample. Experimental results tally with the simulation results verifying the proposed methodology and the sample-detection scheme.

Simulations show a bandwidth of $f_0/4$ (f_0 being the natural frequency of the micro-cantilever) in detecting small time scale tip-sample interactions. The result is corroborated in

experiments and sample profiles appearing at $10KHz$ were detected using a cantilever with resonant frequency equal to $70KHz$ in realtime by implementing analog observer. In this method high quality factor does facilitate high resolution but it does not limit the bandwidth as in steady state data based methods.

The transient signal based sample detection scheme is integrated with regular tapping mode operation of AFM to detect high bandwidth content sample features at large scan rates in the case where the tapping mode images failed to produce useful image. As part of ongoing work experiments to ascertain sub-nanometer scale features like atomic layers of graphite and the lattice structure of mica are being conducted at a scan rate an order higher than the regular tapping mode operation with encouraging preliminary results.

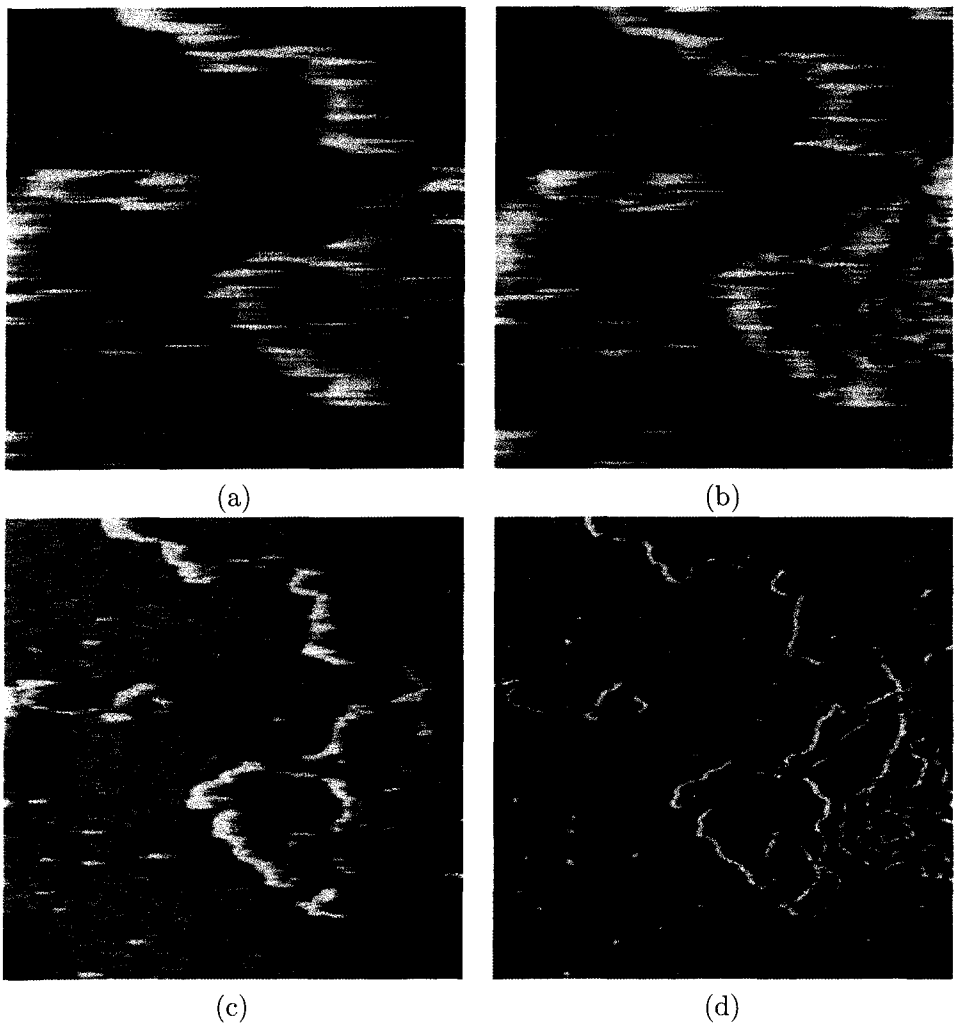


Figure 1.13 AM-AFM (a) height, (b) phase, (c) amplitude and (d) TF-AFM images of Lambda DNA at 12.21 Hz. Scan size $= 2\mu$m and height = 1.5 nm.

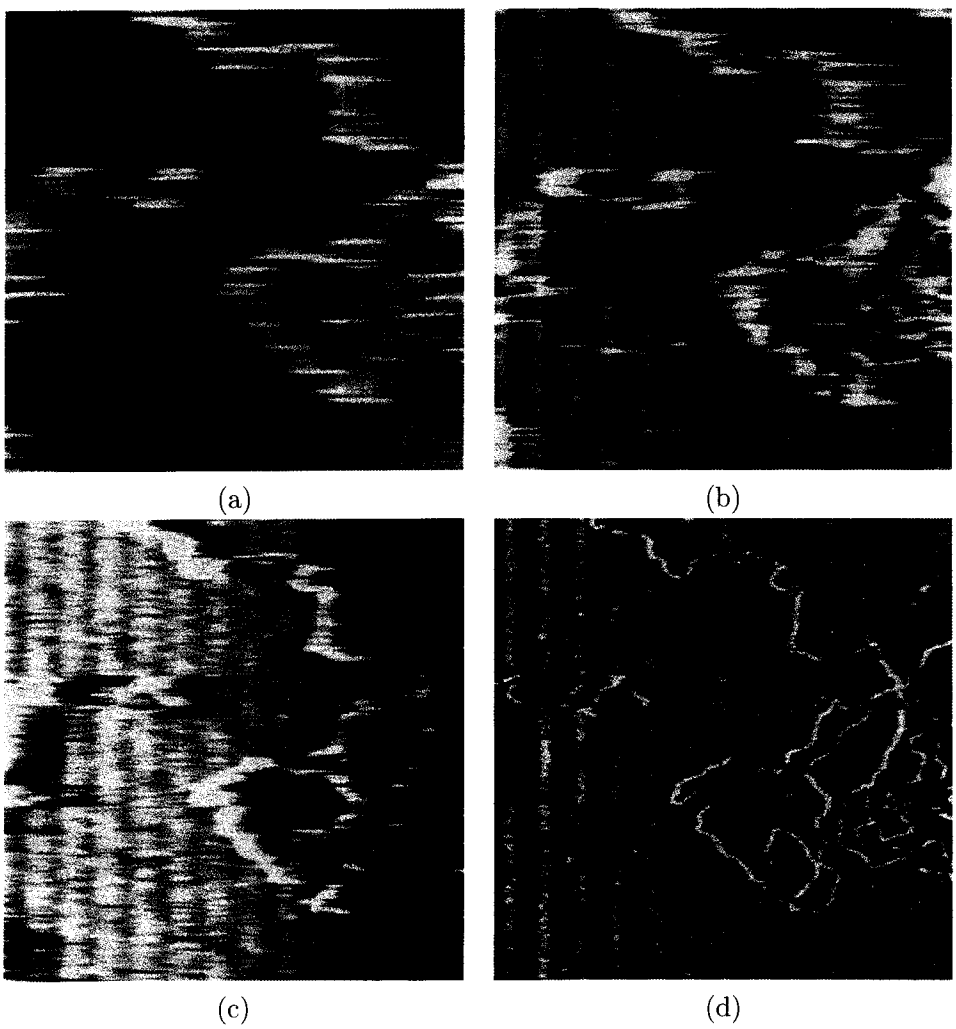


Figure 1.14 AM-AFM (a) height, (b) phase, (c) amplitude and (d) TF-AFM images of Lambda DNA at 20.35 Hz. Scan size $= 2\mu$m and height $= 1.5$ nm.

CHAPTER 2 Dual Quality Factor (Q) Control

In atomic force microscopy, bandwidth or resolution can be affected by active quality factor (Q) control. However, in existing methods the trade off between resolution and bandwidth remains inherent. Observer based Q control method provides greater flexibility in managing the tradeoff between resolution and bandwidth during imaging. It also facilitates theoretical analysis lacking in existing methods.

Steady state signals like amplitude and phase are slowly varying variables that are suited to image low bandwidth content of the actual sample profile. Observer based transient imaging scheme with Q control has the promise of detecting high bandwidth content of the sample features during scanning. Transient detection also has the advantage of high sensitivity to small features.

2.1 Introduction

There is considerable interest in interrogation and manipulation of surface properties of inorganic and biological materials at molecular level using atomic force microscope (AFM) [1]. AFM is often operated in the dynamic mode (i.e. amplitude modulation and frequency modulation) to image with low lateral force and high force sensitivity. A schematic of AFM operating in dynamic mode is shown in Fig. 2.1. In amplitude modulation mode (AM-AFM) operation a sinusoidal excitation is applied by the dither piezo at the base of the cantilever. Frequency of excitation signal is set very close to the resonant frequency of cantilever. The cantilever tip deflection is measured by a photo diode sensor. When tip of the oscillating cantilever interacts with the sample the spring constant and damping coefficient of cantilever get altered. Therefore the amplitude (phase and frequency) of deflection signal changes. In

AM-AFM a set point amplitude of deflection signal is maintained by feeding back the amplitude signal to vertical sample positioning system (z-piezo). The sample is raster scanned by the xy-scanner. Images of the sample are obtained from the control signal to z-piezo, and the amplitude and phase of deflection signal.

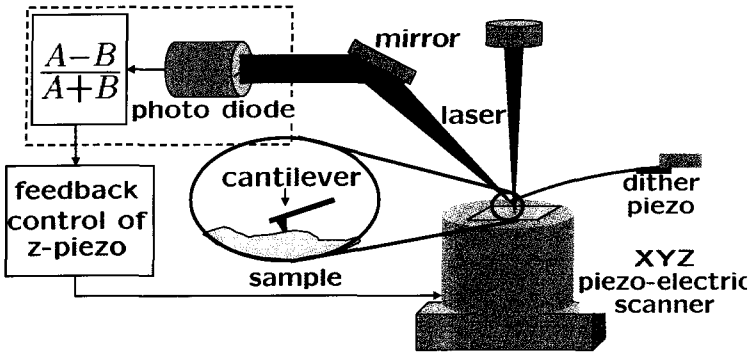


Figure 2.1 A schematic of AFM.

The scan speed in dynamic mode AFM is dictated by the mechanical bandwidths of the xyz scanner and cantilever. Given a xyz scanner, imaging resolution is better if cantilever with high quality factor is used. However, high quality factor results in large settling times for the amplitude and phase signals. Thus imaging speed of amplitude (or phase) based methods are compromised by high quality factor of cantilever.

Recently several methods have been proposed to increase the bandwidth or the resolution of imaging in dynamic mode atomic force microscopy. In [21], scan speed while imaging in air is increased by using a z-piezo with high bandwidth and active reduction of quality factor of cantilever. In [22] the force sensitivity of cantilever (resolution of imaging) in fluid is enhanced by 3 orders of magnitude by active Q-enhancement. However, given a xyz scanner and a cantilever, the bandwidth and resolution of imaging in dynamic methods is dictated by quality factor of cantilever.

Imaging under fluids with Q-enhancement has numerous advantages. While imaging in air, molecular forces (in piconewton range) are not accessible due to capillary forces (in nanonewton range) from the moisture layer covering the sample. Moisture layer problem is absent when sample is under fluid. Atomic resolution images have been obtained by imaging under fluid

[23]. Also biological samples are soft and they often need to be imaged in a buffer solution. The low lateral forces in fluids facilitates imaging of biological samples as they are not displaced or destroyed [24], [25]. However, in fluids quality factor and hence force sensitivity of cantilever reduces by approximately two orders of magnitude compared to its value in air. By active Q-control quality factor of cantilever and consequently force sensitivity can be enhanced in order to sense molecular level forces. In [22] quality factor of a cantilever is enhanced by three orders of magnitude and force sensitivity is improved to piconewton regime.

In existing Q-control methods, the deflection signal is phase shifted (or time delayed), amplified and added to standard excitation signal applied to dither piezo. However, in these methods the trade off between bandwidth and resolution remains inherent [26]. It is always desirable to improve both bandwidth and resolution together. In current methods when the feedback loop is set to achieve a particular value of Q, the resolution and bandwidth get fixed according to the effective value of Q. Observer based Q-control method provides flexibility in the feedback loop. It improves resolution in the case of active damping and improves bandwidth in the case of active Q-enhancement.

Another advantage of observer based approach is its amenability to the transient detection technique. Note that imaging signals like the amplitude and phase being slow varying, are inadequate to construct the high bandwidth (fine) features of sample surface. Based on these methods, at higher scan rates or scan sizes the sample features appear at higher temporal frequency to the cantilever and the image may not portray the fine features on the sample. On the other hand, transient detection method in [9] is able to detect tip-sample interactions at a very high bandwidth (not limited by high Q). Transient detection method adapted to sample imaging with observer based Q control can detect high bandwidth content of the sample at high scan rates that is not possible with amplitude and phase based imaging schemes.

2.2 AFM model

From a system perspective, the input signal to AFM is the excitation signal applied to the dither-piezo and the output signal is the deflection of the cantilever-tip obtained from the

photo-diode. The frequency of the excitation signal is set close to the first resonant frequency of the cantilever. It is observed from experiments that near first resonant frequency, the transfer function from dither-piezo-input to photo-diode-output is described by a second order model with a right half plane zero, given by:

$$G(s) = \frac{k/m(c_1 + c_2 s)}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2}, \tag{2.1}$$

where k, m, ω_0 and Q are equivalent stiffness, equivalent mass, resonance frequency and quality factor of cantilever. Note that $k/m = \omega_0^2$. By putting $c_1 = 1$ and $c_2 = 0$ in Equation(2.1), transfer function corresponding to the point mass description of cantilever is recovered.

In AFM setup, as described in Figure 2.2, thermal noise η enters as a process noise, photo-diode noise appears at the output as a measurement noise and tip-sample force appears at the input of the cantilever. A state-space representation of AFM is given by,

$$\begin{aligned} \dot{x} &= Ax + Bu + B_1 h + B_1 \eta;\ x(0) = x_0, \\ y &= Cx + v, \end{aligned} \tag{2.2}$$

where state matrices A, B, B_1 and C are realized from the frequency response from dither-piezo-input to photo-diode-output. x is the dynamic state of the cantilever, u is the dither-piezo-input and y is the photo-diode-output. The canonical realization of the transfer function (fitted to the frequency response of AFM) in Equation(2.1) gives $A = [0, 1; -\omega_0^2, -\omega_0/Q]$, $B = [0; k/m]$, $B_1 = [0; 1/m]$ and $C = [c_1,\ c_2]$. Cantilever-tip position $p = Cx$ and velocity $v = [0,\ 1]x$.

Tip-sample interaction force h is a function of tip-position p and tip-velocity $\dot{p}$, i.e. $h = \Phi(p, \dot{p})$. The function Φ is a nonlinear function of p and depends on sample. Thus in presence of sample the interconnection of the cantilever and sample is described by:

$$\begin{aligned} p &= \Phi_{hp} h + \Phi_{gp} g, \\ h &= \phi(p, \dot{p}), \end{aligned}$$

where Φ_{hp} and Φ_{gp} are the maps from $h \rightarrow p$ and $g \rightarrow p$, respectively. g is typically a sinusoidal input to the dither-piezo. This feedback interconnection is evidently quite complex. However, by treating h as an independent input to the cantilever, qualitative arguments about bandwidth and resolution during imaging can be obtained. It is then verified through simulation where a piece wise linear (spring-mass-damper model) tip-sample interaction potential is assumed.

2.3 Observer based Q-control

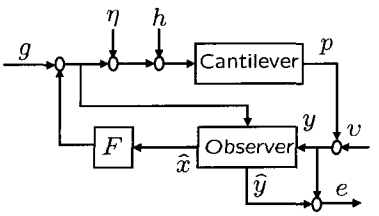


Figure 2.2 Block diagram of AFM combined with observer.

An observer is designed based on the AFM model given by Equation(2.2) that provides the estimated position $\hat{p}$ and velocity $\hat{v}$ of the cantilever tip. The block diagram of AFM combined with an observer is given in Figure 2.2. In observer based Q-control paradigm $\hat{p}$ and $\hat{v}$ are added in feedback to standard excitation signal g. The observer dynamics is given by,

$$\begin{aligned} \dot{\hat{x}} &= A\hat{x} + Bu + L(y - C\hat{x});\ \hat{x}(0) = \hat{x}_0, \\ u &= (g + F\hat{x}), \\ \hat{p} &= C\hat{x}, \\ \hat{v} &= D\hat{x}, \end{aligned} \tag{2.3}$$

where L is the observer gain and F is the state feedback gain.

Let $\tilde{x} = x - \hat{x}$ denote the state estimation error. The combined cantilever-observer dynamics

is given by,

$$\begin{bmatrix} \dot{x} \\ \dot{\tilde{x}} \end{bmatrix} = \begin{bmatrix} A+BF & -BF \\ 0 & A-LC \end{bmatrix} \begin{bmatrix} x \\ \tilde{x} \end{bmatrix} + \begin{bmatrix} B \\ 0 \end{bmatrix} g + \begin{bmatrix} B_1 \\ B_1 \end{bmatrix} (h+\eta) - \begin{bmatrix} 0 \\ L \end{bmatrix} v,$$
$$\begin{bmatrix} y \\ \hat{p} \\ \hat{v} \end{bmatrix} = \begin{bmatrix} C & 0 \\ C & -C \\ D & -D \end{bmatrix} \begin{bmatrix} x \\ \tilde{x} \end{bmatrix} + \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} v. \tag{2.4}$$

It can be shown from Equation(2.4) that the transfer function from excitation signal g to cantilever-tip deflection p is given by,

$$\begin{aligned} (g \to p)(s) &= C[sI - (A+BF)]^{-1}B \\ &= \frac{k/m(c_1 + c_2 s)}{s^2 + s\frac{\omega_{qc}}{Q_{qc}} + \omega_{qc}^2}, \end{aligned} \tag{2.5}$$

where $F = [F_1,\ F_2](m/k)$, $\frac{\omega_{qc}}{Q_{qc}} = \frac{\omega_0}{Q} - F_2$ and $\omega_{qc}^2 = \omega_0^2 - F_1$. Note that ω_{qc} and Q_{qc} are the modified resonant frequency and quality factor of cantilever.

