Chapter 5
Atomic Force Microscopy: Principles and Systems Viewpoint Enabled Methods

Srinivasa Salapaka and Murti Salapaka

5.1 Introduction

In 1959, Richard Feynman [12] offered an impressive vision of engineered devices at the nanoscale, where he asserted that there are no fundamentally limiting reasons that would disallow manipulation of matter at the nanoscale. He offered an elaborate hierarchy of contraptions that would facilitate altering matter at the nanoscale from a macroscale environment.

A significant milestone in this direction, the invention of the atomic force microscope (AFM) that realized an elegant and simple means of interrogating as well as manipulating matter at the nanoscale, was first reported in [6]. Here, a cantilever beam with a sharp tip at one end (see Fig. 5.1) was shown to be an effective means of sensing interatomic forces between the atoms on the tip and the atoms on the sample being interrogated. The AFM borrows a number of operating principles from its predecessor, the scanning tunneling microscope (STM). Similar to STM operation, the positioning of the sample with respect to the main probe (the cantilever-tip in the case of the AFM) is provided by using piezoelectric actuation, where the actuating material deforms when an external voltage is applied. A remarkable and enabling discovery by the inventors of STM [7] is that it is possible to deform piezoelectric material with Angstrom (an Angstrom is roughly the dimension of an atom) precision reliably by applying voltages that can be easily generated. In a typical AFM setup (see Fig. 5.1), the sample is positioned using piezoelectric scanners that provide motion of the cantilever tip with respect to all three directions; the two lateral x and y directions as well as the vertical z

S. Salapaka (✉)
University of Illinois, Urbana Champaign, Champaign, IL 61820-5711, USA
e-mail: salapaka@illinois.edu

M. Salapaka
University of Minnesota, Minneapolis, MN 55455, USA
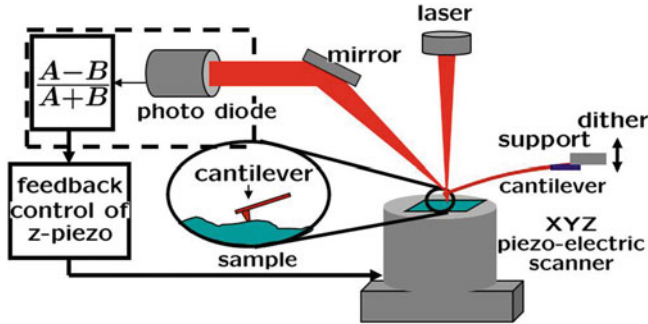e-mail: murtis@umn.edu

Fig. 5.1 An atomic force microscope schematic illustrates a cantilever with a tip that probes a sample. The support of the cantilever can be actuated by a dither piezo. The cantilever deflection is sensed by a laser beam that reflects from the cantilever surface into a split photodiode. The sample can be positioned with sub-nanometer accuracy using piezoactuated material

direction. The first AFM reported in [5] used an STM to sense the motion of the cantilever. However, using an STM to sense the cantilever motion proved to be too cumbersome. The laser beam-bounce method, where a laser beam incident on the cantilever reflects into a split photodiode, is the preferred means of sensing the cantilever deflection. The beam-bounce method leverages the large optical path that effectively amplifies a small motion of the cantilever into a relatively large motion of the laser at the photodiode. Photodiodes provide bandwidths in MHz regime with high accuracy, and the associated shot noise is dominated by other noise sources that limit the AFM operation.

At the heart of the AFM is a force sensor; a micro-cantilever that has a sharp tip at its free end. The atoms on the tip of the cantilever and the atoms on the surface of the sample exert a force on each other. These forces are typically in the piconewton range and qualitatively have the characteristics of an attractive nature for large separations and sharp repulsive nature at short ranges. A good qualitative model for the interatomic forces is the Lennard–Jones potential, where the force F between the atoms is given by $F(r) = -A/r^7 + B/r^{13}$, where r represents the separation between the atoms (see Fig. 5.2). For the cantilever flexure to have a large enough deflection (above the dominant noise sources) due to the interatomic forces the cantilevers need to be soft. Also, the first resonant frequency of the cantilever has to be away from the frequency content in the 0–2 kHz range of the disturbances of the ambient environment, which include building vibrations. Thus, a high first resonant frequency is needed for the cantilever. A small stiffness (in the range 0.06– 100 N/m) and a large resonant frequency implies small mass for the cantilevers and therefore the cantilevers employed in AFM have dimensions at the micrometer scale (a length, width, and thickness on the order of 100, 10 and 5 μm, respectively).

A prevalent use of AFMs is in unraveling the force profile between various materials, where a quantitative description of the force variation as a function of the separation between the material is reconstructed. Such a description is obtained by moving the sample relative to the tip in the vertical z direction while
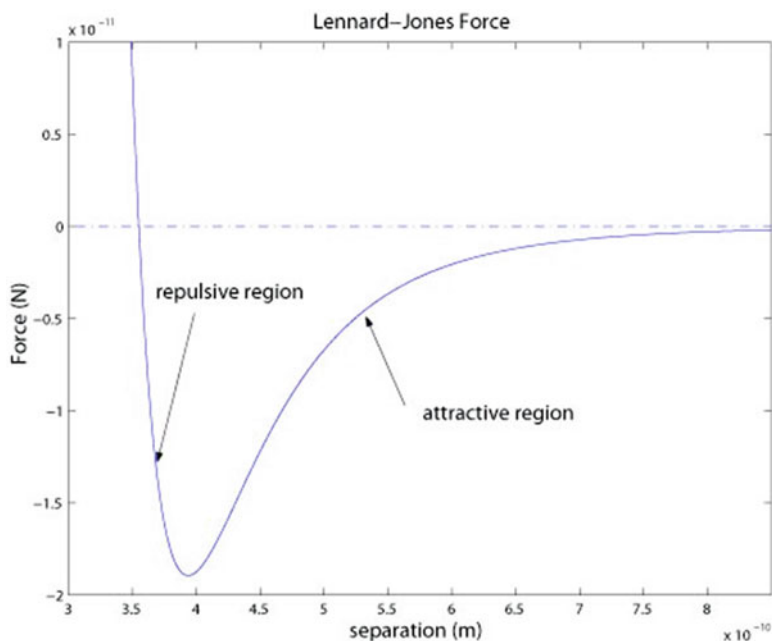
Fig. 5.2 The qualitative nature of forces between tip atoms and atoms on the sample is shown. These forces are characterized by weak attractive forces for large tip-sample separations and strong repulsive forces for small tip-sample separations

keeping the lateral position fixed. Often the cantilever tip is suitably modified to characterize forces between different materials of interest. For example, in a number of studies the cantilever tip is modified biologically with a molecule of interest and the interaction forces between this molecule and the molecules on the sample are interrogated with respect to the separation between them to obtain the force profile (often such force separation relationships differ even qualitatively from the Lennard–Jones force characteristics).

AFMs are mainly used to obtain the topography of the sample with nanometer or sub-nanometer resolution. The primary means of obtaining the topography of the sample is to move (scan) the sample laterally with respect to the tip. As the sample is scanned, at each lateral coordinate (x, y) the tip feels the interatomic force $h(x, y)$ that results in a deflection of the tip. Assuming that the material properties are the same along the lateral directions, the changes in the forces felt by the cantilever tip are due to the tip atoms moving closer or farther from the sample atoms due to the topography change in the sample. Thus the forces felt by the cantilever tip, as the sample is scanned, do not vary linearly with respect to the sample topography since the dependence of the tip-sample interaction force on the separation is qualitatively described by the nonlinear Lennard–Jones potential. Also, since there is no quantitative description of the nonlinear dependence, the

problem of extracting the sample-surface topography from the forces felt by the cantilever becomes challenging. The inventors of AFM employed feedback (the same principle was employed for STM) to overcome these challenges, which also provided a significant enabling method for interrogation at the nanoscale.

Feedback plays a pivotal role in the operability of the AFM, particularly for estimating the topography of the sample. In methods based on the *force-balance principle*, the sample is moved vertically in the z direction by a controller (see Fig. 5.1) to maintain a constant deflection of the cantilever (which is equivalent to maintaining a constant force on the cantilever) while the sample is scanned laterally in the x and y directions. The vertical movement of piezo scanner is considered an estimate of the topography of the sample with the reasoning that the sample is appropriately moved to negate the variations of the topography to regulate a constant force on the cantilever tip. This methodology proves remarkably effective and elegantly overcomes challenges in estimating the topography without the knowledge of the nonlinear force profile that exists between the tip and the sample atoms.

The demands on AFM technology are considerable, many of which primarily stem from the need to interrogate large samples with resolution at the nanometer scale. The AFM technology enables interrogation of the sample at the atomic scale with unparalleled ease. However, it interrogates the sample a single location at a time which makes it impractical for high-throughput applications. This shortcoming can be addressed partly through parallel deployment of multiple cantilevers. Another strong motivation for high-speed scanning requirements arise from studies that investigate nanoscale *dynamics* of samples. For example, a significant effort in AFM research is the study of the dynamics of bio-molecules, where the motion of molecular motors is typically examined [1]. These needs cannot be met by introducing parallelism. Thus, the need to improve the interrogation bandwidth remains central to AFM instrumentation research. Such an improvement in bandwidth implies related challenges on the positioning systems and on cantilever related technologies.

Another important objective of future AFM technologies is to provide a measure of the fidelity of the data for the sample that is being generated. Current commercially available AFMs provide scant or no information on the interpretation accuracy of the data.

As will be seen in this chapter, control systems viewpoints provide an effective means of addressing these challenges. In this chapter, after presenting the basic operational principles of AFM, we present research related to nanopositioning followed by research on the cantilever dynamics in the presence of sample forces.

5.2 Operational Principles

As described earlier, the AFM (see Fig. 5.1) uses a micro-cantilever with a sharp tip at one end to probe the sample being interrogated. The other end is fixed to a support, which can be oscillated by using a *dither piezo*. The laser beam that bounces

off the cantilever surface is collected by a quadrant photodiode that provides a measure of the cantilever's vertical deflection as well as any torsional twist of the cantilever. The controller is typically realized through analog components or digital components such as DSP and FPGA boards. The controller uses the photodiode signal to regulate the relative separation of the tip and the sample in the vertical z direction and also provides the capability to control the lateral x and y motions. In earlier AFMs, feedback was used for the z direction but no sensors were used for the lateral positioning and therefore the lateral positioning was achieved in open-loop. Lateral closed-loop operation was first realized in [9].

5.2.1 Noise Sources

The main noise sources are the laser, the photodiode, structural vibrations of the surroundings such as building sways and floor vibrations, acoustic sources, and thermal noise sources. Considering the cantilever as a linear time-invariant system that processes various input forces to yield the cantilever deflection as the output, some of the noise sources act at the input while others affect the measurement of the output of the cantilever system. The thermal noise, acoustic, and building noise sources appear at the input of the cantilever system and thus these inputs get processed by the cantilever system before these effects are perceived at the measurement. These form the main contributors to the process noise input to the cantilever system. Since cantilever behavior is well approximated by a one-mode model for most studies, the dynamics are described by

$$\ddot{p} + \frac{\omega_0}{Q}\dot{p} + \omega_0^2 p = f, \tag{5.1}$$

where p is the deflection of the cantilever, ω_0 is the first resonant frequency, Q is the quality factor of the cantilever that characterizes the damping, and f is the net external force action on the cantilever system. The associated transfer function is

$$G = \frac{1}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2}. \tag{5.2}$$

The laser noise and photodiode noise are sources that affect the measurement of the cantilever deflection and therefore appear as measurement noise ϑ. The measurement is given by

$$y = p + \vartheta. \tag{5.3}$$

One simple and elegant method of assessing the force resolution of the micro-cantilever is to obtain the power spectral density of the photodiode output when the laser beam bounces off the cantilever surface into the photodiode without any sample present. Assuming that appropriate vibration isolation is in place, the main
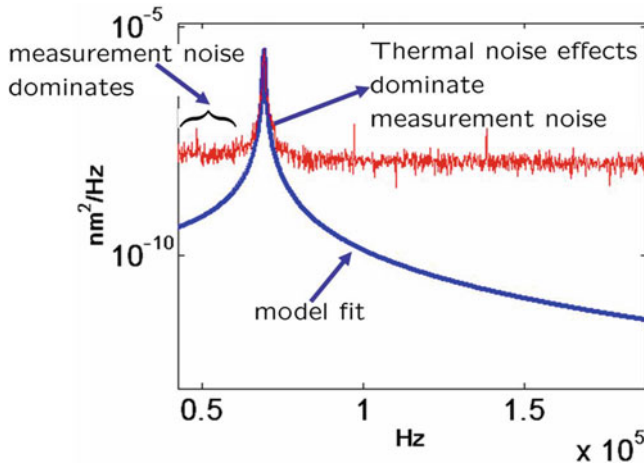
Fig. 5.3 The power spectral density of the cantilever deflection when the cantilever is forced by thermal noise is shown. The response of the cantilever due to thermal noise peaks above the noise floor only near the first resonant frequency of the cantilever demonstrating that the cantilever can sense forces as small as the thermal Langevin force near the resonant frequency. Away from the resonant frequency of the cantilever, measurement noise dominates