In Equation(2.5), $(g \to y)(s)$ is an exact second order transfer function. It is independent of observer gain L. In this paradigm the modified transfer function has the same structure as the unmodified transfer function (compare Equation(2.1) with Equation(2.5)) with an equivalent quality factor Q_{qc} and equivalent resonant frequency ω_{qc}. Thus one can imagine that the new dynamics is to mimic the dynamics of a new cantilever with changed quality factor Q_{qc} and resonant frequency ω_{qc}. This provides considerable advantages in analyzing the Q-controlled cantilever dynamics. It needs to be noted that the equivalent transfer function is not of higher order. In the above architecture suitable pole-zero cancelation has taken place to obtain a second order description of the closed-loop system even though the state space dimension is four. Note that the transfer function $(g \to p)(s)$ in existing methods (with a phase shifter or time delay in the feedback link) is of high order or an infinite dimensional linear system [27]. Therefore in existing methods one has to *approximate* the resulting transfer function to a second-order behavior with the purpose of carrying the intuition of a modified cantilever to

the interpretation of the cantilever-sample interconnection.

For quantifying Q in existing Q-control methods it is assumed that the cantilever is a second order transfer function system with a constant numerator and the cantilever deflection is purely sinusoidal so that a 90 degrees phase shift yields the velocity of the cantilever. Both these assumptions typically do not hold. In the proposed method the feedback signal is the *estimated* tip-velocity. From Equation(2.5) it can be seen that by choosing F appropriately a desired quality factor and resonance frequency of the cantilever can be obtained. Note that $\omega_{qc}^2 = \sqrt{k_{qc}/m}$, k_{qc} being the modified stiffness of the cantilever. Therefore a desired stiffness of the cantilever can be obtained by feeding back the estimated tip position signal. A desired quality factor can be achieved by varying F_2 only. In experiments and simulations active Q-control of cantilever by feeding back estimated tip velocity $\hat{p}$ is considered (i.e. $F_1 = 0$).

When the cantilever is not interacting with sample, bandwidth is entirely determined by the transfer function $(g \rightarrow y)(s)$ as $h = 0$. Therefore off-sample behavior of cantilever can be tuned by choosing F appropriately. Typically, one would need large damping (low Q_{qc}) off-sample for faster imaging. Resolution does not have any significance off-sample and therefore high bandwidth is the only objective. On-sample behavior is analyzed next.

2.4 Effect of tip-sample force

The effect of tip-sample force h on tip deflection p during imaging is given by the following transfer function:

$$\begin{aligned}(h \rightarrow p)(s) &= C[sI - A_{BF}]^{-1}B_1 - C[sI - A_{BF}]^{-1}BF[sI - A_{LC}]^{-1}B_1, \\ &= \frac{1/m(c_1 + c_2 s)}{s^2 + s\frac{\omega_{qc}}{Q_{qc}} + \omega_{qc}^2}\left[1 - \frac{F_1(1 - l_1c_2) + F_2(s + l_1c_1)}{s^2 + s\frac{\omega_{kf}}{Q_{kf}} + \omega_{kf}^2}\right],\end{aligned}$$

where $A_{BF} = A + BF$, $A_{LC} = A - LC$, $\frac{\omega_{kf}}{Q_{kf}} = \frac{\omega_0}{Q} + l_1c_1 + l_2c_2$ and $\omega_{kf}^2 = \omega_0^2(1 - l_1c_2) + l_1c_1\frac{\omega_0}{Q} + l_2c_1$.

The bandwidth and resolution during imaging depends on the maps $(h \rightarrow p)(s)$ and $(g \rightarrow p)(s)$. If gain of $(h \rightarrow y)(s)$ is large, the effect of h on y is large and consequently resolution

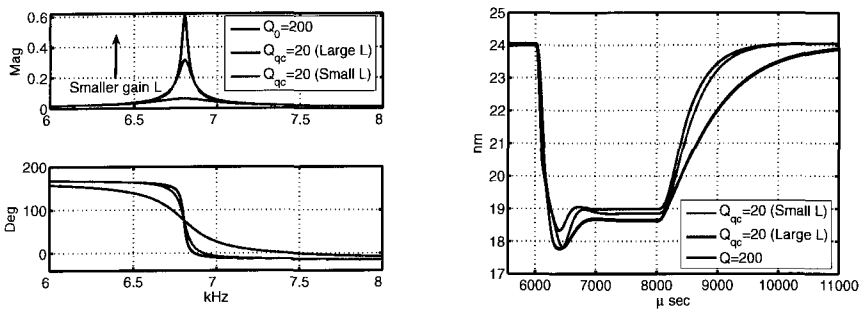
is high. If bandwidth of $(h \rightarrow y)(s)$ is large, tip deflection due to h settle down faster and consequently bandwidth is high. From Equation(2.6) it follows that as L is increased from 0, $(h \rightarrow y)(s)$ changes from original uncontrolled cantilever dynamics given by Equation(2.1) to a modified dynamics given by Equation(2.5). Thus one can achieve a wide range of on-sample behavior of cantilever by choosing L appropriately.

In effect, observer based methodology provides a "dual Q" control; (a) choice of F determines off-sample behavior and (b) choice of L determines on-sample behavior.

While imaging in air Q-control is often employed to decrease quality factor in order to increase scan speed. While imaging in buffer Q-control is often employed to increase quality factor in order to increase force sensitivity. Observer based Q-control approach is explored both in air and under buffer.

In a simulation, a cantilever having $f_0 = 68.1$ kHz while oscillating in air with $Q = 200$ is damped to $Q_{qc} = 20$ by choosing $F = k/m[0, 9\omega_0/Q]$. $(h \rightarrow y)(s)$ for undamped cantilever and damped cantilever for different values of observer gain L are plotted in Fig. 2.3(a) in black, red and blue, respectively. As the values of L increases, $(h \rightarrow p)(s)$ approaches to damped $(g \rightarrow p)(s)/k$. By choosing smaller values of L the effect of tip-sample forces on tip deflection y can be enhanced. In Fig. 2.3(b) the amplitude of deflection signal y is plotted when the cantilever encounters a step sample profile of 5 nm height between time instants 6000 μsec and 8000 μsec while freely oscillating at 24 nm. For smaller values of L the cantilever settles down to a lower amplitude further away from the sample compared to larger values of L. The tip-sample interaction force is smaller for smaller L. Thus a higher force sensitivity and resolution can be achieved by using a smaller value for L without compromising the off-sample bandwidth. When the step sample profile goes away at 8000 μsec i.e. $h = 0$, the deflection signal builds up to the free oscillation amplitude at the desired rate. Thus one can conclude for a fixed off-sample damping (fixed F), tip-sample force resolution can be improved by choosing smaller value for L.

In another simulation, the Q of a cantilever ($f_0 = 15$ kHz) oscillating under water with $Q = 40$ is enhanced to $Q_{qc} = 400$ by choosing $F = k/m[0, -0.9\omega_0/Q]$. $(h \rightarrow p)(s)$ for



(a) (b)

Figure 2.3 (a) Transfer function from tip-sample interaction h to tip deflection p is plotted for different values of observer gain L (shown in red and blue) when the cantilever is damped in air. (b) Amplitude signal for different values of observer gain L is plotted when the cantilever encounters a step sample profile of 5 nm height, and between 6000 and 8000 μseconds while freely oscillating in air with amplitude 24 nm. Off-sample the amplitude increases at a faster rate when the cantilever is damped for different values of L and fixed state feedback gain F. On-sample the cantilever settles to a lower magnitude away from the sample when observer gain L is low.

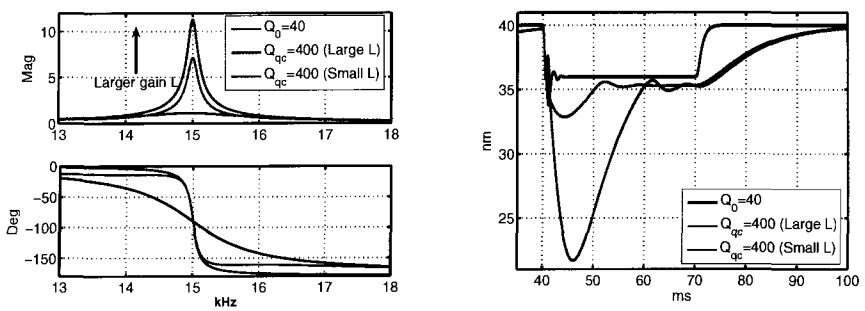


Figure 2.4 (a) The transfer function from tip-sample interaction h to tip deflection p is plotted for different values of observer gain L (shown in red and blue) when quality factor the cantilever is enhanced under water. (b) Amplitude signal for different values of observer gain L is plotted when the cantilever encounters a step sample profile nm height, and between 40 and 70 milliseconds while freely oscillating under water with amplitude 40 nm. Off-sample the amplitude increases at same rate when the Q is increased using different values of L and fixed feedback gain F. On-sample the cantilever settles faster when observer gain L is low.

unchanged cantilever and Q-controlled cantilever with different values of observer gain L are plotted in Fig. 2.4(a) in black, blue and red respectively. As L is increased, $(h \rightarrow p)(s)$ approaches $(g \rightarrow p)(s)/k$. In Fig. 2.4(b) amplitude of the deflection signal y is plotted when the cantilever encounters a step sample profile of 5 nm height between time instants of 40 ms and 70 ms while freely oscillating at 40 nm. The cantilever settles down to similar amplitudes on the sample for different values of L. When a smaller value for L is chosen the cantilever settles down faster. By choosing small value for L the force sensitivity or resolution is not compromised; however the bandwidth is enhanced in presence of sample. When the step sample profile goes away at 70 ms i.e. $h = 0$, the deflection signal builds up to the free oscillation amplitude at the bandwidth corresponding to the new enhanced Q . The phase of the deflection signal y shows a similar behavior as the amplitude signal. Thus one can conclude that for a fixed Q-enhancement (fixed F) for increased force resolution, the bandwidth can be improved by choosing smaller value for L.

2.5 Effect of noise

The effect of thermal noise η and photo-diode noise υ on the deflection signal y is given by the transfer functions:

$$(\eta \rightarrow y)(s) = C[sI - A_{BF}]^{-1}B_1 - C[sI - A_{BF}]^{-1}BF[sI - A_{LC}]^{-1}B_1, \tag{2.6}$$

and

$$(\upsilon \rightarrow y)(s) = I + C[sI - A_{BF}]^{-1}BF[sI - A_{LC}]^{-1}L, \tag{2.7}$$

respectively.

The effect of thermal noise η and photo-diode noise υ on estimated cantilever-tip position $\hat{p}$ is given by the transfer functions:

$$(\eta \rightarrow \hat{p})(s) = C[sI - A_{BF}]^{-1}L - C[sI - A_{LC}]^{-1}B_1 - C[sI - A_{BF}]^{-1}BF[sI - A_{LC}]^{-1}B_1, \tag{2.8}$$

and

$$(\upsilon \rightarrow \hat{p})(s) = C[sI - A_{LC}]^{-1}L + C[sI - A_{BF}]^{-1}BF[sI - A_{LC}]^{-1}L, \tag{2.9}$$

respectively.

The severity of the effect of thermal noise η and photo-diode noise υ on deflection signal y and tip-position estimate $\hat{p}$ can be obtained by analyzing Eq.(2.6), (2.8), (2.7) and (2.9) for different values of gain F and L.

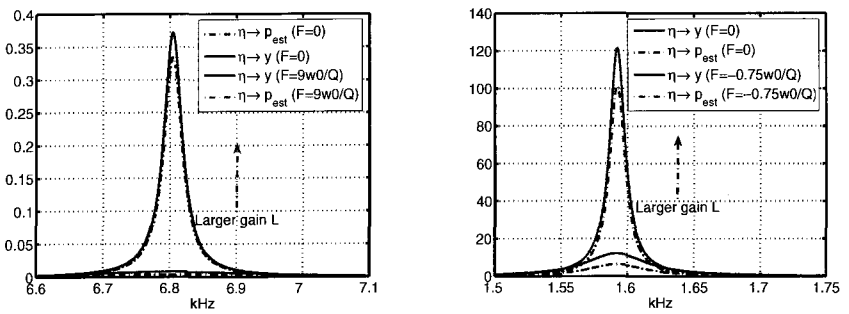


Figure 2.5 Transfer function from thermal noise η to deflection signal y and tip-position estimate $\hat{p} = p_{est}$ are plotted (a) when the cantilever is in air and damped and (b) when the cantilever is under water and its Q is enhanced.

In Fig. 2.5(a) and (b) thermal noise power in tip-deflection y and tip-position estimate $\hat{p}$ are shown when the cantilever is in air and under water, respectively. In Fig. 2.6(a) and (b) photo-diode noise power in tip-deflection y and tip-position estimate $\hat{p}$ are shown when the cantilever is in air and under water, respectively. It can be observed that when the cantilever is damped by an order of magnitude the noise contribution is also reduced by an order of magnitude. For fixed damping (fixed F) noise in $\hat{p}$ is lesser if value of L is smaller. It can be observed that when quality factor of the cantilever is enhanced by 4 times the noise contribution increases by over an order. Therefore, photo-diode noise may have drastic effects while imaging under fluid with Q enhancement. For fixed Q-enhancement (fixed F) noise in $\hat{p}$ is lesser if value of L

is smaller.

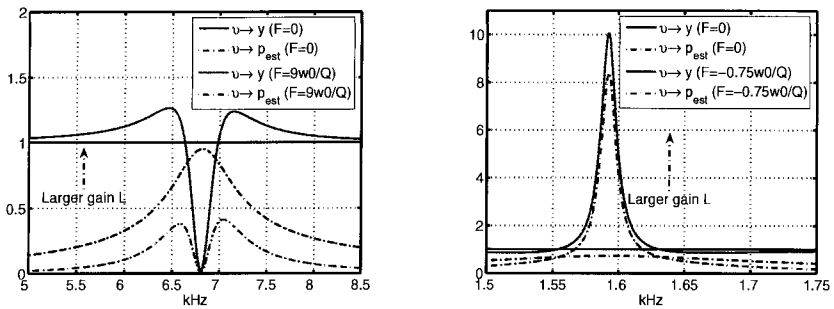


Figure 2.6 Transfer function from photo-diode noise v to deflection signal y and tip-position estimate $\hat{p} = p_{est}$ are plotted (a) when the cantilever is in air and damped and (b) when the cantilever is under water and its Q is enhanced.

2.6 TF-AFM with Q-control

Observer based imaging has numerous benefits with respect to resolution, bandwidth and noise reduction. However the imaging signals i.e. the amplitude and phase of deflection signal y and estimated tip position $\hat{p}$ remain slow. Therefore at a given scan speed these signals represent a low bandwidth content (slow varying features) of the actual sample profile. As the scan speed is increased the image looks smoother and the fine details of the sample is lost. Transient force atomic force microscopy (TF-AFM) technique has promise of detecting high bandwidth content of sample profile i.e. small tip-sample interactions corresponding to sub-nanometer level sample features with a high probability. It is shown that TF-AFM and dual Q-control can be integrated independently.

The effect of excitation signal g on innovation signal $e = y - C\hat{x}$ is given by transfer function: $(g \rightarrow e)(s) = 0$. The effect of tip-sample interaction force h on innovation signal e

with Q-control is given by transfer function:

$$\begin{aligned}(h \rightarrow e)(s) &= C[sI - (A - LC)]^{-1}B_1, \\ &= \frac{1/m(c_1 + c_2 s)}{s^2 + s\frac{\omega_{kf}}{Q_{kf}} + \omega_{kf}^2}.\end{aligned} \tag{2.10}$$

Therefore, transient signal is not affected by the choice of F. TF-AFM technique can be implemented independent of the Q-control scheme. From Eq.(2.10) one can conclude that the bandwidth of detecting tip-sample interaction ($BW_{kf} \propto \frac{\omega_{kf}}{Q_{kf}}$) can be increased (which is independent of Q [9]) independent of state feedback gain F.

2.7 Experimental Results

In the experiments frequency response of AFM setup is obtained via a HP 3563A control system analyzer. A model of AFM is obtained from frequency response data by fitting a second order transfer function with a right half plane zero, given by Equation(2.1).

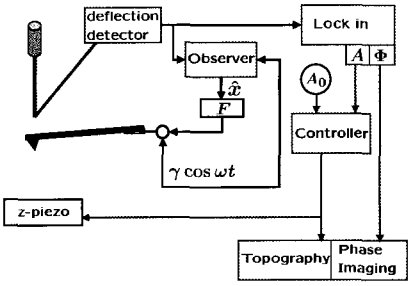


Figure 2.7 An observer is implemented into the standard tapping mode imaging set up. The state estimates are used for Q-control and imaging.

Noise power of measurement and thermal noise are obtained experimentally. Measurement noise (mainly photo-diode noise) is calculated from the deflection signal when cantilever is forced to remain in contact with a hard sample surface. The thermal noise is calculated from deflection signal of freely oscillating cantilever with no excitation signal ($g = 0$).

A conventional tapping mode AFM set up with an observer is shown in Figure 2.7. This

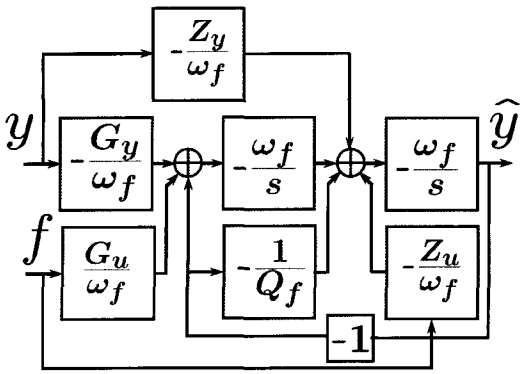


Figure 2.8 The observer is implemented as an analog circuit (Tow-Thomas biquad). It has a second order transfer function with gain G_u and a right half plane zero Z_u from input u, and gain G_y and a left half plane zero Z_y from input y. ω_f and Q_f are the resonant frequency and quality factor of the observer. A similar circuit is implemented that gives the cantilever-tip velocity estimate $\hat{v}$.

setup is used for the experiments. The cantilever is excited near its resonant frequency ω_0 with $g = \gamma \cos \omega_0 t$. A PI controller actuates z-piezo to control the sample position in order to maintain a constant set-point amplitude (A_0) of cantilever deflection signal. The observer provides the estimates of cantilever tip position and velocity. Quality factor Q of cantilever is regulated by feeding back estimated tip-velocity with feedback gain F ($F_1 = 0$).