process noise source is the thermal noise. The thermal noise is a white noise forcing term at the input to the cantilever system caused by the interaction of the cantilever at temperature T in thermal equilibrium with its environment (see [22] for a detailed multi-mode analysis of thermal noise in cantilevers). Considering the thermal forcing as the signal of interest, it is interesting to assess if this signal can be deciphered from the measurement of the cantilever deflection. The power spectral density (psd) plot of the cantilever deflection (when the sample is absent) is shown in Fig. 5.3. A sharp peak is evident in the psd plot of thermal response. This peak is at the first resonant frequency ω_0 of the cantilever. Indeed, this also confirms that the effect of the thermal signal can indeed be deciphered from the photodiode measurement since thermal noise is dominant over other noise at frequencies near the cantilever resonant frequency ω_0. The observed cantilever deflection in Fig. 5.3 is predominantly a response to thermal noise. The other noise sources only additively corrupt the deflection measurement and are dominant only at frequencies away from the resonant frequency of the cantilever.

5.2.2 Traditional Modes of Imaging

5.2.2.1 Contact Mode Imaging

In contact mode imaging, the cantilever is not excited externally and the dither piezo is not used. The cantilever tip deflects due to interatomic forces.

In lift-mode contact-mode imaging, there is no feedback in the vertical z direction and the deflection of the cantilever is considered as an estimate of the sample topography. This method is suitable for imaging relatively flat samples over small areas since any incline or high aspect ratio features will lead to the cantilever tip either losing contact with the sample or crashing into the sample. It also has the disadvantage of not providing a quantitative estimate of the topography. However, it has the advantage of high-bandwidth operation since no feedback is required that obviates the need for high-speed positioning of the sample with respect to the tip.

In the constant-force contact-mode imaging scheme, a constant deflection of the cantilever is regulated by the controller that positions the sample with respect to the tip. The controller effort is considered as an estimate of the topography of the sample. This method provides a highly reliable estimate of the topography particularly when the frequency content of the sample topography is below the bandwidth of the z direction positioning system. It also allows for easy interpretation of data. Contact mode imaging has the drawback of imposing large lateral and vertical forces on the sample that precludes its use when imaging soft samples.

5.2.2.2 Dynamic Modes

In the dynamic modes of imaging, the cantilever is forced externally near the resonant frequency of the cantilever and information about the sample is obtained by monitoring how the cantilever's nominal motion (for example, in air) is altered under the influence of the sample. Typically, the information of the sample is modulated in a frequency range near the resonant frequency of the cantilever, where, as described before, the cantilever probe is thermally limited and thus has better signal to noise ratio.

The traditional dynamic modes of imaging can be classified into two classes: amplitude-modulation and frequency-modulation dynamic modes. In the amplitude-modulation scheme, the dither piezo drives the cantilever support sinusoidally near the resonant frequency of the cantilever that sets the cantilever tip into a periodic motion (see Fig. 5.4). The amplitude of the first harmonic of the drive frequency in the measured cantilever deflection is monitored to obtain the topography of the sample. In the constant-amplitude mode the amplitude is regulated at a set-point amplitude by the controller that alters the tip-sample separation by moving the sample with respect to the tip. Similar to the constant-force contact-mode operation, the control signal forms an estimate of the sample topography. In another mode (sometimes referred to as the error-mode), the feedback is rendered ineffective and the error between the measured amplitude and the set-point amplitude is considered as the imaging signal. In the frequency modulation mode (see Fig. 5.5), the cantilever is forced in a manner that maintains the drive at a phase of $\pi/2$ with respect to the cantilever oscillation, so that the cantilever is always maintained at the resonant condition (note that for a second order under-damped system, at resonance the output of the system is at a phase lag of $90°$ with respect to the input). The resonant frequency of the cantilever-sample system is different from
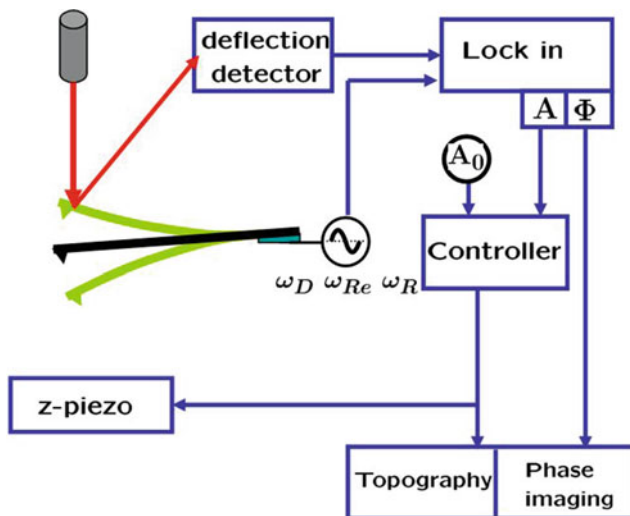
Fig. 5.4 A schematic describing the amplitude modulation AFM is shown. In this mode the cantilever is forced at a frequency ω_D that is near the first resonant frequency ω_0 of the cantilever. The amplitude and the phase of the first harmonic of the cantilever deflection are found by lock in methods. The controller moves the z piezo vertically to maintain a constant setpoint A_0 and the control signal forms an estimate of the sample topography
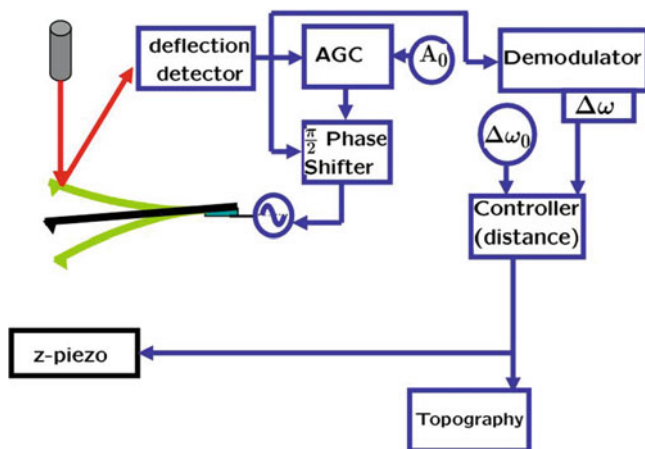

Fig. 5.5 The figure describes the frequency modulation AFM scheme where the cantilever is driven by the cantilever deflection that is phase shifted by $90°$. This scheme maintains the cantilever forced at the equivalent resonant frequency of the cantilever. The automatic gain controller alters the magnitude of the driving force to maintain a set amplitude A_0. The resonant frequency shift $\Delta\omega$ is measured and the controller positions the sample vertically to maintain a reference frequency shift $\Delta\omega_0$. The control signal forms an estimate of the sample topography

the resonant frequency of the cantilever alone as the sample introduces a periodic force every oscillation cycle on the cantilever tip. The changed resonant frequency of the cantilever-sample system is called the equivalent resonant frequency of the cantilever. The measured equivalent resonant frequency is compared to a set-point frequency and the controller positions the sample with respect to the tip to regulate the set-point frequency. As in previously discussed approaches, the control signal provides the estimate of the sample topography.

It is amply evident that feedback control forms an essential part of the traditional imaging modes. The feedback controllers employed are primarily of proportional, derivative, and integral (PID) controllers. However, as will be seen later in the chapter, modern control theory has a lot to offer to AFM instrumentation.

5.3 Control Design for Nanopositioning Systems

In both static and dynamic modes of operation, the lateral motion required for scanning is obtained by moving the sample. The positioning system, which provides this motion, plays a vital role in AFM. A significant aim of positioning-system design in AFM is to preserve the high vertical resolution provided by the cantilever sensor. High-resolution, high-bandwidth, and reliable positioning are the main performance criteria that most nanoscientific studies and applications seek from the positioning systems. One of the main challenges with current systems is achieving high tracking bandwidth. Even though the cantilever sensors have large resonant frequencies on the order of 100 kHz, the AFMs are severely limited by the vertical and lateral positioning systems that have bandwidths only on the order of 1 kHz. Another challenge that is typically neglected in typical positioning system design, is reliability. Reliability in terms of repeatability of experiments is essential for validation of the underlying studies. The delicate nature of AFM experiments, the diverse operating conditions, and lack of tools for obtaining accurate models for AFM emphasize the importance of reliable positioning. The challenge, in this context, arises in developing methods for characterizing, evaluating, and designing positioning systems for reliability. Typical nanopositioning systems comprise a flexure stage that provides frictionless motion through elastic deformation, an actuator typically made from piezoelectric material that provides the required force to deform the flexure stage, and a sensing system along with the control system. The main obstacles in the design of robust broadband nanopositioning systems stem from flexure-stage dynamics that limit the bandwidth of the positioning stage, nonlinear effects of piezoelectric actuation such as hysteresis and creep that are difficult to model, and sensor noise issues that can potentially hamper the tracking resolution of the device.

Fig. 5.6 A block diagram
schematic for a typical
nanopositioning system
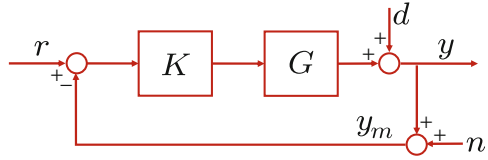
5.3.1 Performance Criteria and Limitations

One of the important contributions of control systems theory to the design of
nanopositioning systems is the quantification of performance objectives and funda-
mental design limitations. Figure 5.6 shows a block-diagram schematic of a typical
nanopositioning system. The transfer function G represents the *scanner*, which
comprises the actuator, the flexure stage, and the sensor. It is the transfer function
from the voltage input u applied to the actuator to the flexure-stage displacement
y. The signals r, d, n, and y_m, respectively, represent the reference trajectory for
tracking, the *mechanical noise*, or the effects of dynamics that are not incorporated
in the model G, the sensor noise, and the noisy measurement, while the transfer
function K represents the feedback law. The main objective for the design of the
controller K is to make the *tracking error* small, that is to make the difference $r - y$
between the desired and actual motions small.

The performance criteria are quantified by characterizing the tracking error. For
a given controller K, the tracking error for the system in Fig. 5.6 is given by

$$e = r - y = S(r - d) + Tn, \tag{5.4}$$

where the *sensitivity transfer function* $S = (1 + GK)^{-1}$ and the *complementary
sensitivity transfer function* $T = 1 - S = (1 + GK)^{-1}GK$. Thus, high resolution can
be achieved by designing the feedback law K such that S and T are small in ranges
where the frequency contents of r and n, respectively, are large. The resolution of the
closed-loop positioning system is determined by the term Tn, whereby low values of
T over a larger range of frequencies guarantee better resolution. More specifically,
the resolution of a positioning system is determined by the standard deviation σ of
the position signal when reference signal is identically zero, where

$$\sigma^2 = \int_0^\infty |T(j\omega)|^2 P_n(\omega) \, d\omega, \tag{5.5}$$

and $P_n(\omega)$ denotes the power spectral density of the noise signal n. Thus the
smaller the bandwidth of T, which is characterized by the roll-off frequency ω_T,
the smaller the standard deviation σ, and hence the better the resolution of the
closed-loop device. The tracking bandwidth is determined by the bandwidth ω_{BW}
of the sensitivity transfer function. The reliability criterion translates to robustness
of positioning systems to modeling uncertainties and operating conditions. That

are insensitive to diverse operating conditions give repeatable measurements, and are hence reliable. The peak value $\|S\|_\infty$ of the magnitude of the sensitivity function serves as a good measure to characterize the robustness of the positioning system. Thus the performance specifications translate to control design objectives of achieving high values of ω_{BW} for high tracking bandwidth, high roll-off rates of T, smaller values of ω_T for better positioning resolution, and low values of $\|S\|_\infty$ for better robustness to modeling uncertainties.