The following experimental plots demonstrate the tuning of analog kalman filter from frequency response data of the cantilever. The frequency response of a cantilever oscillating in air with $f_0 = 68.1$ kHz and $Q = 110$ in plotted in Figure 2.9(a). A Kalman observer is designed in Matlab and the transfer functions from observer inputs dither excitation g and deflection y to observer outputs estimated tip position and velocity are calculated. Impulse response of cantilever and observer are plotted in Figure 2.9(b) at top and bottom, respectively. Due to high quality factor cantilever takes approximately 2000 μseconds to settle down. Observer is designed such that it tracks within 40 μseconds. The observer transfer functions are implemented in analog circuit and the corresponding frequency responses were measured using HP3563A control systems analyzer. The calculated and measured transfer functions are

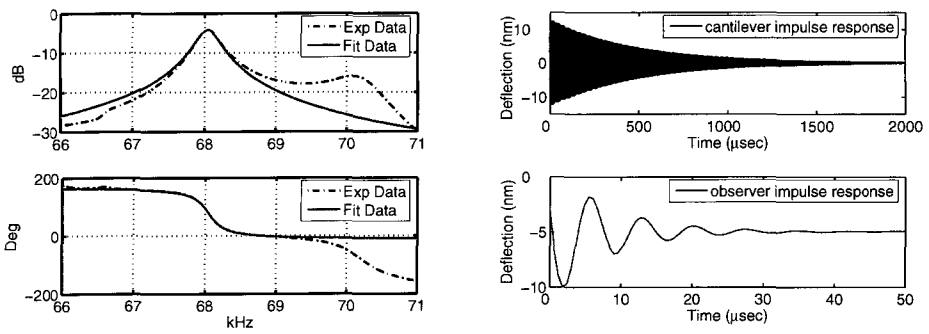


Figure 2.9 Frequency response data is fitted with a second order transfer function with a right half plane zero. Cantilever has resonant frequency $f_0 = 68.1$ kHz and quality factor $Q = 110$. (b) Impulse response of cantilever (top) and observer (bottom) are plotted. Due to high quality factor cantilever takes nearly 2000 μseconds to settle down. Observer tracks within 40 μseconds.

plotted in Figure 2.10.

In another experiment a cantilever with resonant frequency $f_0 = 71.25$ kHz and quality factor $Q = 110$ was damped to $Q_{qc} = 70$ by using different observers implemented in the analog circuit and fixed feedback gain F_2. The undamped and damped frequency response data of $g \rightarrow y$ are plotted in Figure 2.11(a). Therefore off-sample quality factor of cantilever is independent of observer gain L and depends only on feedback gain F (see Equation(2.5)). In a similar experiment the frequency response for $h \rightarrow y$ was measured for different observer gains L (see Figure 2.11(b)). For different values for L a variety of response varying between damped cantilever response and undamped cantilever response was obtained. Therefore on-sample quality factor of cantilever depends on observer gain L for fixed state feedback gain F. Thus on-sample behavior of cantilever can be changed independent of its off-sample behavior and a better bandwidth resolution trade off can be achieved.

In another experiment off-sample behavior of cantilever during imaging was investigated for different values of observer gain L and fixed feedback gain F. A cantilever with resonant frequency $f_0 = 68.1$ kHz was damped from quality factor $Q = 110$ to $Q_{qc} = 70$ in air using

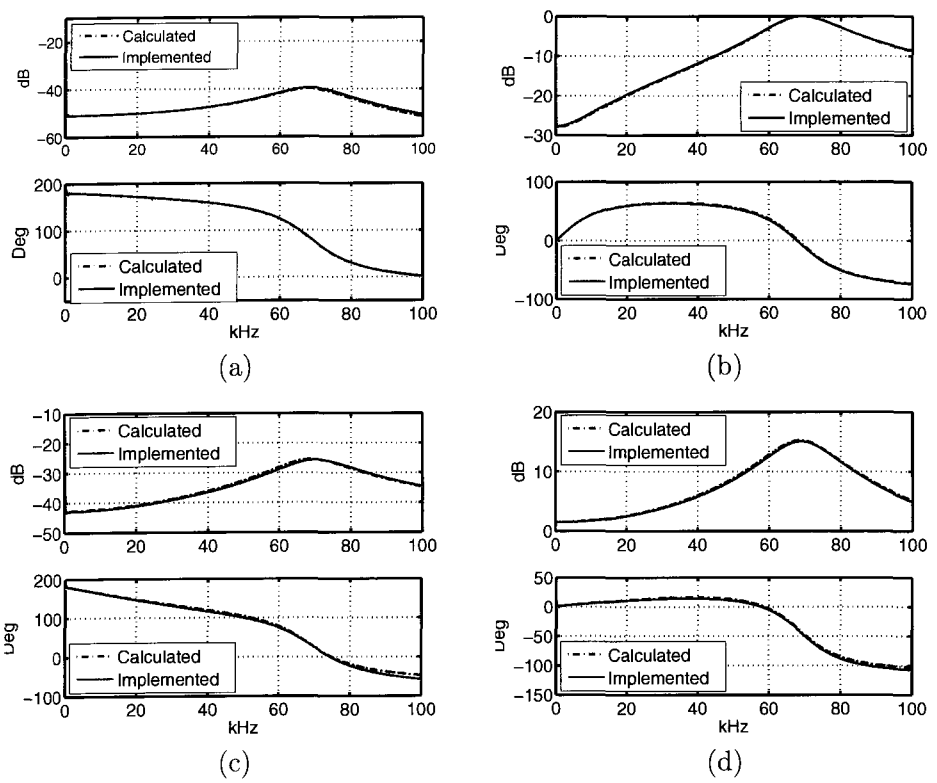


Figure 2.10 Calculated transfer function of Kalman observer from inputs dither excitation and deflection signal to outputs estimated tip position and velocity are plotted with the corresponding transfer functions from implemented analog circuit.

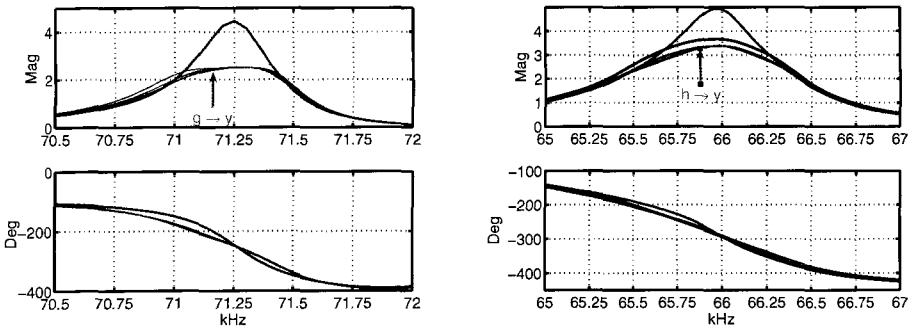


Figure 2.11 (a)Frequency response from dither excitation to deflection $g \rightarrow y$ is plotted when observers with different gain values L were employed. (b)Frequency response for $h \rightarrow y$ is plotted when observers with different gain values L were employed.

analog observer circuit. The amplitude and phase profile while imaging a grating sample is plotted in Figure 2.12 for different values of L and fixed F. The grating sample had 20 nm high steps. Free oscillation amplitude was 97.3 nm and set point amplitude was 87.3 nm. Proportional and integral feedback gains for set point amplitude control were $K_p = 0$ and $K_i = 0.2$, respectively. In Figure 2.12 the amplitude and phase build to free oscillation amplitude and phase values faster when the cantilever is damped. The rate of growth of amplitude and phase profiles off-sample are equal for different values of L.

In another experiment quality factor of a cantilever was increased under water. A soft cantilever (Olympus Biolever-B) with resonant frequency $f_0 = 15.9$ kHz, nominal quality factor $Q = 5$ and stiffness $k = 0.005$ N/m in water was used on Asylum Research MFP3D AFM. The frequency response showed forest of peaks due to acoustic excitation from surrounding fluid. Thermal noise response of the cantilever was taken and the peak near the thermal peak was identified as the resonant peak of the cantilever in water. A second order transfer function with a right half plane zero was fit to the frequency response data near the resonant peak of cantilever. A second order transfer function (with a right half plane zero) was fitted to the frequency response data. An analog observer was designed and the effective quality factor

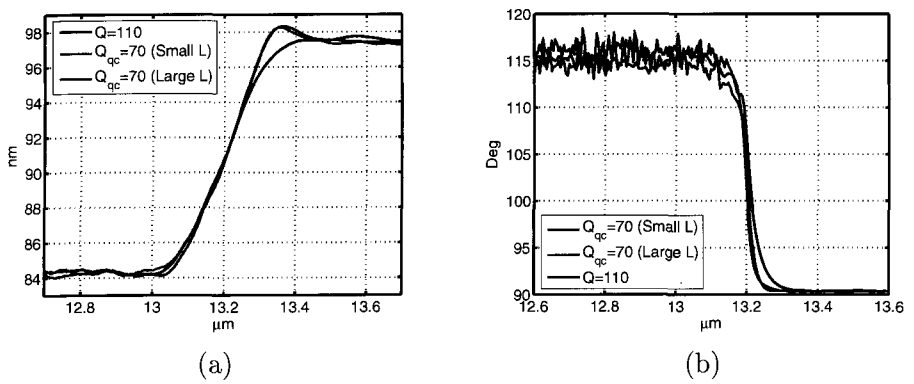


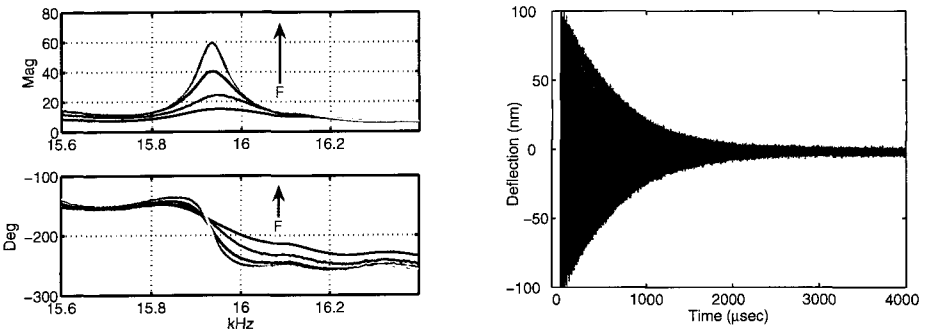
Figure 2.12 (a) Amplitude and (b) phase profiles are plotted when a cantilever comes off a 20 nm step in a calibration sample near 13 μm. Amplitude and phase dynamics are fast when the cantilever is damped. Off sample amplitude and phase dynamics are independent of observer gain L.

of the cantilever was increased by gradually increasing feedback gain F. Quality factor of the cantilever was increased to 350 under water (see Figure 2.13(a)). Effective Q change was verified by measuring the settling time of cantilever when dither excitation was withdrawn. When effective quality factor was high cantilever took longer time to settle down.

Analysis shows that observer based Q-control provides flexibility in managing bandwidth and resolution trade off. All the benefits of traditional Q-control are achieved by choosing suitable feedback gain F to shape off sample behavior of the cantilever independent of observer gain L. On sample behavior of the cantilever is managed by choosing suitable value for L. It is verified in experiments that observer based Q-control achieves all the benefits of traditional Q-control methods. The benefits of dual Q-control (i.e. on sample behavior of cantilever) is extensively investigated in simulations.

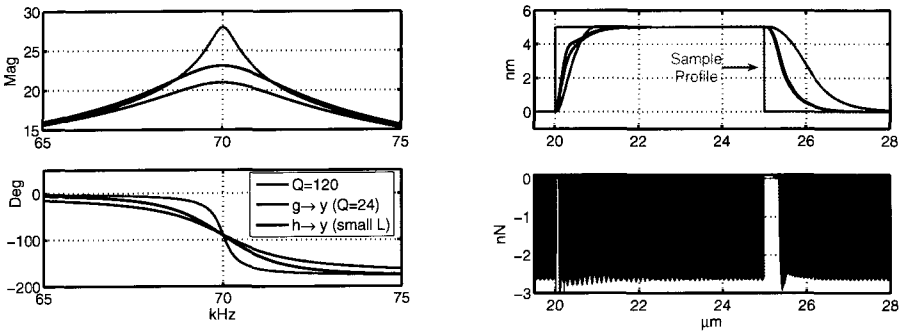
In simulations an experimentally obtained z-piezo model as shown in Figure 1.5 was used. Cantilever model was also experimentally obtained by fitting a second order transfer function to frequency response data near resonant peak. The tip-sample interaction model was a piece wise linear model.

In Figure 2.14 quality factor of a cantilever was reduced from $Q = 120$ to $Q = 24$ by both



(a) (b)

Figure 2.13 (a) Frequency response of a cantilever in water is plotted when its quality factor is gradually increased from a nominal value of 5 to 300 by feeding back estimated velocity signal from an analog observer to dither excitation signal. (b) Settling time of cantilever with Q-control is plotted when dither excitation was withdrawn. When effective quality factor was high cantilever took longer time to settle down.



(a) (b)

Figure 2.14 (a) Cantilever with $Q = 120$ (red) is damped to $Q = 24$ (blue). The transfer functions from dither to deflection $g \rightarrow y$ and tip-sample force to deflection $h \rightarrow y$ are plotted in blue and black, respectively. For traditional methods $h \rightarrow y = g \rightarrow y$. (b)Top: Images of a step sample profile of height 5 nm are plotted where red, blue and black correspond to $Q = 120$, $Q = 24$ traditional and $Q = 24$ dual Q method. When Q was reduced image became sharper. Bottom: Tip-sample interaction forces are plotted. Tip-sample interaction force is small in dual Q method.

traditional and dual Q method. The transfer functions from dither to deflection $g \rightarrow y$ and tip-sample force to deflection $h \rightarrow y$ are plotted in blue and black in Figure 2.14(a), respectively. For traditional methods $h \rightarrow y = g \rightarrow y$. The images and tip-sample interaction forces of a step sample profile of 5 nm height is plotted in Figure 2.14(b). The image is sharp when the cantilever is damped (blue and black). Both Q control methods provided similar images. However, tip-sample interaction force was small when dual Q-control was employed.

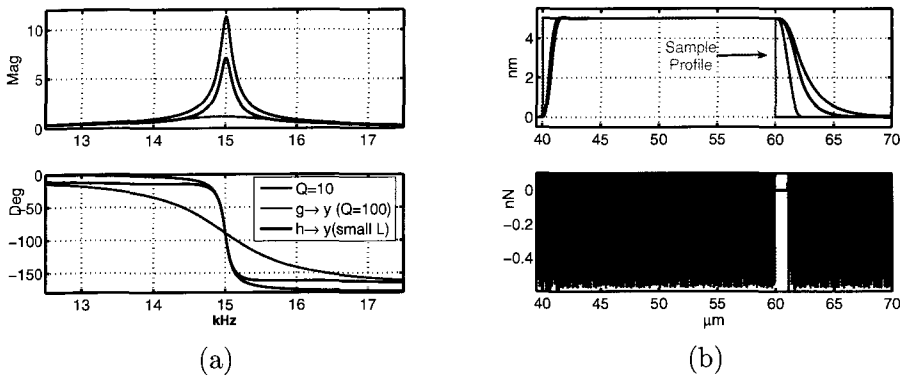


Figure 2.15 (a) Quality factor of a cantilever with $Q = 10$ (red) is enhanced to $Q = 100$ (blue). The transfer functions from dither to deflection $g \rightarrow y$ and tip-sample force to deflection $h \rightarrow y$ are plotted in blue and black, respectively. For traditional methods $h \rightarrow y = g \rightarrow y$. (b)Top: Images of a step sample profile of height 5 nm are plotted where red, blue and black correspond to $Q = 10$, $Q = 100$ traditional and $Q = 100$ dual Q method. When Q was increased images lost sharpness. However, dual Q method produced a better image than traditional method. Bottom: Tip-sample interaction forces are plotted. When Q was increased tip-sample force decreased by both traditional and dual Q method.

In Figure 2.15 quality factor of a cantilever was increased from $Q = 10$ to $Q = 100$ by both traditional and dual Q method. The transfer functions from dither to deflection $g \rightarrow y$ and tip-sample force to deflection $h \rightarrow y$ are plotted in blue and black in Figure 2.15(a), respectively. For traditional methods $h \rightarrow y = g \rightarrow y$. The images and tip-sample interaction forces of a step sample profile of 5 nm height is plotted in Figure 2.14(b). The image lost sharpness

when Q was increased (blue and black). However, dual Q method produced better image than traditional method. Tip-sample interaction force was reduced when Q was increased. Both Q control methods reduced tip-sample force similarly.

2.8 Conclusions

Observer based method provides an exact design and performance assessment for active Q control in AFM. The transfer function from dither input to photo-diode output is independent of the observer so that the cantilever effectively behaves like a spring-mass-damper system unlike the complex behavior in existing methods. The effective quality factor and stiffness of the cantilever can be changed by appropriately choosing the state feedback gain. The effect of tip-sample interaction force on the bandwidth and resolution during imaging is qualitatively analyzable in this framework. The observer provides flexibility in the state feedback loop to improve the bandwidth or resolution during imaging. Dual Q-control performed as good as traditional Q-control methods in experiments. Analysis and simulation showed that dual Q-control can manage bandwidth and resolution trade off better than traditional methods. Transient force atomic force microscopy (TF-AFM) is married to observer based Q control method. TF-AFM performs independent of dual Q-control method.

CHAPTER 3 FEEDBACK CONTROL FOR REAL-TIME SCHEDULING

Most of real-time scheduling algorithms are open-loop algorithms as the scheduling decisions are based on the worst-case estimates of task parameters. They do not continuously observe the performance of the system and therefore do not dynamically adjust the system parameters accordingly for improving the performance. In many cases, it is preferable to base scheduling decisions on *average-case* workload parameters and be ready to deal with bounded transient overloads dynamically. In recent years, the "closed-loop" scheduling has gained importance due to its applicability to many real-world problems wherein the feedback information can be exploited efficiently to adjust task and/or scheduler parameters, thereby improving the system's performance.

In this dissertation, we discuss an open-loop dynamic scheduling algorithm that employs a notion of task overlap in the scheduler in order to provide some flexibility in task execution time. Then we present a novel closed-loop approach for dynamically estimating the execution time of tasks based on both deadline miss ratio and task rejection ratio in the system. This approach is highly preferable for firm/soft real-time systems since it provides a firm performance guarantee in terms of deadline misses while achieving a high guarantee ratio. We design proportional-integral controller and H-infinity controller for closed loop scheduling. We evaluate the performance of the open-loop and the closed-loop approaches using simulation studies. We show that the closed-loop dynamic scheduling offers a better performance over the open-loop scheduling under all the practical conditions.

3.1 Introduction

Modeling a computing system as a dynamic system or as a controller is an approach that has proved to be fruitful in many cases. For example, the step-length adjustment mechanism in numerical integration algorithms can be viewed as a PI-controller [28], and the traveling salesman optimization problem can be solved by a nonlinear dynamical system formulated as a recurrent neural network. In this dissertation we develop a similar notion for dynamic task schedulers for a real-time system.

In real-time systems, the system's performance depends not only on the logical correctness of the result, but also on the time at which the results are produced [29]. Based on the time criticality, these systems are classified into : hard, firm or soft real-time systems, in the decreasing order of damages caused due to missing a deadline. So, in hard real-time systems, missing a deadline can lead to catastrophic conditions whereas in firm/soft real-time tasks, missing a deadline will not have such serious consequences. Multiprocessors and multicomputer based systems have emerged as a powerful computing means for these real-time systems, primarily due to their capability for high performance and reliability. The problem of scheduling real-time tasks in such systems is to determine when and on which processor a given task is executed [29]. These real-time scheduling algorithms, which does this assignment of tasks to their processors, fall into two categories: static and dynamic scheduling. In static scheduling, the assignment of tasks to processors is done apriori. These algorithms require an apriori knowledge of task's characterisitics and cannot schedule tasks at run-time (such as aperiodic tasks), which requires dynamic scheduling.

In dynamic scheduling, the scheduling algorithm does not require the complete knowledge of the task set and its constraints.Among the dynamic scheduling algorithms, some of them operate under resource-sufficient environments and others operate under resource-insufficient environments [30, 31]. Though significant results have been achieved in real-time scheduling with algorithms such as RMS, EDF and Spring Scheduling, these algorithms perform well when the workload of the system can be accurately modeled, and does not perform well for unpredictable workloads. Further, all the scheduling algorithms discussed, work based on the

worst case execution time (WCET). Practically, identifying the correct estimate for WCET is not easy and many analysis tools come up with over-estimation strategies. Thus, scheduling algorithms based on WCET leads to under-utilized system. In many cases, it is preferable to base scheduling decisions on average execution time and to be ready to deal with bounded transient overloads dynamically. This approach is especially preferable in firm/soft real-time systems since it provides a firm performance guarantee in terms of deadline misses while achieving high utilization and throughput (guarantee ratio) at the same time.

The requirements of an ideal firm real-time scheduling algorithm are to (1) provide (firm) performance guarantees to admitted tasks, i.e., maintain low deadline miss ratio among admitted tasks; and (2) admit as many tasks as possible, i.e., achieve high guarantee ratio. In an unpredictable environment, it is impossible for a system to achieve 100% utilization and a 0% miss ratio all the time and a tradeoff between miss ratio and utilization is unavoidable. Two approaches can be used to deal with this tradeoff. The first approach, which uses an admission control based on worst case estimation, represents the pessimistic approach, where deadline misses are avoided at the cost of low utilization and throughput. This approach has been widely used in hard real-time systems. The second approach, as followed in [32], uses the closed-loop scheduling based on average case estimation, represents the optimistic approach, where a low (but possibly non-zero) miss ratio is maintained with high utilization. When high misses happen due to underestimation of the execution time, the scheduler corrects the system's state back to the satisfactory state, i.e., a state with low miss ratio and high utilization and throughput. This optimistic approach would work as follows: start with a schedule based on the nominal assumptions of the incoming tasks. The system would then measure the actual performance of the schedule, compare it to the system requirements and detect differences. The system would call control functions to assess the impact of these differences and apply a correction to keep the system within an acceptable range of performance.

The key issue addressed in this dissertation is the relaxation of the requirement on a known *worst-case* workload parameters. Our main approach to this issue is to design a closed-loop scheduling algorithm, in which, the tasks are scheduled using their *average-case execution time*

(AvCET) as estimation for their *actual execution time (AET).* A feedback from the system is used to generate the error term and this error term is used for future estimation of execution time. We model and analyze the closed-loop scheduling algorithm using existing control theory. The result is that this scheduling paradigm has low miss ratio and high guarantee ratio thereby improving the productivity of the firm/soft real-time systems. Specifically in this dissertation, we use the dynamic planning based schedulers (e.g spring schedulers) and schedule the tasks based on the estimated execution time, based on the estimation factor obtained by the feedback controller. The feedback controller does its estimation with the two parameters, miss ratio and the rejection ratio, being constantly fed back to the controller for estimation.

Specifically, in this dissertation, we use an open-loop dynamic scheduling algorithm that employs a notion of task overlap in the scheduler in order to provide some flexibility in task execution times. This algorithm, dynamically guarantees incoming firm tasks via on-line admission control and planning based on their *AvCET.* We use feedback control theory to design a closed-loop scheduling algorithm derived from the open-loop algorithm. The loop is closed by feeding back both deadline miss ratio and task rejection ratio.

The rest of the chapter is organized as follows. In Section 2, we define the system model, which includes the task model and scheduler model, and the performance metrics of the system. In Section 3, we present the related work and also discuss the system model and its limitations on performance. In Section 4, we discuss the design of the PI controller and a H_∞ controller as a part of the closed loop scheduling. We also study the response of the PI controller and H_∞ controller and verify the models. In Section 5 we study the performance of the closed loop algorithms and compare it with the open loop scheduler. Then we conclude in section 6.

3.2 System Model

3.2.1 Task Model

The following assumptions form the task model.

- Tasks are aperiodic, i.e., task characteristics are not known a priori. Every task T_i has the following attributes: arrival time (a_i), ready time (r_i), worst-case execution time

($WCET_i$), best-case execution time ($BCET_i$), and firm deadline (d_i).

- Tasks are non-preemptable, i.e., when a task starts execution on a processor, it finishes to its completion.
- Tasks may have resource and/or precedence constraint.

3.2.2 Scheduler Model

In our dynamic multiprocessor scheduling, all the tasks arrive at a central processor called the scheduler, from where they are dispatched to other processors in the system for execution. These processors are identical and are connected through a shared medium. The communication between the scheduler and the processors is through the dispatch queues. Each processor has its own dispatch queue and the scheduler will be running in parallel with the processors, scheduling the newly arriving tasks and periodically updating the dispatch queues. The scheduler considered in this dissertation is a dynamic scheduler, which does an admission check, wherein the tasks are rejected if they are found non-schedulable and if the task is accepted by the admission controller, then the scheduler constructs the appropriate schedule and arranges the tasks in their appropriate dispatch queues.

3.2.3 Performance Metrics

- Task hit ratio (HR): This is the ratio of the number of admitted tasks that meet their deadlines to the total number of tasks admitted in the system. The missing of deadlines happens at the processors when they (the tasks) are executed. Though the schedulability check of tasks are performed while admitting them, tasks can still miss their deadline when their actual execution time (AET) is greater than their estimated execution time (EET) or due to the occurrence of unanticipated faults in the system. Task miss ratio (MR) is the ratio of the number of admitted tasks that missed their deadlines to the total number of tasks admitted in the system which is equal to one minus the hit ratio.
- Task guarantee ratio (GR): This is the ratio of the number of tasks admitted into the

system to the total number of tasks that arrived in the system. The rejection of tasks happens at the scheduler and depends on the schedulability check algorithm, estimated execution time, and the time at which the schedulability check is performed for a task. Task rejection ratio (RR) is the ratio of the number of tasks rejected to the total number of tasks arrived in the system which is equal to one minus the guarantee ratio.

- Task effective ratio (ER): This is the ratio of the number of tasks that meet their deadlines to the total number of tasks arrived in the system which is equal to the product of the task hit ratio (HR) and the task guarantee ratio (GR).

3.3 Feedback Controlled Scheduling

Estimation of task execution time is very important in dynamic scheduling of firm real time tasks. An underestimation of execution times may jeopardize the desired behavior of the system and an overestimation will cause performance degradation by wasting system resources [33]. The actual execution time of a task varies between its $BCET$ and $WCET$ due to non deterministic behavior of several low-level processor mechanisms (e.g. caching, prefetching, and DMA data transfer), and also due to the fact that the actual execution time for these tasks are function of the system state, and the amount, nature, and the value of input data [34, 35]. In this section, we present a novel approach in which the actual execution time of tasks is dynamically estimated based on the current deadline miss ratio (MR) and task rejection ratio (RR) in the system. This output feedback scheme improved the scheduling performance of the algorithm by accepting significantly more number of tasks and meeting more deadlines. Thereby it can improve the productivity of the firm/soft real-time systems.

3.3.1 Related Work

The idea of feedback has been used informally in scheduling algorithms for applications where the dynamics of the computation workload cannot be characterized accurately for a long time. For example, the VMS operating system [36] used multi-level feedback queues to improve system throughput. Internet protocols use feedback to help solve the congestion

problems. Recently, in [37] the authors presented a feedback-based scheduling scheme that adjusts CPU allocation based on application-dependent progress monitors (e.g., the fill-level of buffers). However, the performance of real-time tasks is not addressed in this work. In the area of real-time databases, in [38] the authors proposed *Adaptive Earliest Deadline* (AED), a priority assignment policy based on EDF. In order to stabilize the performance of EDF under overload conditions, AED features feedback control loop that monitors transactions' deadline misses ratio and adjusts transaction priority assignments accordingly. The feedback control in these algorithms has mostly been done in an ad hoc manner and no analysis has been done.

Recently, under the title of quality of service, in multimedia scheduling R&D, the idea of feedback has been used for traffic flow control (e.g. ABR service in ATM networks). Several papers [39, 40, 41] presented feedback control architectures and algorithms for QoS control in communication systems. These works are targeted at supporting end-to-end QoS in multimedia communication and they are not scheduling algorithms. In [39] the authors used a PI-controller and web-content adaptation mechanism for web server QoS resource allocation (in terms of hit rate and bandwidth). This work targeted at a different (non-real-time) application domain.

Feedback scheduling in real-time systems has been a relatively new area and very few works have been done in this area. Both in [42] and [43] the authors proposed to integrate the design of the system controller with the scheduling of real-time control systems. Both papers aim at providing design tools that enable control engineers to take into consideration scheduling in the early design stage of control systems. These algorithms are off-line algorithms. The scheduling algorithms in both of these cases are still open-loop scheduling algorithms such as RMS and EDF. Several works [44, 45, 46, 47] utilized the flexible timing constraints as a mechanism for graceful performance degradation in control systems. In [44, 45] an elastic task model [48] is used to design an elastic control approach. This approach integrates the *continuous design with digitization* (CDD) [45] and the *direct digital design* (DDD) [44] with adjustable frequencies. However, all of the works assumed the execution times are known and focused on how to reassign the periods for tasks to satisfy the utilization constraints. Instead, our work will focus on using feedback control loops to maintain satisfactory system performance when

the task execution times change dynamically.

A notable work in real-time feedback scheduling is [49, 50] where the authors propose the use of a PID controller as an on-line scheduler under the notion of Feedback Control-EDF (FC-EDF). The measured signal (the controlled variable) is the deadline miss ratio for tasks and control signal is the requested CPU utilization. Changes in the requested CPU utilization are effectuated by two mechanisms (actuators). An admission controller is used to control the flow of workload into the system and a service level controller is used to adjust the workload into the system. A simple liquid tank model is used as an approximation for the scheduling system. Recently, in [33] the authors proposed an approach in which task periods can be dynamically adjusted based on the current load. Load is estimated by monitoring the actual computation time of each task. When the estimated load is found to be greater than a certain threshold (e.g. 1 under EDF), the elastic theory is used to enlarge the task periods to find a feasible configuration.

In both works [50, 33] the authors assume that each task has multiple versions that differ in their execution time, the higher the execution time the better the quality (imprecise computation). Instead, our work assumes only one version for each task which is more general and realistic. Moreover, in [50] the feedback mechanism is used to reject tasks in order to keep the total number of missed deadlines below a desired value. Instead, in our work the feedback mechanism is used to obtain a trade off between the guarantee ratio and the hit ratio in order to improve the effective ratio. Also, in this work the admission controller uses the *WCET* of task versions to perform the schedulability test. Instead, in our work we use an estimated execution time for each task to perform the schedulability test. The estimated execution time of the tasks are adjusted using the feedback mechanism. In [33] the control mechanism has not been analyzed to prove its correctness and stability. Instead, in our work we analyze the control mechanism and we present the block diagram for the system and the analytical equations that connect the miss ratio and the rejection ratio with the estimated execution time of tasks.

3.3.2 System and Model Identification

Before applying feedback control techniques, we present the measured variables, the control output and the open-loop characteristics of the system.

3.3.2.1 Measured Variables

The choice of the measured variables depends on the system goal. The performance of firm real-time systems usually depends on: (1) how many tasks it admits (rejects); (2) how many tasks among the accepted task make (miss) their deadlines. Therefore the deadline miss ratio (MR) and the tasks rejection ratio (RR) are natural choices of the measured variables.

In closed-loop approach it is important to be able to measure the appropriate signal on-line. The instantaneous miss ratio (MR_{kT}) is calculated every sample period (T), which is defined as the ratio between the number of tasks that missed their deadlines within the time interval $[(k-1)T,\ kT]$ and the total number of tasks that finished execution within the same interval, where k is the current time instant.

$$MR_{kT} = \frac{\#\ of\ missed\ tasks_{[(k-1)T,\ kT]}}{\#\ of\ finished\ tasks_{[(k-1)T,\ kT]}}. \tag{3.1}$$

In contrast, the average miss ratio (MR) is defined as the time average of the instantaneous miss ratio (MR_{kT}) for the entire run-time.

The instantaneous rejection ratio (RR_{kT}) is calculated every sample period (T), which is defined as the ratio between the number of tasks that are rejected in the time interval $[(k-1)T,\ kT]$ and the total number of tasks that arrived to the system within the same interval.

$$RR_{kT} = \frac{\#\ of\ rejected\ tasks_{[(k-1)T,\ kT]}}{\#\ of\ arrived\ tasks_{[(k-1)T,\ kT]}}. \tag{3.2}$$

In contrast, the average rejection ratio (RR) is defined as the time average of the instantaneous rejection ratio (RR_{kT}) for the entire run-time.

3.3.2.2 Control Output

The control output must be able to affect the value of the measured variable. In the non-preemptive multiprocessor firm real-time system, it is a widely known fact that the deadline miss ratio and the tasks rejection ratio, highly depend on the estimated execution time of tasks. Thus the estimated execution time (EET) of tasks is an appropriate choice for the control output. The estimated execution time (EET_i) of task (T_i) is calculated using the following equation:

$$EET_i = AvCET_i + etf_{kT}\left[AvCET_i - BCET_i\right], \tag{3.3}$$

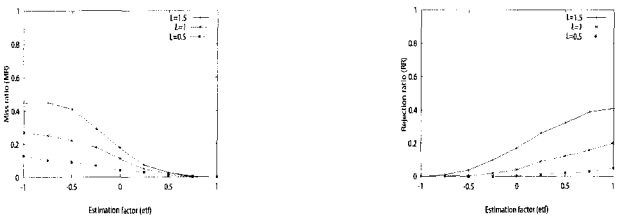
where etf_{kT} is the estimation factor at time instant kT which can have a value in the interval $[-1,\ 1]$. For $etf_{kT} = -1$ the tasks estimated execution time are equal to their best case execution time $BCET$, for $etf_{kT} = 0$ the tasks estimated execution time are equal to their average case execution time $AvCET$, for $etf_{kT} = 1$ the tasks estimated execution time are equal to their worst case execution time $WCET$.

In the case of open-loop scheduling algorithms etf is fixed which is equal to 0 for the algorithms that use the $AvCET$ of tasks as an estimation for their actual execution time, and it is equal to 1 for the algorithms that use the $WCET$ of tasks as an estimation for their actual execution time. In the case of closed-loop scheduling algorithms the estimated factor (etf) is determined using the controller.

3.3.2.3 Open Loop Characteristics

After identifying the measured variables and control output, we analyze the open loop characteristics of the system. From simulation the open-loop characteristic of the real-time task scheduler system was obtained (See figure 3.1).

In figure (3.1), we observe that the relation of the estimation factor,etf with task rejection ratio RR, and deadline miss ratio MR, is not linear for any task load (L). Also the nonlinear relation varies with the system task load. It can be noticed that MR starts from a maximum



(a) MR as etf varies (b) RR as etf varies

Figure 3.1 MR and RR vs. etf

value when $etf = -1$ and reduces quadratically as etf increases. For all task loads that are plotted MR reaches zero when $(etf > 0.75)$. RR starts from approximately zero when $etf = -1$ and increases quadratically to a maximum value as etf increases to 1.