The challenges in achieving the above design objectives mainly originate from hardware and fundamental algebraic limitations on the design of the feedback law K. Good control design trades off one objective for another in the needed frequency ranges. For instance, the simple algebraic constraint of $S(j\omega) + T(j\omega) \equiv 1$ clearly implies that S and T cannot be made small simultaneously in all frequencies, which reflects the conflict between the bandwidth and the resolution objectives. Besides, for scanner G with phase margin less than $90°$, which is true for most practical systems, the bandwidth ω_{BW} cannot be larger than ω_T [31]. This limitation prevents the feedback control to achieve noise attenuation over the target reference frequency range. Another fundamental limitation that imposes a trade-off between bandwidth, resolution, and robustness requirements can be explained in terms of the Bode integral law [8, 13], which imposes the following constraint on any stable system G with the relative degree of the transfer function $K(s)G(s)$ greater than or equal to two,

$$\int_0^\infty \log|S(j\omega)|\,d\omega = 0. \tag{5.6}$$

For positioning systems, the condition on the relative degree is typically satisfied. This is so since T needs a sufficiently fast roll-off rate at high frequencies for noise attenuation (and therefore better resolution). The open-loop transfer function $K(s)G(s)$ is designed such that it has relative degree greater than or equal to 2. Also when the control is discrete and the system is analog, then the relative degree condition is inherently satisfied [18]. The limitations from this algebraic law can be explained in terms of the *waterbed effect* [11] – since the area under the graph of $\log|S(j\omega)|$ over the entire frequency range is zero, $S(j\omega)$ made small at a frequency range has to be compensated by making it large in some other frequency ranges. One direct consequence of this law is that $S(j\omega)$ cannot be made less than 1 over all frequencies. Therefore $\|S\|_\infty$, the measure for robustness is at least 1. Moreover, for positioning systems with real non-minimum phase zeros and design objectives that demand high roll-off rates on T, the following stricter fundamental algebraic limitation can be derived [13],

$$\int_0^\infty \log|S(j\omega)|W(z,\omega)\,d\omega = 0, \tag{5.7}$$

$W(z,\omega) = 2z/(z^2 + \omega^2)$ for a real positive zero z. Typical scanner systems have non-collocated actuators and sensors that are separated by flexure stages. The transfer function models of such systems generally exhibit non-minimum phase zeros. In this
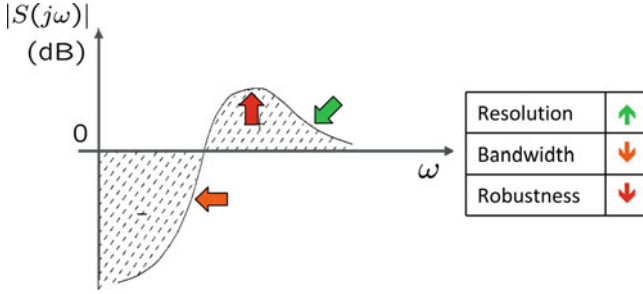
Fig. 5.7 Trade-offs due to the finite-waterbed effect. The Bode integral laws manifest themselves as waterbed effects, where decreasing the magnitude of the sensitivity function at a certain frequency range results in its increase in some other frequency range. For instance, making the sensitivity function small (near to 1) at high frequencies to ensure a high roll-off rate of the complementary sensitivity function (since $T = 1 - S$) for better resolution results in lower robustness to modeling uncertainties (due to higher values of the peak ($\|S\|_{\infty}$)) and lower values of tracking bandwidth (since ω_{BW} decreases). Similar trade-offs where one performance objective is sacrificed at the cost of others can be analyzed by studying the finite-waterbed effect (© IOP 2009), reprinted with permission

case, it can be shown that the integral of $\log|S(j\omega)|$ over a *finite* frequency range can be bounded from below, thus manifesting a waterbed effect over a finite frequency range. Thus the simultaneous requirements of low $|S(j\omega)|$ over a large frequency for a high tracking bandwidth, high order roll-off rates of T at high frequencies for high resolution, and small peaks of the $S(j\omega)$ compete against each other. For instance, small $|S(j\omega)|$ over a specified bandwidth might not leave out enough frequency range to be compatible with the integral bound in the *finite-waterbed effect* even with $S(j\omega)$ at the allowed peak value for the remaining frequencies (see Fig. 5.7).

Besides these algebraic limitations, further constraints come in the form of hardware constraints. For instance, high-order controllers require larger computation times by digital signal processor (DSP), which limit sampling rates, and therefore the tracking bandwidth. Another important limitation on control design arises from saturation limits on the actuation signals imposed by the hardware.

5.3.2 Design for Resolution, Bandwidth, and Robustness

The algebraic and practical limitations on the control design severely restrict the space of achievable performance specifications. The model-free based designs (such as proportional-integral-derivative (PID) designs) that are typically used in the nanopositioning industry, as well as designs based on loop-shaping of the open-loop transfer functions, further restrict the achievable range of specifications due to their inherent structural limitations. These techniques are inadequate to achieve simultaneously the multiple objectives of resolution, bandwidth, and robustness under the design challenges and fundamental limitations described above. The

robust optimal control theory provides an apt framework for control design for nanopositioning systems [15, 24, 29]. In this framework, it is possible to determine if a set of design specifications are feasible, and when feasible the control law K is obtained by posing and solving an optimization problem. The main advantage of using this optimization framework is that it incorporates performance objectives directly into its cost function. This eliminates the tedious task of tuning gains (in trial-and-hit manner) as in the PID designs, where even the exhaustively tuned gains may fail to yield acceptable performance. The robust control optimization problems are of the form

$$\min_{K} \|\boldsymbol{\Phi}(K)\|, \tag{5.8}$$

where $\boldsymbol{\Phi}$ is a matrix transfer function whose elements are in terms of closed-loop transfer functions in (5.4) and $\|(\cdot)\|$ represents a metric on transfer functions. The design specifications are interpreted in terms of closed-loop signals z (such as tracking error e in (5.4)) and these set $\boldsymbol{\Phi}$ as the transfer function from external variables w (such as reference signal r and noise n) to signals z. Demonstration of this framework and discussion of its advantages are presented through experimental results in the following section.