The open loop scheduler is designed for a firm real-time system and works according to the algorithm given in [32]. The open loop algorithm schedules the incoming tasks by taking into account their $AvCET$. The closed loop algorithm schedules the incoming tasks by estimating their EET. The goal of our closed loop scheduler design is to trade off between MR and RR in order to achieve a high value of hit ratio and guarantee ratio. In that case, the MR and RR values will be close to each other and the system will operate in the linear region of the curve near $etf = 0$. Then the corresponding system model in that region is given by,

$$\begin{aligned} MR(z) &= -(mgf(L))etf(z) + mdf(L), \\ RR(z) &= (rgf(L))etf(z) + rdf(L), \end{aligned} \tag{3.4}$$

where mgf and rgf are miss ratio gain factor and rejection ratio gain factor respectively. mdf and rdf are miss ratio disturbance factor and rejection ratio disturbance factor respectively. The values for the mgf, rgf, mdf and rdf varies with, the scheduling algorithm that is used, the system load, and the system parameter (e.g. number of processors). The values of these factors can be found by studying the simulation model.

This linear model approximation may not be accurate since scheduling system typically contains non-linear factors which are not presented in the current simulator. These figures have been generated using the simulation model. For each point in Figures 3.1, the system was simulated for 10,000 tasks. This number of tasks have been chosen to have a 98% confidence interval within ± 0.0017, and ± 0.0022 around each value of MR, and RR, respectively.

3.3.2.4 Limitation on Performance

From equation (3.4) it can be derived that

$$\frac{MR(z)}{mgf(L)} + \frac{RR(z)}{rgf(L)} = \frac{mdf(L)}{mgf(L)} + \frac{rdf(L)}{rgf(L)}. \tag{3.5}$$

Since $MR(z)$ and $RR(z)$ must satisfy equation (3.5) for a given task load, we can not simultaneously make them arbitrarily small. This is a limitation on performance exerted by the system for any task-load. However we can trade off between the miss ratio and rejection ratio by applying feedback control or by designing a closed loop dynamic scheduler and obtain an optimal performance of the scheduler in terms of effective ratio. Note that since the parameters mgf, mdf, rgf and rdf are functions of load L, miss ratio and rejection ratio after trade off remains high or low according to the task-load.

From equation (3.5), it is clear that a low rejection ratio corresponds to a high miss ratio and vice versa. In other words a high guarantee ratio corresponds to a low hit ratio and vice versa. For a stable real-time system both the guarantee ratio and hit ratio should be high or effective ratio should be high. Therefore for closed loop scheduler design we feedback both miss ratio and rejection ration to obtain an high effective ratio for optimal performance.

3.4 Developing Closed-Loop Scheduling Algorithms with Controllers

In this section, we will incorporate control theory into the open loop scheduling algorithm to develop the dynamic closed loop scheduling algorithms. We will design PI and H_∞ controllers to develop the closed loop model.

The closed loop architecture is given in Figure 3.2. The new scheduler is composed of a controller, an execution time estimator and an overlap scheduler. In this approach the task rejection ratio (RR_{kT}) and deadline miss ratio (MR_{kT}) are periodically fed back to the controller after each sample time T. The controller computes the estimate factor etf and calls the execution time estimator to change estimated execution time of tasks. The overlap scheduler schedules the arrived tasks according to their estimated execution time.

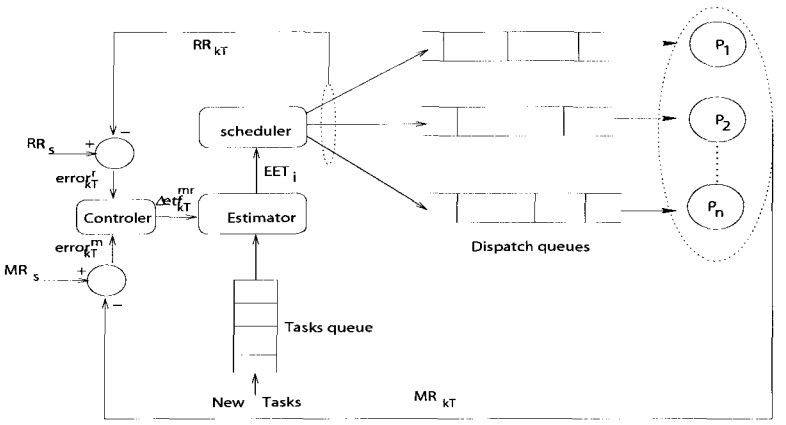


Figure 3.2 Architecture of Closed Loop Algorithms

3.4.1 PI-Controller

Our primary goal behind using a feedback controller is to obtain a trade off between hit ratio and guarantee ratio, by controlling/estimating the execution time of tasks, in order to improve the effective ratio of the system.

In the section we propose an ad hoc PI-Controller based closed loop scheduler algorithm (See figure (3.3)).

The suggested controller in Z-domain is given by,

$$etf(z) = \frac{z}{1-z}\Delta etf(z), \tag{3.6}$$

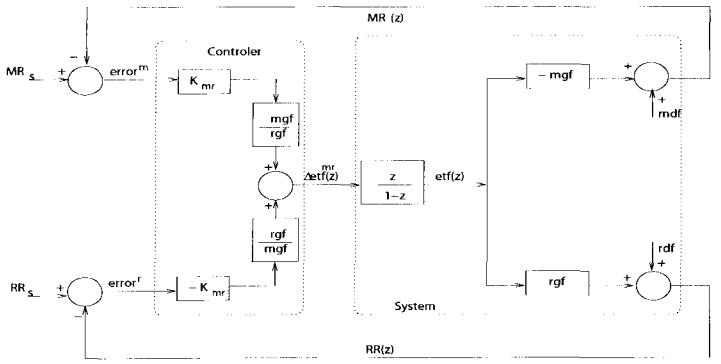


Figure 3.3 Block diagram of the Closed Loop System with the PI-Controller

where

$$\Delta etf(z) = (nf_{m/r})(K_{mr})error^m(z) - (nf_{r/m})(K_{mr})error^r(z). \tag{3.7}$$

In equation (3.7), $nf_{r/m} = \frac{rgf}{mgf}$ and $nf_{m/r} = \frac{mgf}{rgf}$ are normalization factors for task rejection ratio and deadline miss ratio respectively. The idea behind using these normalization factors is that the sensitivity of miss ratio (MR_{kT}) to some variation in the estimated factor (Δetf_{kT}) is different from the sensitivity of rejection ratio (RR_{kT}) to the same variation.

Note that in s-domain the PI controller is given by using bilinear-transforms as,

$$etf(s) = \frac{-\frac{1}{F} - \frac{1}{2}s}{s}\Delta etf(s), \tag{3.8}$$

where F is the sampling frequency of the scheduler.

The control rule is obtained from equation (3.6) and (3.7) by using inverse Z-transform and is given by,

$$\Delta etf_{kT}^{mr} = (nf_{m/r})(K_{mr})error_{kT}^m - (nf_{r/m})(K_{mr})error_{kT}^r etf_{kT} = etf_{(k-1)T} - \Delta etf_{kT}, \tag{3.9}$$

$$etf_{kT} = etf_{(k-1)T} - \Delta etf_{kT}. \tag{3.10}$$

From Figure 3.3, the functions for the deadline miss ratio ($MR(z)$) and the task rejection ratio ($RR(z)$) are given by,

$$MR(z) = \frac{z\,\frac{(RR_s - rdf)\ nf_{m/r}\ G_r + (mdf - MR_s)\ G_m}{G_{mr}\ (1+G_{mr})}}{z - \frac{1}{1+G_{mr}}} + \frac{z\,\frac{MR_s\ G_m + \left((rdf - RR_s)\ nf_{m/r} + mdf\right)\ G_r}{G_{mr}}}{z-1}, \tag{3.11}$$

$$RR(z) = \frac{z\,\frac{(MR_s - mdf)\ nf_{r/m}\ G_m + (rdf - RR_s)\ G_r}{G_{mr}\ (1+G_{mr})}}{z - \frac{1}{1+G_{mr}}} + \frac{z\,\frac{RR_s\ G_r + \left((mdf - MR_s)\ nf_{r/m} + rdf\right)\ G_m}{G_{mr}}}{z-1}. \tag{3.12}$$

where $G_{mr} = G_m + G_r$ with $G_m = (mgf)(K_{mr})(nf_{m/r})$ and $G_r = (rgf)(K_{mr})(nf_{r/m})$. Using Inverse Z-transform, the instantaneous value of task rejection ratio, RR_{kT} and the deadline miss ratio, MR_{kT} in the discrete time domain is given by,

$$MR_{kT} = \begin{cases} 0 & \text{for } kT < 0 \\ \left(\frac{(RR_s - rdf)\ nf_{m/r}\ G_r + (mdf - MR_s)\ G_m}{G_{mr}\ (1+G_{mr})}\right)\left(\frac{1}{1+G_{mr}}\right)^{kT} & \\ \quad + \frac{MR_s\ G_m + \left((rdf - RR_s)\ nf_{m/r} + mdf\right)\ G_r}{G_{mr}} & \text{for } kT \geq 0 \end{cases} \tag{3.13}$$

$$RR_{kT} = \begin{cases} 0 & \text{for } kT < 0 \\ \left(\frac{(MR_s - mdf)\ nf_{r/m}\ G_m + (rdf - RR_s)\ G_r}{G_{mr}\ (1+G_{mr})}\right)\left(\frac{1}{1+G_{mr}}\right)^{kT} & \\ \quad + \frac{RR_s\ G_r + \left((mdf - MR_s)\ nf_{r/m} + rdf\right)\ G_m}{G_{mr}} & \text{for } kT \geq 0 \end{cases} \tag{3.14}$$

From Equations (3.13) and (3.14) we notice that the steady state values for miss ratio and rejection ratio (MR_f and RR_f) are $\frac{MR_s\ G_m + \left((rdf - RR_s)\ nf_{m/r} + mdf\right)\ G_r}{G_{mr}}$ and $\frac{RR_s\ G_r + \left((mdf - MR_s)\ nf_{r/m} + rdf\right)\ G_m}{G_{mr}}$, respectively. Therefore, the steady state effective ratio (ER_f) is equal to:

$$ER_f = \begin{array}{l} \left(1 - \frac{MR_s\ G_m + \left((rdf - RR_s)\ nf_{m/r} + mdf\right)\ G_r}{G_{mr}}\right) \\ \times\left(1 - \frac{RR_s\ G_r + \left((mdf - MR_s)\ nf_{r/m} + rdf\right)\ G_m}{G_{mr}}\right). \end{array} \tag{3.15}$$

If we look at the system given in figure (3.3) as a MIMO-system with two inputs and two

outputs, then the input-output relation is given by

$$\begin{bmatrix} MR'(z) \\ RR'(z) \end{bmatrix} = K^{mr}\frac{z}{1-z}\begin{bmatrix} -\frac{mgf^2}{rgf} & rgf \\ mgf & -\frac{rgf^2}{mgf} \end{bmatrix}\begin{bmatrix} error^m(z) \\ error^r(z) \end{bmatrix}, \tag{3.16}$$

where $MR'(z) = -(mgf)etf(z)$, $RR'(z) = (rgf)etf(z)$, $error^m(z) = MR_s(z) - MR(z)$ and $error^r(z) = RR_s(z) - RR(z)$. The characteristic equation is given by

$$\lambda(\lambda + \frac{mgf^2}{rgf} + \frac{rgf^2}{mgf}) = 0. \tag{3.17}$$

The eigen values are given by

$$\lambda_1 = 0 \; and \; \lambda_2 = (\frac{mgf^2}{rgf} + \frac{rgf^2}{mgf}). \tag{3.18}$$

Therefore the closed loop system is stable when

$$(\frac{mgf^2}{rgf} + \frac{rgf^2}{mgf}) < 1. \tag{3.19}$$

Qualitatively we can say that the closed loop will be unstable when $mgf \ll rgf$ or $rgf \ll mgf$. We observed from simulation that the system is stable for a wide range of task-load variation.

The PI controller is an ad hoc design and it needs tuning of controller gain K^{mr} and sampling period T for optimal performance.

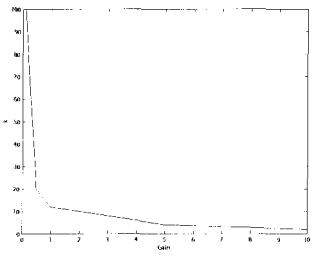


Figure 3.4 Settling time Vs Controller Gain

Note from equation (3.8) that the bandwidth of the system is decided by T. If the sampling period T is low, the bandwidth ω of the system will be high. Thereby the transient response i.e. the settling time of the system (T_s) will be low. But it is not possible to obtain arbitrarily high bandwidth as the system will become unstable when T is smaller than a certain value $T_0 = \max(\theta,\ C)$, where C is the mean computation time of the system and θ is the mean arrival time of tasks. That is because, in the case of feeding back the rejection ratio, if the sample period is less than the mean arrival time of tasks, there is a chance that the number of tasks arrived in the sample interval $[kT,\ (k+1)T]$ is zero or a very small number. This results in feeding back an inaccurate estimation for the rejection ratio. In the case of feeding back the miss ratio, if the sample period is less than the mean computation time of the system , there is a chance that the number of tasks finished in the sample interval $[kT,\ (k+1)T]$ is zero or a very small number. This results in feeding back an inaccurate estimation for the miss ratio.

Apart from the sample period T, we have to tune the controller gain (K^{mr}) in order to obtain an optimal performance by the PI-Controller based closed loop scheduler. From classical control theory, we know that the controller gain K^{mr} decides the transient response of the system. Figure 3.4 shows the plot of no of sample periods the time response (MR_{kT} and RR_{kT}) of the closed loop model (in matlab) took to settle for different controller gains. It shows that a low value of K^{mr} can make the system extremely sluggish.

After tuning we chose $K^{mr} = 0.5$ for our simulations.

3.4.2 H_∞ Controller

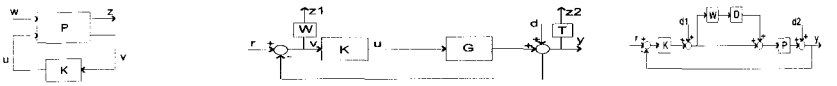
In this section we design a model based closed loop scheduler using a H_∞ controller.

The open loop characteristics of the scheduler system is given by figure (3.1). The corresponding linear model, for a nominal load, is given by

$$\begin{aligned} MR(z) &= -(mgf)etf(z) + mdf, \\ RR(z) &= (rgf)etf(z) + rdf. \end{aligned} \tag{3.20}$$

For H_∞ design, the system block diagram is given in figure (5 b). This closed loop scheduler

gives an optimal performance near the nominal task-load. Note that a change in task-load corresponds to perturbation in the system parameters mgf, rgf, mdf and rdf. Hence a robust performance analysis of the perturbed model (given in figure 5 c) is required to verify the optimal performance of the closed loop scheduler.



(a) (b) (c)

Figure 3.5 (a) Configuration of a H_∞ Controller (b) Block Diagram of the system with the H_∞ Controller where $G = [-mgf,\ 0;\ 0,\ rgf]$, $d = [mdf\ rdf]'$, $y = [MR\ RR]'$, $r = [MR_s\ RR_s]'$ and $u = etf$ (c) Uncertainty model for a closed loop dynamic real-time task scheduler.

The H_∞ controller was designed using the function 'hdfsyn' in matlab. The state space representation of the controller is given by

$$A = \begin{bmatrix} 0.0848 & 0 & 0 & 0 \\ 0 & 0.0848 & 0 & 0 \\ 0.1427 & -0.0856 & 0.3005 & -0.2300 \\ -0.0856 & 0.0514 & -0.2300 & 0.0552 \end{bmatrix}, \quad B = \begin{bmatrix} 1.1949 & 0 \\ 0 & 1.1949 \\ 0.2584 & -0.3768 \\ -0.3768 & -0.1435 \end{bmatrix},$$

$$C = \begin{bmatrix} -0.3861 & 0.2317 & -1.0371 & 0.6223 \end{bmatrix} \text{ and } D = \begin{bmatrix} -1.6992 & 1.0195 \end{bmatrix}.$$

3.5 Model Verification and Performance Evaluation

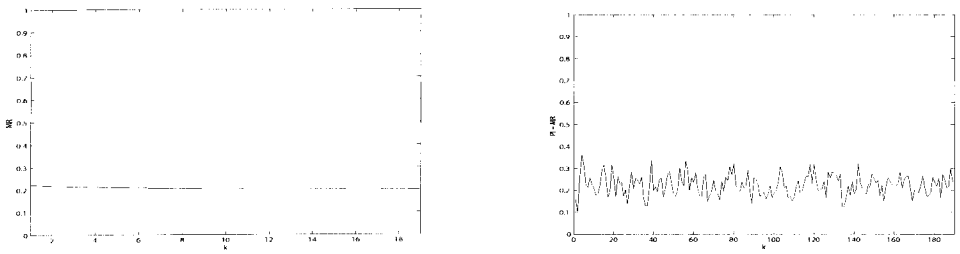
In this section, we have verified the closed-loop scheduling models by comparing the results from the matlab simulink models and the results from the simulator for the real-time scheduler. The simulator software simulates the open loop and closed loop algorithms (PI and H_∞) for the

real-time schedulers for a multiprocessor firm real-time system . The tasks for the simulator are generated as per the rules given below:

- The best-case execution time($BCET_i$) of a task T_i are chosen uniformly between 10 secs and 20 secs.
- The worst-case execution time ($WCET_i$) of a task T_i is chosen to be 4 times its $BCET_i$.
- The average-case execution time ($AvCET_i$) of a task is the average of its $BCET_i$ and $WCET_i$.
- The actual execution time is computed as an uniform random variable in the interval [BCET, WCET]
- The firm deadline of a task T_i is uniformly chosen between $2 + WCET_i$ and $4 * WCET_i$.
- The arrival time of tasks follow exponential distribution with mean θ.
- The task load L is defined as the expected number of task arrivals per mean service time and is approximately equal to C/θ, where C is the mean computation time of the system.

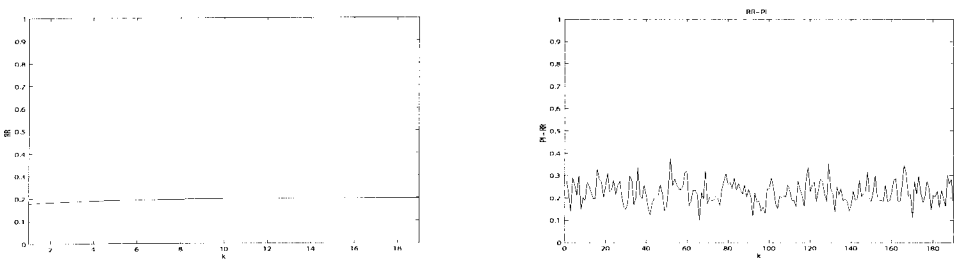
Figures 3.6 to 3.9 show the time response of the closed loop schedulers implemented in the matlab simulink and the real-time simulator. They verify that closed loop scheduler algorithm in the simulator behaves consistently as predicted by the closed loop model in simulink. The instantaneous values of steady state miss ratio (MR_{kT}) and rejection ratio (RR_{kT}) are used as the performance metrics to verify accuracy of the modeling.