The fundamental limitations on control design presented in the previous section can be effectively used for designing better nanopositioning systems in addition to designing trade-offs between performance objectives in control design. These limitations are on the control design for a *given* scanner stage G. The study of the limitations can be used to design new scanners which result in a larger space of achievable performance specifications. This requires integration of control design into the device design, which is rarely done in the nanopositioning industry. For instance, most of the nanopositioning systems designs focus on single degree of freedom axis modules and multi-axis systems are realized by stacking individual units together. One of the main reasons for the popularity of these serial-kinematic mechanisms is that in contrast to single-mass systems, having a separate mass for each axis avoids the coupling between different axial motions. However, these multi-mass nanopositioning systems are heavier and therefore provide lower tracking bandwidth than parallel-kinematic (single-mass) mechanisms. By delegating the decoupling of axial-motions to the control design, single-mass positioning systems can be made that achieve significantly higher bandwidths for similar resolution as guaranteed by multiple-mass devices.

5.3.3 Experimental Demonstration of Optimal Control Framework

In Fig. 5.8, a parallel-kinematic *xyz* nanopositioning scanner stage and a schematic of its kinematic model are shown. The detailed design, kinematics, and dynamics analysis are described in [10]. The mechanical component of the stage is fabricated
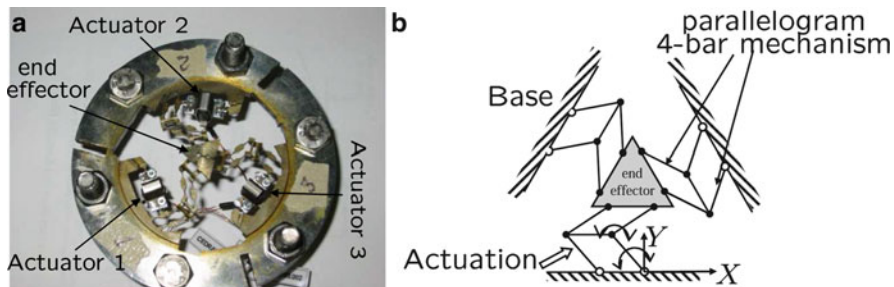
Fig. 5.8 (**a**) A prototype of a parallel-kinematic *xyz* nanopositioning system and (**b**) its schematic. In this design, three independent kinematic chains, which are piezoactuated, connect the base and the end effector. Each kinematic chain is composed of two parallelogram four-bar mechanisms, which makes the connector always parallel to the base. Together the three kinematic chains restrict all rotational degrees-of-freedom at the table, leaving it with three translations to satisfy the constraints imposed by the three kinematic chains (© ASME 2008), reprinted with permission

as a monolithic structure using electro-discharge machining (EDM). The triangular end effector at the center undergoes translation in the *x*, *y*, and *z* directions when the kinematic chains are actuated. A set of piezoelectric actuators were chosen for this stage (APA35XS by CEDRAT, free stroke 55 μm, blocking force 27 N, and maximum driving voltage 150 V) to actuate each of these kinematic chains at the first four-bar mechanism, that is connected to the base. The position sensing system consists of three capacitance displacement sensors which have a measuring range of ±50 μm and sub-nanometer resolution (0.3 nm up to 1 KHz bandwidth). Time-domain identification techniques resulted in a 13th order 3×3 multi-input multi-output (MIMO) transfer function from the input voltages to piezoelectric actuators to the sensor outputs. The six dominant modes are in the range 150–900 Hz, all of which are lightly damped with damping factors less than 0.01 (see [10] for details). Several models were identified at different operating points to characterize the modeling uncertainty. The uncertainties were dominant at low frequencies (<5 Hz). This uncertainty is primarily due to nonlinear effects of piezoelectric actuation such as hysteresis and creep.

The device design of this scanner system is such that the motion along the *x, y* and *z* directions are strongly coupled. Hence the off-diagonal terms in the MIMO transfer function are not small. For this system, it is extremely difficult and impractical to design controllers based on tuning-based or open-loop shaping techniques. The strongly coupled MIMO system makes optimal-control based design almost a necessity. The primary objective of the control design is to achieve a stage with high precision positioning and high bandwidth tracking capability that is robust to uncertainties in the operating conditions. Further, these objectives are to be met under the hardware limitation that requires the control signal (actuator input) to be within the range −1–7.5 V. The signal *z* in the optimal-control framework was chosen as $[W_P e \ W_T y \ W_u u]'$, where W_P, W_T and W_u are transfer functions that
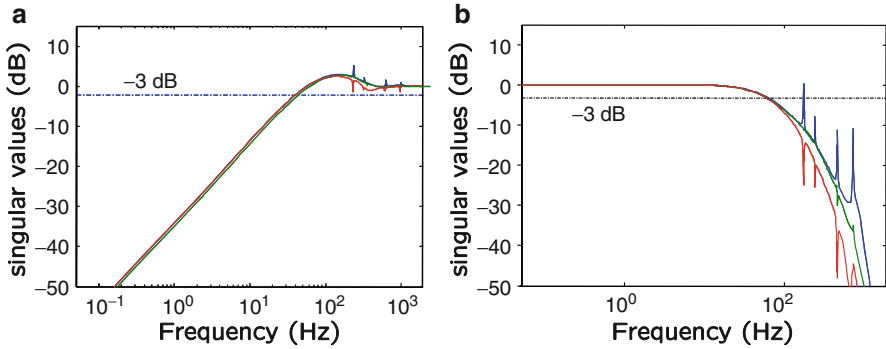
Fig. 5.9 Demonstration of the optimal-control design on a positioning system. (**a**) Singular-value plots of the sensitivity function (*solid*) from models at different operating points. The crossover frequency is 32 Hz for all the plots emphasizing the robustness in the closed-loop design. (**b**) A comparison of the corresponding complementary sensitivity transfer functions (crossover frequency is at 59 Hz) (©️ ASME 2008), reprinted with permission

reflect the relative frequency weighting of the performance objectives. The resulting transfer function Φ from $w = [r\ n]'$ to z for the formulation in (5.8) is given by $[W_P S\ W_T T\ W_u K S]$. We used the $\mathcal{H}_\infty$ norm to determine the control design. The transfer function W_P is chosen to have high gain at low frequency and low gain at high frequency, which forces the solution to the optimal control problem to have low input-to-error gain and achieves a prescribed bandwidth (the weight W_P in [10] was designed for input-to-error gain of 0.01% at low frequencies and a bandwidth of 50 Hz). The high-frequency noise attenuation is imposed by designing the weight W_T for the complimentary sensitivity transfer function T. W_T is chosen to have high gains at high frequencies so as to make T small at high frequencies (W_T was designed to ensure a high-frequency gain of 120 dB and a roll off rate of a 40 dB slope for T). The weighting function W_u is chosen to be a constant, so that the input to the actuator does not cross its saturation limits.