Two experiments were carried out to study the performance of the closed-loop scheduling algorithms. In the first experiment (Experiment A), the hit ratio, guarantee ratio and effective ratio of the system were observed from simulation for different task loads (L). In the second experiment (Experiment B), the task load (L) for each simulation run is dynamically changed using a step task load profile. A step task load SL is a task load that instantaneously jumps from a nominal task load L_{nom} to task load L_{max} and stay constant after the jump. The step task load is represented with a tuple $SL(L_{nom}, L_{max})$. In the context of real-time systems, the



(a) From Matlab (b) From Simulator

Figure 3.6 MR_{kT} for the PI-Controller based closed loop algorithm



(a) From Matlab (b) From Simulator

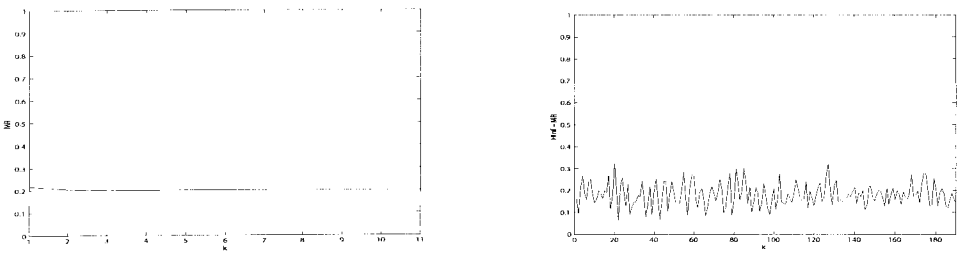
Figure 3.7 RR_{kT} for the PI-Controller based closed loop algorithm

step task load represents the worst-case task load variation. Each point in the performance plots is the average of 20 runs. In each run the system was simulated with 10,000 tasks. This number of runs has been chosen to have a 98% confidence interval within ± 0.013, and ± 0.015, around each value of MR, and RR respectively.

3.5.1 Experiments A: Effect of Task Load on HR, GR and ER

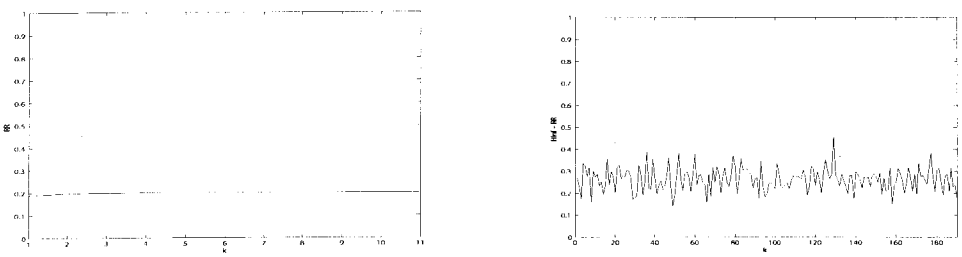
Experiment A studies the behavior of the three scheduling algorithms (*Open loop (OL),PI and* H_∞) for a given task load (L). Figure 3.10 shows the effect of task load on hit ratio, rejection ratio and effective ratio for all the three algorithms.

As shown in the figure, all the three algorithms perform equally well when the system load is less than 80%, with respect to GR, whereas the open loop scheduler's HR decrease with



(a) From Matlab (b) From Simulator

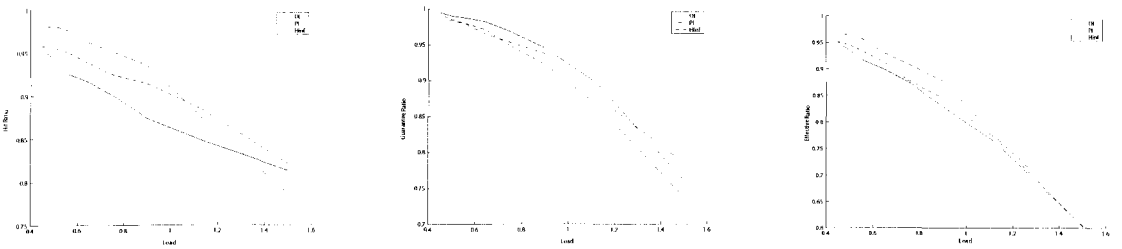
Figure 3.8 MR_{kT} for the H_∞-Controller based closed loop algorithm



(a) From Matlab (b) From Simulator

Figure 3.9 RR_{kT} for the H_∞-Controller based closed loop algorithm

increase in task load L. Further, during overload cases the open loop schedulers gives a poor MR, as it is insensitive to the change in workload conditions, but the closed loop schedulers observe the change in system dynamics by constant feedback of the MR and RR and give a better performance during overload scenarios also. In the context of real-time systems, hit ratio is important and completing all the accepted tasks successfully is highly desirable, which is done efficiently by the closed loop schedulers. Among the closed loop schedulers, the PI controller gives a better HR and GR than the H_∞ controller for all task loads. The PI-controller is an ad hoc design and it can adapt to the system dynamics that changes with the task load. Whereas the H_∞-controller, which is a model based design, performs well near the nominal task load $L = 1$. Figure 3.10 shows the effective ratio under different load conditions for all the three algorithms. It can be seen from the figure that the closed loop schedulers perform



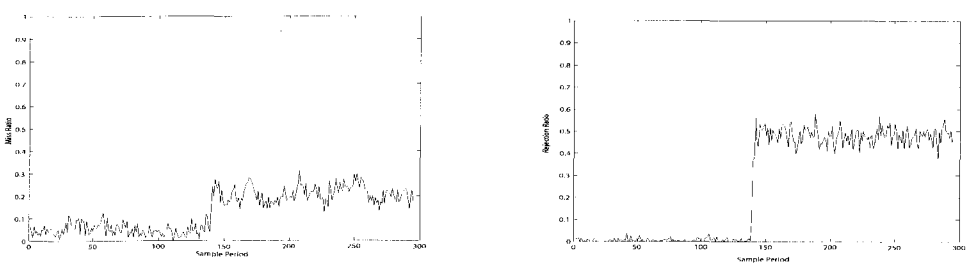
(a) Hit Ratio (b) Guarantee Ratio (c) Effective Ratio

Figure 3.10 Hit Ratio, Guarantee Ratio and Effective Ratio of the open loop and the two closed loop algorithms

better than the open loop schedulers for all load scenarios. This proves the effectiveness of the closed scheduling algorithms developed in this dissertation. Further, among the closed loop schedulers it can be seen that PI controller performs better than the the H_∞ controller over a wide range of task-load.

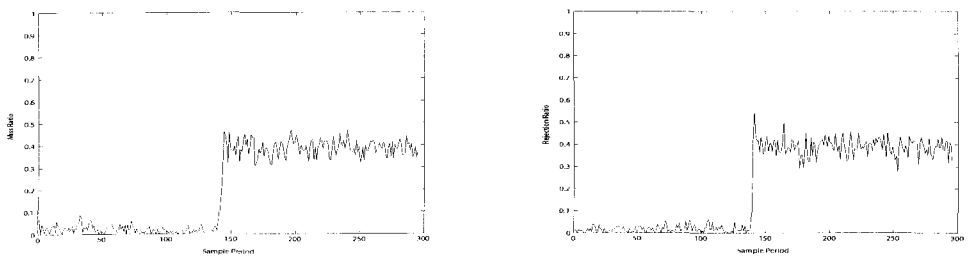
3.5.2 Experiments B: Dynamic Task Load

Experiment B studies the behavior of the three closed loop scheduling algorithms when the task load (L) for each simulation run is dynamically changed using a step task load profile. In this experiment, we change the task load of the system from 0.5 to 1.5 within one simulation run and study the MR_{kT} and RR_{kT} responses for all the three algorithms.



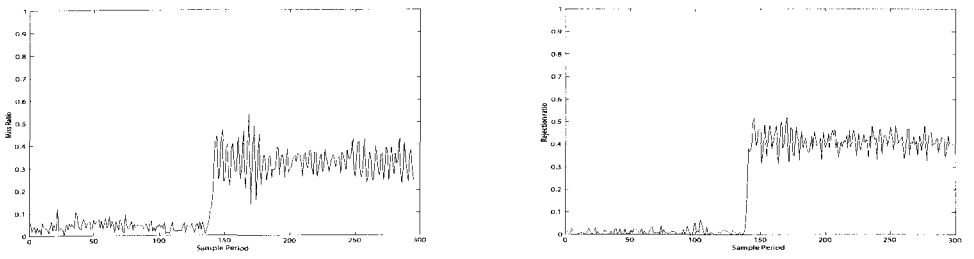
Miss Ratio Rejection Ratio

Figure 3.11 MR_{kT} and RR_{kT} for the open loop algorithm



Miss Ratio Rejection Ratio

Figure 3.12 MR_{kT} and RR_{kT} for the PI-control based closed loop algorithm



Miss Ratio Rejection Ratio

Figure 3.13 MR_{kT} and RR_{kT} for the H_∞ control based closed loop algorithm

Figure 3.11 to 3.13 show the MR_{kT} and the RR_{kT} for the open loop algorithm and the two closed loop algorithms when $MR_s = 0.01\%$, $RR_s = 0.01\%$, $T = 100$ *sec*, and step load profile that jumps form $L = 0.5$ to $L = 1.5$ and $k = 0.5$ (PI).

From simulation we found that at the step of the load for the open loop algorithm, MR_{kT} jumped from almost 0 level to 50% value which can lead to extremely poor performance of the system. The open loop algorithm is insensitive to change in the task load and hence gives a poor guarantee ratio and a relatively low hit ratio. But the closed loop algorithms were able to find a right trade off between the miss ratio and the rejection ratio giving a better performance with respect to guarantee ratio and the hit ratio. Further, when the task load shoots from 0.5 to 1.5, the closed loop algorithms estimate the change in the system parameters and minimize

the miss ratio leading to a more stable performance of the system as compared to their open loop counterpart. Therefore, we find the closed loop scheduling algorithms to be more effective in varying load cases giving more stability and performance to the real-time system.

3.6 Conclusion

In this dissertation, we have viewed the problem of scheduling firm real-time tasks in multiprocessor systems as an input/output system. We have used feedback control theory to design two closed-loop scheduling algorithms, which works based on estimated execution time, estimated by feeding back both the deadline miss ratio and task rejection ratio. We have shown that the proposed closed-loop dynamic scheduling algorithms perform better than the open-loop scheduling algorithm for all the stated performance metrics. Further, we would also like to point out that the open loop scheduling algorithms will perform well only when the tasks execute with average case execution time and will perform poorly when the task's execution times are more or less than its average execution time. However, the closed loop schedulers will react to the change in the task workload and will estimate the task load characteristics resulting in the correct estimation of the task's execution time, resulting in a better system performance under dynamic workload settings. Further, we have shown through our experiments that the closed loop algorithms perform better than the open loop algorithm for a dynamic task load, such as a step task load profile. Thus, we feel that closed loop scheduling algorithms are a proper solution to real-time scheduling, leading to improved performance, especially when the workload cannot be characterized easily.

CHAPTER 4 CONSTRUCTIVE CONTROL OF QUANTUM-MECHANICAL SYSTEMS

In this chapter we develop a method for exact constructive controllability of quantum mechanical systems. The method has three steps, first a path from the initial state to the final state is determined and intermediate points chosen such that any two consecutive points are close, next small sinusoidal control signals are used to drive the system between the points, and finally quantum measurement technique is used to exactly achieve the desired state. The methodology is demonstrated for the control of spin-half particles in a Stern-Gerlach setting.

4.1 Introduction

Control of quantum mechanical systems has taken particular importance due to the recent theoretical and technological advances. It was shown in [51] that quantum mechanical principles can be exploited to solve problems that are intractable if classical means are employed. This has given significant boost to the field of quantum computing where quantum control is needed. At the same time laser technology has sufficiently advanced to prompt the study of control of molecular systems at the quantum regime.

Most of the studies on control of quantum-mechanical systems utilize the description of the dynamics given by Schrodingers equation:

$$i\hbar\frac{d\psi}{dt} = H(t)\psi; \quad \psi(0) = \psi_0, \tag{4.1}$$

where ψ is the state and H represents the Hamiltonian. In many studies it is assumed that the quantum mechanical system can be described in terms of the eigenstates of an observable

with the observable having a discrete spectrum. In such a case, the underlying state space has a finite dimensional basis. In this dissertation we will restrict the study where such an approximation is valid. It can then be shown that the study of control of (4.1) is equivalent to the study of the following:

$$i\hbar\frac{dU}{dt} = H(t)U; \quad U(0) = I, \tag{4.2}$$

where $H = X_0 + \sum_{i=1}^{m} X_i u_i(t)$ with X_i being skew-Hermitian matrices. In this form control of quantum-mechanical systems can borrow results from the control of bilinear systems evolving on Lie groups. Even though tests based on the classical result given in [52] to assess the controllability of a bilinear system can be applied to quantum mechanical systems (see [53]) no result is known on how to synthesize a control law that drives an initial state to a desired final state. Thus the question of constructive controllability remains largely unsolved.

Recent results have been obtained on constructive controllability for specific systems. For example in [54] it was shown that in the case of a single spin $\frac{1}{2}$ particle sinusoidal control authority can be utilized to obtain any desired state in an optimal manner when the system is drift-less.

In [55] averaging techniques were utilized to achieve approximate constructive controllability of drift-less left-invariant bilinear systems evolving on Lie groups. In this method small amplitude sinusoidal signals drive the system state to any desired final condition approximately. Once the state is near the desired state linear feedback strategies are employed to reach the desired state.

For quantum mechanical systems traditional feedback strategies cannot be directly applied.This is due to the quantum mechanical principle that a measurement results in the state collapsing into one of the eigenstates of the observable corresponding to the measurement, thus destroying the state.

In this dissertation we derive the result in [55] by a simple and direct application of a classical result based on two time scale separation. We apply this result to quantum-mechanical systems. Finally we provide control techniques for spin states where a new method is introduced to obtain *exact* constructive controllability of the quantum-mechanical system. We

demonstrate the effectiveness of the proposed strategy for the control of spin $\frac{1}{2}$ particles in a Stern-Gerlach experiment.

The rest of the chapter is organized as follows: section 2 presents mathematical and quantum mechanical preliminaries, section 3 presents the application of two-time scale method to quantum mechanical systems, section 4 describes the Stern-Gerlach experiment to which the developed methodology is applied, and finally we conclude in section 5.

4.2 Preliminaries

In this section we first present relevant quantum mechanical preliminaries outlining the basic axioms of the theory. We then present the needed mathematical preliminaries.

4.2.1 Quantum-mechanical preliminaries

In quantum mechanics it is assumed that the state of system can be described by a Hilbert space $\mathcal{H}$ with the complex numbers as the field. Each element of the Hilbert space is called a *ket* and the Hilbert space is called the *ket* space. Associated with the ket space is the *bra* space which is the dual space of the ket space.

We will assume that a complete orthonormal basis for the ket space is available. Note that the basis corresponds to the states that result when an observation is done. In other words they correspond to the eigenstates of an observable. Let this basis be given by $\{e_i\}_{i\in\alpha}$. One of the axioms of quantum-mechanics is that the probability that a given ket x, is in the eigenstate e_i after a measurement is given by $| < x, e_i > |^2$. Also the total probability that the ket x is in any of the eigenstates e_i after an measurement equals one, that is $\sum_{i\in\alpha} | < x, e_i > |^2 =< x, x >= 1$.

In this dissertation we restrict our study to cases where the ket space is a finite dimensional Hilbert space. Suppose $\psi(t)$ in the ket space describes the quantum mechanical state of a system. Then an axiom of quantum-mechanics is that the evolution of the ket is governed by Schrodingers equation

$$i\hbar\dot{\psi}(t) = H(t)\psi, \tag{4.3}$$

where H is the Hamiltonian operator which maps $\mathcal{H}$ to $\mathcal{H}$. The Hamiltonian can be written as

$H_0 + \sum_{k=0}^{m} u_i(t)H_i(t)$, where the functions $u_i : R \to R$ are the control signals. If we assume that the ket space $\mathcal{H}$ is a n dimensional Hilbert space then it follows that H admits a matrix representation also (with some abuse of notation) denoted by H, where H is Hermitian, and ψ can be represented by a n dimensional vector in R^n also denoted by ψ. The matrix H can be decomposed into $H_0 + \sum_{k=0}^{m} u_i(t)H_i(t)$ where H_0 and H_i are all Hermitian. Associated with Equation (4.3) is the equation for the transition matrix

$$\begin{aligned} \dot{U} &= X_0 U + \sum_{k=1}^{m} u_k X_k U \\ U(t_0) &= I, \end{aligned} \tag{4.4}$$

where $X_0 := \frac{1}{i\hbar} H_0$ and $X_k := \frac{1}{i\hbar} H_k$. Note that X_0 and X_k, $k = 1, \ldots, m$ are skew-Hermitian matrices. If $U(t, t_0, u_i, \ldots, u_m, I)$ denotes the solution of (4.4) then $U(t, t_0, u_i, \ldots, u_m, I)\psi(t_0)$ is the solution of (4.3) with initial condition at time t_0 given by $\psi(t_0)$. It is to be noted that often the Hamiltonian is such that X_0 and X_k can be restricted to traceless skew-Hermitian matrices.

4.2.2 Mathematical preliminaries

Lemma 1 (Wei-Norman) *(see [56]) Let $\{X_i\}_{i=1}^{\ell}$ be a basis for Lie algebra $\mathcal{L}$ with $[X_i, X_j] = \sum_{k=1}^{\ell} \gamma_{ij}^k X_k$, then*

$$\left(\Pi_{j=1}^{r} e^{x_j X_j}\right) X_i \left(\Pi_{j=\ell}^{r} e^{-x_j X_j}\right) = \sum_{k=1}^{\ell} \xi_{ki} X_k, \tag{4.5}$$

where $\xi_{ki}(x_1, x_2, \ldots, x_r)$ is an analytic function of x_j for $j = 1, \ldots, r$.

By expanding the left hand side of (4.5) and comparing coefficients of X_k we can see that

$$\begin{aligned} \xi_{ki}(x) = \delta_{ik} &+ \sum_{m=1}^{r} x_m \gamma_{mi}^k + \sum_{m=1}^{r} \sum_{n=1}^{\ell} \frac{x_m^2}{2} \gamma_{mi}^n \gamma_{mn}^k \\ &+ \sum_{m=1}^{r-1} \sum_{n=m+1}^{r} \sum_{p=1}^{\ell} x_m x_n \gamma_{ni}^p \gamma_{mp}^k + O(x_i x_j x_k). \end{aligned} \tag{4.6}$$

Theorem 1 *(see [56]) Let $A(t) = \sum_{i=1}^{m} u_i(t) X_i$ and $\mathcal{L}$ be the Lie algebra generated by $\{X_i\}_{i=1}^{m}$*

with the basis formed by $\{X_i\}_{i=1}^{\ell}$. *Then there exits* $\epsilon > 0$ *such that for all* t *with* $|t| < \epsilon$ *the solution to*

$$\dot{U} = A(t)U, \quad U(0) = I;$$

is given by

$$U(t) = e^{x_1(t)X_1} e^{x_2(t)X_2} \dots e^{x_\ell(t)X_\ell},$$

where $x(t) := (x_1, x_2, \dots, x_\ell)$, *satisfies the differential equation*

$$\dot{x} = \xi^{-1} u, \tag{4.7}$$

with $u = (u_1 \ u_2 \ \dots \ u_m \ 0 \ 0 \ \dots 0)' \in R^\ell$ *and* ξ *is a matrix with elements* ξ_{ij} *defined by (4.5) with* $r = i - 1$.

We will say $\delta_1(\epsilon) = O(\delta_2(\epsilon))$ for $\epsilon \to 0$ if there exists a constant $k > 0$ and $\epsilon_0 > 0$ such that $|\delta_1(\epsilon)| \leq k|\delta_2(\epsilon)|$ for all $0 < \epsilon < \epsilon_0$. Also, $e(t, \epsilon) = O(\epsilon)$ on a time scale $\frac{1}{\epsilon}$ as $\epsilon \to 0$ if there exists constants L, k and ϵ_0 such that $\sup_{0 \leq t \leq \frac{L}{\epsilon}} |e(t, \epsilon)| \leq k\epsilon$, for all $0 < \epsilon < \epsilon_0$.

The following result is based on two-variable expansion procedure [57].

Theorem 2 *(see [58]) Consider the initial value problem*

$$\dot{x} = \epsilon f(t, x), \ x(0) = x_0;$$

with f, $\frac{\partial f}{\partial x}$ *and* $\frac{\partial^2 f}{\partial x^2}$ *defined, continuous and bounded in* $D \times [0, \infty)$, $D \subset R^n$. *Let* $\tau := \epsilon t$ *and* $y(\tau)$ *be the solution of*

$$\frac{dy}{d\tau} = f_{av}(y); \ y(0) = x_0 + O(\epsilon^2).$$

Let $x_1 = u^1(t, y(\tau)) + z(\tau)$, *with*

$$u^1(t, y(\tau)) = \int_0^t [-\frac{dy}{d\tau} + f(s, y(\tau))] ds,$$

and $z(\tau)$ is the solution to

$$\frac{dz}{d\tau} = \frac{\partial f_{av}}{\partial y} z + \frac{1}{T}\int_0^T [\frac{-\partial u^1}{\partial \tau} + \frac{\partial f}{\partial y} u^1] dt, \; z(0) = 0$$

If $y(\tau)$ belongs to the interior of D then

$$x(t) = y(\tau) + \epsilon x_1(t, y(\tau)) + O(\epsilon^2)$$

on the time scale $\frac{1}{\epsilon}$.

Corollary 1 *Let $v(t)$ be periodic with period T and $v_{av} := \frac{1}{T}\int_0^T v(t)dt = 0$. Define $\tilde{v}(t) = \int_0^t v(\tau)d\tau$. Then the solution to the following dynamic system*

$$\dot{x} = \epsilon M(x) v(t); \; x(0) = x_0. \tag{4.8}$$

is given by,

$$x(t) = x_0 + \epsilon M(x_0)\tilde{v}(t) + \epsilon^2 t \sum_{i=1}^{\ell} G^i(x_0)\alpha^i + O(\epsilon^2)$$

on the time scale $\frac{1}{\epsilon}$, where $G^i(x) = \frac{\partial M_i(x)}{\partial x} M(x)$ with $M_i(x)$ denoting the i^{th} column of $M(x)$ and $\alpha^i = \frac{1}{T}\int_0^T v_i(t)\tilde{v}(t)dt$. Furthermore if $x_0 = O(\epsilon^2)$ then

$$x(t) = \epsilon\tilde{v}(t) + \epsilon^2 t \sum_{i=1}^{\ell} G^i(0)\alpha^i + O(\epsilon^2) \tag{4.9}$$

on the time scale $\frac{1}{\epsilon}$.

A drift-less quantum-mechanical system is described by equation (4.4) with $X_0 = 0$. If we assume that $\{X_k\}_{k=1}^m$ are independent such that $\{X_k\}_{k=1}^m$ can be extended to $\{X_k\}_{k=1}^\ell$ which forms the basis for the Lie algebra containing $X_1, X_2, \ldots, X_m$, we can write equation (4.4) as

$$\dot{U}(t) = \sum_{k=1}^{\ell} u_k(t) X_k U(t), \; U(0) = I. \tag{4.10}$$

with the understanding that $u_k = 0$ if $k > m$.

It is clear from Theorem 1 that for $|t| \leq \epsilon$ where ϵ is some positive number, the solution to (4.10) is given by

$$U(t) = \Pi_{k=1}^{\ell} e^{x_k(t)X_k}, \tag{4.11}$$

where $x := (x_1, x_2, \ldots, x_\ell)$ satisfies the differential equation (4.7). If we assume that the controls are small and periodic with period T that is $u(t) = \epsilon v(t)$ with $v(t)$ being a periodic signal with period T, from Corollary 1 we can write

$$x_n(t) = \epsilon \tilde{v}_n(t) + \epsilon^2 t \sum_{i=1}^{\ell} \sum_{j=1}^{\ell} \gamma_{ji}^n \alpha_j^i + O(\epsilon^2) \tag{4.12}$$

on the time scale $\frac{1}{\epsilon}$, with $x = (x_1, x_2, \ldots, x_\ell)'$, $\tilde{v} = (\tilde{v}_1, \tilde{v}_2, \ldots, \tilde{v}_\ell)'$ and $\alpha^i = (\alpha_1^i, \alpha_2^i, \ldots, \alpha_\ell^i)$. Note that $M(x) := \xi^{-1}(x)$ is defined by (4.5) with $r = i - 1$ and from equations (4.6) and (4.7) $\frac{\partial M_{ni}(x)}{\partial x_j} = -\frac{\partial M_{ni}(x)}{\partial x_j} = \gamma_{ji}^n$ and $G^i(0) = \frac{\partial M_i(x)}{\partial x}\Big|_{x=0} M(0) = \frac{\partial M_i(x)}{\partial x}\Big|_{x=0}$.

Note that the above result agrees with the result obtained in [55] where the averaging technique was modified to obtain the same result. The higher order approximations derived in [55] can be directly derived using the study in [59].

4.3 Small amplitude sinusoidal control

In this section we present the methodology to obtain exact controllability for quantum mechanical system. That is we develop a technique to drive an initial quantum-mechanical state ψ^0 to a desired final state ψ^d. The basic steps involved are given below.

Step 1 Identify a path that connects ψ^0 to ψ^d in the state-space.

Step 2 On the path, choose intermediate points, ψ^1, ψ^2, ψ^3, ...,ψ^N such that the distance between the consecutive points is small.

Step 3 Construct control strategies to drive system state from ψ^{i-1} to ψ^i. When this control is applied to the quantum-mechanical system, it will result in a final state ψ^f which will be close to ψ^i.

Step 4 Perform a measurement which collapses the state ψ^f to ψ^i.

Step 5 Repeat the steps 3 and 4, N times, until state ψ^d is reached.

The procedure in Step 3 is done first by finding the unitary matrix U^i such that $\psi^i = U^i\psi^{i-1}$ and U^i is within $O(\epsilon)$ of I. Then devise the control law to drive (4.4) from $U(0) = I$ to U^f which is $O(\epsilon^2)$ close to U^i over a time scale of $\frac{1}{\epsilon}$. When this control is applied to the quantum-mechanical system, it will result in a final state ψ^f which will be close to ψ^i.

For the control of quantum-mechanical systems in the above methodology two questions remain to be answered:

- Given an arbitrary vector x^d which is in a $O(\epsilon)$ neighborhood of the origin, what is the control law $u(t) = \epsilon v(t)$, with $v(t)$ periodic with zero average, a time t_f such that $x(t_f) = x^d + O(\epsilon^2)$.

- Is there a mechanism by which quantum mechanical state can be made to reach a state x^d exactly. That is, is there a way to eliminate the error which is $O(\epsilon^2)$.

The first question is analyzed in [55]. For the second question we propose that if x^d happens to be an eigenstate of an observable then performing a measurement corresponding to the observable after $x(t_f)$ is reached (which is close to x^d) will collapse $x(t_f)$ onto x^d with a very high probability. The subsequent question is, whether every state x^d, an eigen state of some observable. We will answer these queries for the Stern-Gerlach setup for the control of a spin-$\frac{1}{2}$ particle.

4.4 Control of the spin-$\frac{1}{2}$ particle and the Stern-Gerlach experiment

The Stern-Gerlach apparatus consists of a beam of neutral paramagnetic particles (for example silver atoms) in a highly inhomogeneous magnetic field (see Figure 4.1) in a particular direction associated with the apparatus. Under the influence of the magnetic field generated by a magnet with a pointed pole tip, it is seen that the beam splits into two parts depending upon the spin of the particle. The particles deflected in the direction of the magnetic field are said to be in the plus state (spin up) state with respect to Stern-Gerlach apparatus S and the particles deflected down are in minus state (spin down) with respect to S. The plus state

and the minus states of S are denoted by $|+S>$ and $|-S>$ respectively. These states

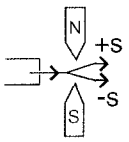


Figure 4.1 +S particles are deflected up and -S particles are deflected down while passing through the magnetic field

would correspond to quantized values of the magnetic spin for the particles. In a modified Stern-Gerlach apparatus, either of the base states ($|+S>$ or $|-S>$) can be blocked and the other base state can be extracted. Consider any other Stern-Gerlach apparatus T which has a magnetic field in a direction different from that of the S device. Then the probability amplitude of $|iS>$ state being in the $|jT>$ state (where i, j can be $+$ or $-$) is denoted by $<jT|iS>$. The probability that a $|iS>$ state results in a $|jT>$ state is given by the square of the magnitude of the probability amplitude $<jT|iS>$. It can be shown that if $<\chi|iS>$

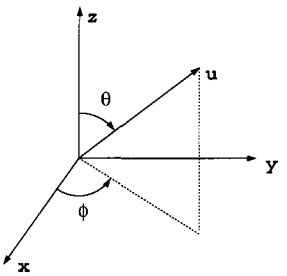


Figure 4.2 The S device has the magnetic field in the direction z whereas the magnetic field of the device T is in the direction u.

is the probability amplitude of the state $|iS>$ being in the state $|\chi>$ after a particular measurement then the corresponding probability of a pure state of a device T whose magnetic field is oriented with respect to the magnetic field of S device as shown in Figure 4.2 is given by

$$\begin{bmatrix} \chi|+T \\ \chi|-T \end{bmatrix} = \begin{bmatrix} e^{-i\frac{\phi}{2}}\cos\frac{\theta}{2} & e^{i\frac{\phi}{2}}\sin\frac{\theta}{2} \\ -e^{-i\frac{\phi}{2}}\sin\frac{\theta}{2} & e^{i\frac{\phi}{2}}\cos\frac{\theta}{2} \end{bmatrix} \begin{bmatrix} \chi|+S \\ \chi|-S \end{bmatrix}$$

This can be deduced based on the symmetry of the space (see [60]). This implies

$$|+T>=\cos\frac{\theta}{2}e^{-i\frac{\phi}{2}}|+S>+\sin\frac{\theta}{2}e^{i\frac{\phi}{2}}|-S>,$$

with χ as any arbitrary state. The direction of the magnetic field of a device whose plus state corresponds to $a|+S>+b|-S>$, where a and b are complex numbers with $|a|^2+|b|^2=1$, is given by,

$$\phi = Arg(b) - Arg(a),\ \theta = arctan(|b|/|a|) \tag{4.13}$$

With this chosen ϕ and θ, we get

$$|+T>=e^{i\beta}(a|+S>+b|-S>),$$

where β is a real number. As $a|+S>+b|-S>$ differs from $|+T>$ by a constant phase factor $e^{i\beta}$ they are the same state physically and thus we can always construct an arbitrary superposition of the pure states of a particular device to correspond to the plus state of some other device. Thus Step 4 in the outline of the method presented in the previous section can be accomplished by performing a measurement which has the magnetic field in the appropriate direction given by equation (4.13).

Consider any particle subjected to an arbitrary magnetic field $\vec{B}$. The vector $\psi(t)$ of probability amplitudes corresponding to the particle being in the $|+S>$ and $|-S>$ satisfies the Schrodinger's equation (4.3) with $H(t)=-\mu\sum_{k=1}^{3}\sigma_k B_k$, where the direction of the magnetic field in the S device is e_3, $\vec{B}=B_1\vec{e}_1+B_2\vec{e}_2+B_3\vec{e}_3$, and σ_i are the Pauli matrices (see [60]). The associated differential equation for the transition matrix is given by following differential equation:

$$\dot{U}=-\frac{\gamma}{2}\sum_{k=1}^{3}u_k(t)X_kU, \tag{4.14}$$

where $X_1=[0,-i;-i,0]$, $X_2=[0,-1;1,0]$ and $X_3=[-i,0;0,i]$. Note that we have identified subscripts $1,2$ and 3 with x,y and z directions where the magnetic field in the Stern-Gerlach experiment is in the z direction. Also, the Pauli matrices are given by $\frac{1}{i}X_i$. $[X_i,X_j]=2\epsilon_{ijk}X_k$,

where $\epsilon_{ijk} = 1$ if ijk is equal to 123, 231 or 312. It is -1 otherwise. X_1, X_2 and X_3 form a basis for the Lie Algebra of traceless skew-Hermitian matrices, $su(2)$. We are interested in solving the following problem;

$$\begin{aligned} \dot{U} &= -\frac{\gamma}{2}\sum_{i=1}^{3} u_i(t)X_iU \\ U(0) &= I,\ U(t_f) = U_d. \end{aligned} \tag{4.15}$$

We will now determine the matrix ξ given in (4.5) for the spin-$\frac{1}{2}$ system. Using the Baker-Hausdorf formula we can get that

$$e^{x_i(t)X_i}X_j e^{-x_i(t)X_i} = \cos 2x_i(t)X_j + \epsilon_{ijk}\sin 2x_i(t)X_k$$

Then using Wei-Norman equation (4.5) we can get that

$$\xi = \begin{bmatrix} 1 & 0 & \sin 2x_2 \\ 0 & \cos 2x_1 & -\cos 2x_2 \sin 2x_1 \\ 0 & \sin 2x_1 & \cos 2x_2 \cos 2x_1 \end{bmatrix},$$

which implies that

$$M = \begin{bmatrix} 1 & \tan 2x_2 \sin 2x_1 & -\tan 2x_2 \cos 2x_1 \\ 0 & \cos 2x_1 & \sin 2x_1 \\ 0 & -\frac{\sin 2x_1}{\cos 2x_2} & \frac{cos2x_1}{\cos 2x_2} \end{bmatrix} \tag{4.16}$$

Lets assume that we know $x^f = O(\epsilon)$ and that we want to drive $x = 0$ to $x = x^f$. We can get from Corollary 1 that $x(t_f) = \epsilon\tilde{v}(t) + \epsilon^2 t_f \sum_{i=1}^{3} G_i(0)\alpha_i + O(\epsilon^2)$ on the time scale $\frac{1}{\epsilon}$, where $G_1(0) = [0\ 0\ 0;\ 0\ 0\ 2;\ 0\ -2\ 0]$, $G_2(0) = [0\ 0\ -2;\ 0\ 0\ 0;\ 0\ 0\ 0]$ and $G_3(0) = 0$. If we assume that $v_1(t) = 0$, $v_2(t) = s_2\cos(\omega t)$ and $v_3(t) = s_3\cos(\omega t + \phi)$, then we can derive that $\tilde{v}(t) = \frac{1}{\omega}[0,\ s_2\sin(\omega t),\ s_3(\sin(\omega t + \phi) - \sin(\phi))]'$ and $\alpha^1 = [0,\ 0,\ 0]'$, $\alpha^2 = -\frac{1}{2\omega}[0,\ 0,\ 2s_2s_3\sin(\phi)]'$

and $\alpha^3 = -\frac{1}{2\omega}[0,\ -2s_2s_3\sin(\phi),\ 0]'$. Thus we have from equation (4.8)

$$x(t_f) = \begin{bmatrix} \frac{\epsilon^2 t_f}{2\omega} 2\sin(\phi) \\ \frac{\epsilon}{\omega}\sin(\omega t_f) \\ \frac{\epsilon}{\omega}(\sin(\omega t_f + \phi) - \sin(\phi) \end{bmatrix} + O(\epsilon^2) \tag{4.17}$$

4.4.1 Illustrative example

We now provide an example of the proposed methodology. We will assume that the state of the spin-$\frac{1}{2}$ particle is described in terms of the basis $|+S>$ and $|-S>$. In this basis we assume that the initial condition is given by $\psi^o = [0.6,\ 0.8]'$. The desired state vector is $\psi^d = [0.5235 - 0.2640i,\ 0.8075 + 0.0652i]'$. It can be verified that $\psi^d = U^d\psi_0$ where,

$$U^d = \begin{bmatrix} 0.9601 - 0.2105i & -0.0657 - 0.1721i \\ 0.0657 - 0.1721i & 0.9601 + 0.2105i \end{bmatrix}.$$

Evaluating the corresponding Wei-Norman parameters we see that $U^d = e^{x_1^d X_1} e^{x_2^d X_2} e^{x_3^d X_3}$, where $x^d = [0.0785,\ 0.0500,\ 0.1000]'$.