Figure 5.9 depicts experimental results when the controller obtained from the design described above is implemented on the positioning system. The maximum peak $\|S\|_\infty$ in the singular-value plot of the sensitivity transfer function is less than 1.85, which indicates good robustness to the modeling uncertainties and operating conditions. The singular-value plots of the sensitivity transfer function and the closed-loop transfer function have crossover frequencies at 32 and 59 Hz, respectively. The closed-loop system demonstrates a good tracking performance, as shown in Fig. 5.10, where the stage tracked the triangular reference signals well for frequencies in the range 5–40 Hz.

Figure 5.11a, b demonstrates the practical elimination of hysteresis and justi-fication of the linear model. When an actuation voltage is applied to an actuator gradually from 0 to 1.65 V in the open-loop configuration, a maximum output hysteresis of about 1.5 μm (15%) and a maximum input hysteresis of about 0.23 V (14%) is observed. However, the feedback control design effectively compensates
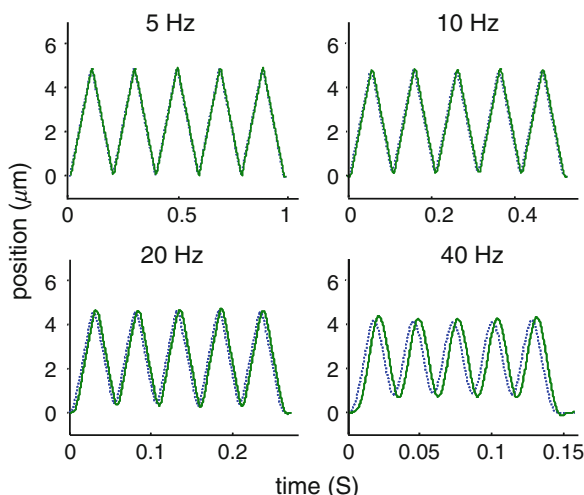
Fig. 5.10 Demonstration of tracking performance. The closed-loop system demonstrates good tracking performance (*solid*), where the stage tracked the triangular reference signals (*dashed*) with frequencies in the range 5–40 Hz. The reference and stage trajectories are practically indistinguishable at low frequencies (5 and 10 Hz) (© ASME 2008), reprinted with permission
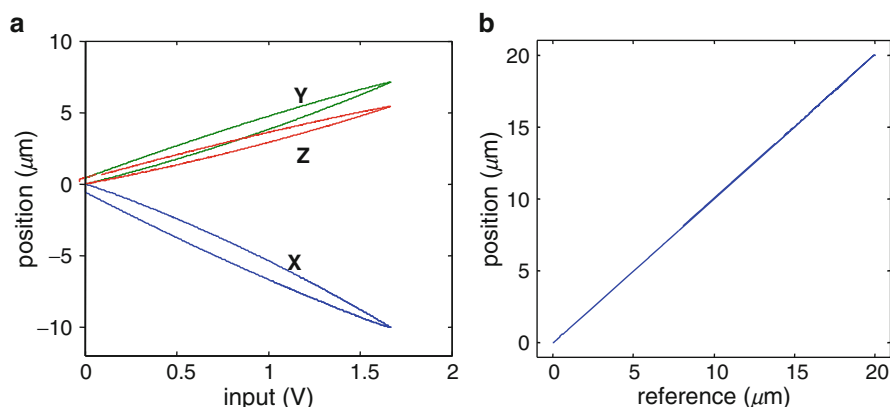

Fig. 5.11 Demonstration of practical elimination of nonlinear effects of piezoelectric actuation. Hysteresis observed in open-loop configuration (**a**) is eliminated in the closed-loop configuration (**b**) (© ASME 2008), reprinted with permission

this nonlinear effect as shown in Fig. 5.11b. A low-frequency (<1 Hz) high-amplitude (20 µm) triangular reference signal is given for the closed loop to track. The resulting experiments show maximum output and input hysteresis of about 20 nm, only 0.1% of the overall input and output range. Also, as evident from these experimental results, the closed-loop system gives a linear input–output characteristic for the entire traversal range, which validates the use of the linear
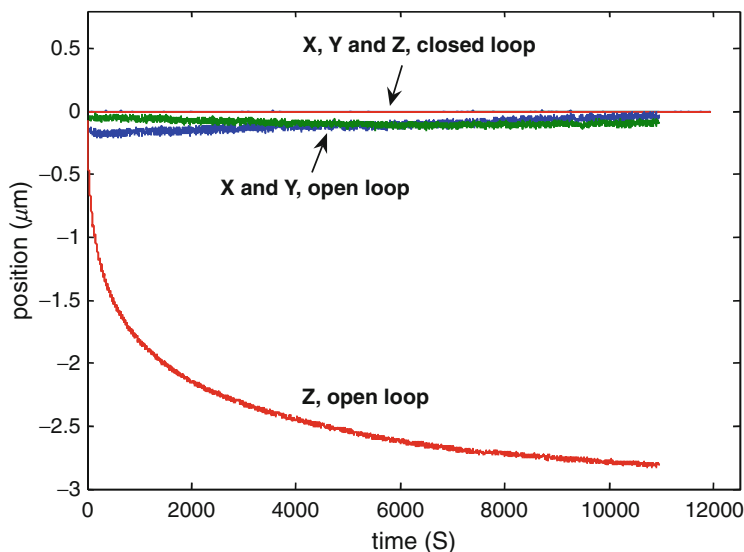
Fig. 5.12 Demonstration of practical elimination of creep effects of piezoelectric actuation by the feedback design (© ASME 2008), reprinted with permission

nominal model G for our control design. Figure 5.12 shows the creep effect of piezoelectric actuation and its elimination by the feedback design. A reference signal is designed to move the stage to the origin in open-loop and closed-loop configurations. The open-loop experiments demonstrate creep in all the actuators, especially prominent in the z direction since similar forcing on the actuators result in motions that are primarily along the z-axis due to a symmetry in the design of the positioning system. The experimental results with the closed-loop system, however, show no creep along any direction, which demonstrates that the control design practically removes this nonlinear effect.