We will assume that x^d is in an $O(\epsilon)$ neighborhood of 0. Thus we need to find small sinusoidal control signals $v_i(t)$ to drive $\dot{x} = \epsilon M(x)v$ from the origin to x^d where $M(x)$ is given by (4.16). We will further impose the condition that $v_1(t) = 0$. From (4.17) we can evaluate the state with ($O(\epsilon^2)$ error) to which the system evolves when small sinusoidal control are used. Thus we can use (4.17) to reach x^d with $O(\epsilon^2)$ error. An algorithm is given in [55]) to achieve this task. We will now devise the control using the algorithm given in [55]. Let $q = \frac{x_1^d}{\pi x_3^d x_2^d \tan(\psi)}$, where $0 < \psi < \pi/2$ is selected such that q is a positive integer. Define $T = \frac{t_f}{q+1}$, $t_1 = \frac{T}{4}$, $t_2 = t_1 + qT$, $t_3 = t_2 + \frac{T}{4}$ and $t_4 = t_f = t_3 + \frac{T}{2}$. We will then construct the

controls as,

$$
\begin{aligned}
\epsilon v_1(t) &= 0 \\
\epsilon v_2(t) &= \begin{cases} x_2^d \omega \sin(\omega t) & 0 \le t \le t_1 \\ \frac{x_2^d \omega}{\cos(\psi)} \cos(\omega(t-t_1) - \psi) & t_1 < t \le t_2 \\ x_2^d \omega \cos(\omega(t-t_2)) & t_2 < t \le t_3 \\ \frac{1}{2} x_2^d \omega \sin(\omega(t-t_3)) & t_3 < t \le t_4 \end{cases} \\
\epsilon v_3(t) &= \begin{cases} x_3^d \omega \sin(\omega t) & 0 \le t \le t_3 \\ -\frac{1}{2} x_3^d \omega \sin(\omega(t-t_3)) & t_3 < t \le t_4 \end{cases}
\end{aligned}
$$

We have chosen $\epsilon = 0.1$ for small amplitude control and $T = 4$ for small frequency control and then $\omega = \pi/2$, $t_f = 24$ and $\phi = \pi/4$.The time plots of the control inputs $v_2(t)$ and $v_3(t)$ are given in the Figure 4.3. The time plots of the Wei-Norman Parameters (as obtained by

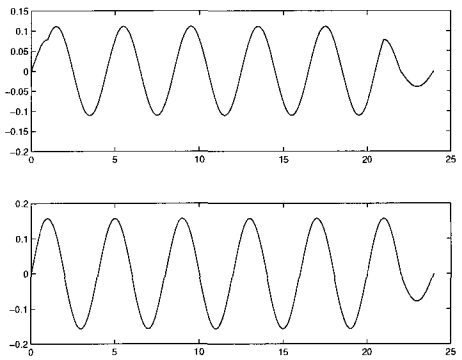


Figure 4.3 Control Input Signals v_2(top) and v_3(bottom)

solving the differential equation $\dot{x} = \epsilon M(x)v(t)$ numerically) $x_1(t)$, $x_2(t)$, and $x_3(t)$) are given in the Figure 4.4. It is seen that the numerical solution $x(t)$ evaluated at t_f is such that

[width=2.5in]qmfigs/x2

Figure 4.4 Actual (solid lines) and Average (dashed lines) plots of the Wei-Norman Parameters x_1(top), x_2(middle) and x_3(bottom)

$x(t_f) = [0.0756,\ 0.0500,\ 0.1039]'$. The matrix $U(t_f)$ where $U(t)$ is a solution to (4.15) is given by

$$U(t_f) = \begin{bmatrix} 0.9594 - 0.2176i & -0.0657 - 0.1670i \\ 0.0657 - 0.1670i & 0.9594 + 0.2176i \end{bmatrix}.$$

Thus the numerical solution of (4.8) yields $\psi(t_f) = U(t_f)\psi(t_0) = [0.5231 - 0.2641i,\ 0.8069 + 0.0739i]'$.

Using the relationship (4.13) we can construct a Stern-Gerlach setup with its magneitic field aligned in such a manner that the plus state of the device corresponds to the vector ψ^d. When a measurement is performed on the state $\psi(t_f)$ the probability that we will obtain the plus state is given by $| < \psi^d, \psi(t_f) > |^2 = 0.98$. Thus with the proposed method we can reach ψ^d state with a probability of 0.98.

4.5 Conclusions

It is evident that the method developed offers a number of advantages the primary one being that we obtain exact controllability. The result on the construction of small sinusoidal control has no restriction on the dimension of the quantum-mechanical state space. The notion of using measurements to obtain the state desired with very high probability has been demonstrated for two state systems. However, it holds considerable promise for generalization to higher dimensional state space. This is the topic of future research.

4.6 Acknowledgements

We appreciate the discussions with Prof. D. D'Allesandro on quantum-mechanical feedback. This work was supported by the CAREER, NSF grant ECS-9733802.

BIBLIOGRAPHY

[1] G. Binnig, C.F. Quate, and C. Gerber. Atomic force microscope. *Physical Review Letters*, 56, no.9:pp. 930–3., 1986.

[2] National Nanotechnology Initiative. The initiative and its implementation plan. *http://www.nano.gov/nni2.pdf*, July 2000.

[3] J. H. Hoh, R. Lal, S. A. John, J. P. Revel, et al. Atomic force microscopy and dissection of gap junctions. *Science*, 253:1405–1408, 1991.

[4] S. Kasas, N. H. Thomson, B. L. Smith, H. G. Hansma, and others. Escherichia coli rna polymerase activity observed using atomic force microscopy. *Biochemistry*, 36(3):461–468, 1997.

[5] D. Rugar, C. S. Yannoni, and J. A. Sidles. Mechanical detection of magnetic resonance. *Nature*, 360:563–566, (1992).

[6] J. A. Sidles. Noninductive detection of single proton-magnetic resonance. *Applied Physics Letters*, 58(24):2854–2856, 1991.

[7] J. A. Sidles. Folded stern-gerlach experiment as a means for detecting nuclear magnetic resonance of individual nuclei. *Physical Review Letters*, 68:1124–1127, 1992.

[8] J. A. Sidles, J. L. Garbini, and G. P. Drobny. The theory of oscillator-coupled magnetic resonance with potential applications to molecular imaging. *Review of Scientific Instruments*, 63:3881–3899, 1992.

[9] D. R. Sahoo, A. Sebastian, and M. V. Salapaka. Transient-signal-based sample-detection in atomic force microscopy. *Applied Physics Letters*, 83(26):5521–5523, December 29,2003.

[10] Mamin, H. J. and Ried, R. P. and Terris, B. D. and Rugar, D. High-density data storage based on the atomic force microscope. *Proceedings of the IEEE*, 87(6):1014–1027, June 1999.

[11] M. V. Salapaka, H. S. Bergh, J. Lai, A. Majumdar, and E. McFarland. Multimode noise analysis of cantilevers for scanning probe microscopy. *Journal of Applied Physics*, 81(6):2480–2487, 1997.

[12] J. N. Israelachvili. *Intermolecular and Surface Forces.* Academic Press, 1985.

[13] A. Sebastian, M. V. Salapaka, D. J. Chen, and J. P. Cleveland. Harmonic analysis based modeling of tapping-mode afm. In *Proceedings of the American Control Conference*, California, June 1999. San Diego.

[14] A. Sebastian, M. V. Salapaka, D. J. Chen, and J. P. Cleveland. Harmonic and power balance tools for tapping-mode atomic force microscope. *Journal of Applied Physics*, 89, no.11:6473–6480, 2001.

[15] Chi-Tsong Chen. *Linear System Theory and Design.* Oxford University Press, 1999.

[16] M. V. Salapaka, D. Chen, and J. P. Cleveland. Linearity of amplitude and phase in tapping-mode atomic force microscopy. *Physical Review B.*, 61, no. 2:pp. 1106–1115, Jan 2000.

[17] Ali H. Sayed Thomas Kailath and Babak Hassibi. Linear Estimation. *Prentice Hall, NJ*, 2000.

[18] Alan S. Willsky and Harold L. Jones. A Generalized Likelihood Ratio Approach to the Detection and Estimation of Jumps in Linear Systems. *IEEE Transactions on Automatic Control, pp. 108-112*, Feb. 1976.

[19] H. Vincent Poor. An Introduction to Signal Detection and Estimation, 2nd Ed. *Springer-Verlag*, 1994.

[20] D. B. Owen. Handbook of Statistical Tables. *Addision Wesley, Reading, MA*, 1922.

[21] T. Sulchek, R. Hsieh, J.D. Adams, G.G. Yaralioglu, S.C. Minne, C.F. Quate, J.P. Cleveland, A. Atalar, and D.M. Adderton. High-speed tapping mode imaging with active q control for atomic force microscopy. *Applied Physics Letters*, 76(11):1473, 2000.

[22] J. Tamayo, A.D.L. Humphris, and M.J. Miles. Piconewton regime dynamic force microscopy in liquid. *Applied Physics Letters*, 77(4):582, 2000.

[23] F. Ohnesorge and G. Binnig. True atomic resolution via repulsive and attractive forces. *Science*, 260:1451, 1993.

[24] P.K. Hansma, J.P. Cleveland, M. Radmacher, D.A. Walters, P.E. Hilner, M. Benzanilla, M. Fritz, D. Vie, H.G. Hansma, C.B. Prater, J. Massie, L. Fukunga, J. Gurley, and V. Eilings. Tapping mode atomic force microscopy in liquids. *Applied Physics Letters*, 64:1738, 1994.

[25] C.A.J. Putman, K.O. Van der Werf, B.G. De Grooth, N.F. Van Hulst, and J. Greeve. Tapping mode atomic force microscopy in liquid. *Applied Physics Letters*, 64:2454, 1994.

[26] Rainer D. Jäggi, Alfredo Franco-Obregòn, Paul Studerus, and Klaus Ensslin. Detailed analysis of forces influencing lateral resolution for q-control and tapping mode. *Applied Physics Letters*, 79:135–137, 2001.

[27] Tomás R. Rodríguez and Ricardo García. Theory of q control in atomic force microscopy. *Applied Physics Letters*, 82(26):4821, 2003.

[28] K. Gustafsson. Control theoretic techniques for stepsize selection in explicit runge-kutta methods. *ACM Transactions on Mathematical Software*, 17(4):533 – 554, 1991.

[29] Krithi Ramamritham and John A. Stankovic. Scheduling algorithms and operating systems support for real-time systems. *Proceedings of the IEEE*, 82(1):55 – 67, 1994.

[30] C. L. Liu and J. W. Layland. Scheduling algorithms for multiprogramming in a hard real-time environment. *Journal of the ACM*, 20(1):40 – 61, 1973.

[31] J. A. Stankovic and Ramamritham K. The spring kernel: A new paradigm for real-time operating systems. *ACM SIGOPS, Operating Systems Review*, 23(2):77 – 83, Jan 1995.

[32] Ra'ed Mohammad S. Al-Omari. Controlling schedulability-realiability trade-offs in real-time systems. *Ph.D thesis, Iowa State University*, 2001.

[33] Giorgio Buttazzo and Luca Abeni. Adaptive rate control through elastic scheduling. *Proceedings of the 39th IEEE Conference on Decision and Control*, 1(2):184 – 194, Dec 2000.

[34] K.-E. *A°*rzén, B. Bernhardsson, J. Eker, A. Cervin, K. Nilsson, P. Persson, and L. Sha. Integrated control and scheduling. *Internal report TFRT-7582, Department of Automatic Control, Lund University*, Aug 1999.

[35] Patrik Persson and Górel Hedin. Interactive execution time predictions using reference attributed grammars. *Proceedings of WAGA '99: Second Workshop on Attribute Grammars and their Applications, Amsterdam, The Netherlands*, Mar 1999.

[36] P. R. Blevins and C. V. Ramamoorthy. Aspects of a dynamically adaptive operating systems. *IEEE Transactions on Computers*, 25(7):713 – 725, Jul 1976.

[37] et. al. David C. Steere. A feedback-driven proportion allocator for real-rate scheduling. *Operating Systems Design and Implementation (OSDI)*, Feb 1999.

[38] M. Livny J. R. Haritsa and M. J. Carey. Earliest deadline scheduling for real-time database systems. *Proceedings of the IEEE Real Time System Symposium (RTSS)*, pages 232 – 242, 1991.

[39] T. F. Abdelzaher and C. Lu. Modeling and performance control of internet servers. 39^{th} *IEEE Conference on Decision and Control*, Dec 2000.

[40] T. F. Abdelzaher and K. G. Shin. End-host architecture for qos-adaptive communication. *IEEE Real-Time Technology and Applications Symposium*, Jun 1998.

[41] Klara Nahrstedt Baochun Li. A control theoretical model for quality of service adaptations. *IEEE International Workshop on Quality of Service*, May 1998.

[42] et. al. D. Seto. On task schedulability in real-time control systems. *IEEE Real-Time Systems Symposium*, Dec 1996.

[43] M. Ryu and S. Hong. Toward automatic synthesis of schedulable real-time controllers. *Integrated Computer-Aided Engineering*, 5(3):261 – 277, 1998.

[44] X. Liu, L. Sha, M. Caccamo, and G. Buttazzo. Online control optimization using load driven scheduling. *Proceedings of the* 39^{th} *IEEE Conference on Decision and Control*, Dec 2000.

[45] M. Caccamo, G. Buttazzo, and L. Sha. Elastic feedback control. *IEEE Proceedings of the* 12^{th} *Euromicro Conference on Real-Time Systems*, pages 121 – 128, Jun 2000.

[46] T. F. Abdelzaher, E. M. Atkins, and K. G. Shin. Qos negotiation in real-time systems and its application to automated flight control. *IEEE Real-Time Technology and Applications Symposium*, Jun 1997.

[47] K. G. Shin and C. L. Meissner. Adaptation and graceful degradation of control system performance by task reallocation and period adjustment. *EuroMicro Conference on Real-Time Systems*, Jun 1999.

[48] Giorgio Buttazzo, Giuseppe Lipari, and Luca Abeni. Elastic task model for adaptive rate control. *Proceedings of the IEEE Real-Time Systems Symposium*, Dec 1998.

[49] J. A. Stankovic, C Lu, S. SOn, and Tao G. The case for feedback control real-time scheduling. *EuroMicro Conference on Real-Time Systems*, pages 11 – 20, Jun 1999.

[50] J. A. Stankovic, C Lu, S. Son, and Tao G. Design and evaluation of feedback control edf scheduling algorithm. *Real-Time Systems Symposium (RTSS)*, pages 56 – 67, Dec 1999.

[51] P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *Proceedings, 35th annual symposium on foundations of computer science*, 1994.

[52] R. W. Brockett. System theory on group manifolds and coset spaces. *SIAM Journal on Control*, 10, no. 2:pp. 265–284, 1972.

[53] V. Ramakrishna, M. V. Salapaka, M. Dahleh, H. Rabitz, and A. Pierce. Controllability of molecular systems. *Physical Review A.*, 51, no. 2:pp. 960–966, Feb 1995.

[54] D. D'Alessandro and M. Dahleh. Optimal control of two level quantum systems. In *Proceedings of the American Control Conference.* Chicago, Illinois, June 2000.

[55] N. E. Leonard and Krishnaprasad. Motion control of drift-free left-invariant systems on lie groups. *IEEE Transactions on Automatic Control*, 40:pp. 1539–54, 1995.

[56] J. Wei and E. Norman. On the global representations of the solutions of linear differential equations as a product of exponentials. *Proceedings of the American Mathematical Society*, 15:pp. 327–334, 1964.

[57] J. A. Morrison. Comparasion of the modified method of averaging and the two variable expansion procedure. *SIAM Review*, 8:66–85, 1966.

[58] J. A. Sanders and F. Verhulst. *Averaging Methods in Nonlinear Dynamical Systems.* Applied Mathematical Sciences, Springer-Verlag, New York, 1985.

[59] L. M. Perko. Higher order averaging and related methods for perturbed periiodic and quasi-periodic systems. *SIAM Journal on Applied Mathematics*, 17:698–724, 1969.

[60] R. P. Feynman, R. B. Leighton, and M. L. Sands. *The Feynman lectures on Physics.* Addison-wesley Publishing Co. Inc., Reading, Massachusetts, 1963.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my appreciation to those who helped me with various aspects of conducting research and the writing of this thesis. First and foremost, Dr. Murti V. Salapaka for his guidance, patience, and support throughout this research and the writing of this thesis. His insights, frank appraisals and words of encouragement have greatly inspired me during my graduate education. I would also like to acknowledge my committee members for their efforts and contributions to this work: Dr. Wolfgang Kliemann, Dr. Degang Chen, Dr. Nicola Elia and Dr. Zhengdao Wang. I would like to thank Dr. Saju Nettikadan from Bioforce Nanosciences and Dr. Irene Revenko from Asylum research for their timely help with DNA imaging. I would additionally like to thank Dr. Govindarasu Manimaran, my seniors Dr. Abu Sebastian and Dr. Anil Gannepalli for their guidance and my colleagues Tathagata De and Pranav Agarwal for their assistance in conducting this research work. I am thankful to Mrs. Durga Kocherlla (Salapaka), my lab mates Vikas Yadav, Hullas Sehgal and Tanuj Aggarwal, all staff members, and friends in my department who helped making my graduate life a success. Above all, this work could not have been completed without the blessings of Almighty, who stood by me and guided me through every phase of my life and gave me the strength and will to walk through every challenging task.