5.3.4 Design for Ultrahigh Resolution

For a nanopositioning scanner, the positioning resolution is the smallest motion that can be differentiated from the noise in the system (signal values greater than three times the standard deviation of noise gives a 99.7% confidence in the signal). A widely held belief is that any feedback-based design strategy can only deteriorate resolution of the device since sensor (electronic) noise is fed back into the system. This argument, however, assumes no modeling uncertainty (which is always present). In fact, in the presence of significant modeling uncertainties, feedback strategies can achieve a better resolution by making better trade-offs between the effects of modeling uncertainties and the sensor noise. For instance,
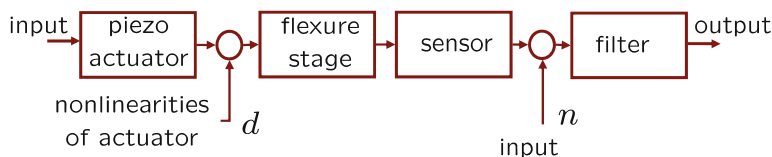
Fig. 5.13 A schematic of a single-axis open-loop system: The piezo actuator drives the flexure stage and its movement is measured by a sensor. The main challenges to achieving high resolution stem from modeling uncertainties d that arise from the nonlinear effects of piezoelectric actuation and the sensor noise n (© AIP 2007), reprinted with permission

in the context of Fig. 5.13, the resolution of a nanopositioning system is determined mainly by the mechanical and sensor noises d and n, respectively. The mechanical noise, which represents modeling uncertainties, mainly consists of the slowly varying drift and creep, which are therefore prominent in slow scans, and the inertial lag at high frequencies that is prominent in high speed scans. These nonlinear effects of drift, creep, and hysteresis are sensitive to changes in operating conditions such as ambient temperature, residual polarization in piezoelectric actuators, and the operating point – that is, the reference value on the nonlinear input–output (input voltage vs stage displacement) graph about which stage motions are calibrated. Therefore, including their precise behavior in device models is practically impossible and hence they are treated as noise. Feedback-based schemes can achieve effective compensation for the creep, drift, hysteresis, and inertial lag problems without requiring precise models. They compensate for the mechanical noise but at the cost of feeding back relatively smaller electronic noise. Hence, designing the feedback law to limit the effect of additional electronic noise from the sensor is critical. In view of the fundamental limitations on the control design presented in Sect. 5.3.1, the feedback design has to achieve the right trade-off between the resolution and the bandwidth of the resulting closed-loop positioning system. While trying to reduce the overall error e in (5.4), the error due to a large range of frequencies in the reference signal can be made small only at the expense of increased errors due to noise at those frequencies. Since the electronic noise has components in practically all the frequencies, obtaining high resolution requires high rejection bandwidth of T implying a very low rejection bandwidth of S. Fortunately, for applications such as raster scanning, the range of appreciable frequency components in reference signals is small (some scans even have only one frequency), thus making it possible to achieve high resolution. For high resolution in low-frequency scans (where r has only low-frequency components), this design can be made to exploit the resolution-bandwidth trade-off by making T have a very small *pass* bandwidth which forces S to have a very small *rejection* bandwidth, which results in a high-resolution output. Making S small at low frequencies over a small bandwidth ensures that the sensor noise is dominant over modeling uncertainties and therefore determines the resolution. This design results in low resolution since T has a very low bandwidth,
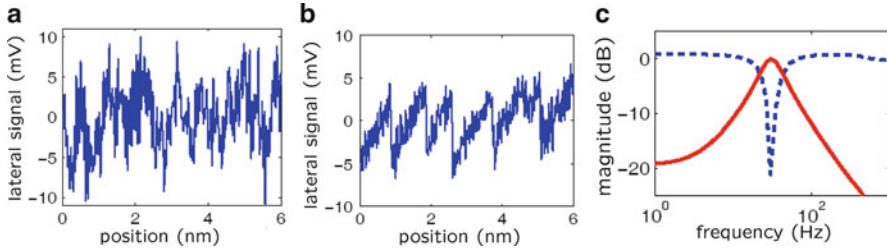
Fig. 5.14 Demonstration of sub-nanometer positioning resolution: In order to prove the high resolution capability, we used mica as the calibration sample. The friction image of mica has been established to show lattice structure (which repeats every 5.2 Å) through stick–slip dynamics that manifests as a sawtooth wave structure. The atomic-scale stick–slip is not observed in (**a**) when a high-bandwidth control design, which does not exploit the trade-off between the bandwidth and resolution, is implemented. The stick–slip is detected in (**b**) when the proposed band-focused control design in (**c**) is implemented on the nanopositioning system. With this design, lateral positioning resolution is very high (sub-nanometer) which enables detection of the atomic-scale stick–slip dynamics (© AIP 2007), reprinted with permission

which results in low value of the integral in (5.5) that determines the standard deviation of the effective noise in the closed-loop system.

The noise-management scheme described above is verified by an experiment performed on the nano-scanning stage of a commercial AFM, Molecular Force Probe-3D (MFP-3D) from Asylum Research Inc., Santa Barbara. In order to prove the high resolution capability, Mica was imaged over a scan range of 6–8 nm at slow scan rates (0.02 Hz). Under these conditions, the friction signal from mica is expected to show a triangular waveform representing the stick–slip phenomenon [4, 33] with a periodicity of 5.2 Å, which can be detected only if the positioning system has a sub-nanometer positioning resolution. The experimental results are shown in Fig. 5.14. The stick–slip phenomenon could not be observed when common (high-bandwidth or open-loop) feedback designs were used as shown in Fig. 5.14a. When a very low bandwidth (2.85 Hz) controller was designed for the scanning axis with the complementary sensitivity T as a lowpass filter (see Fig. 5.14c), sub-nanometer positioning resolution was obtained. This enabled capturing the lattice averaged atomic scale stick–slip phenomenon in the friction signal in Fig. 5.14b (see [30] for details).

This trade-off between the designs of the sensitivity and complementary sensitivity transfer functions can be extended for large scans. This approach, in fact, results in making tracking precision practically independent of the scanning frequency. For large scans, where the amplitude of the reference r is large, the tracking error e in (5.4) is determined by the Sr component which dominates over the Tn term. By making the sensitivity transfer function small over a small bandwidth around the scanning-frequency, the Sr term can be made extremely small which results in small tracking error e. In view of the finite waterbed-effect resulting from the limitation (5.6), since, the constraint on S is only over a small frequency range, a
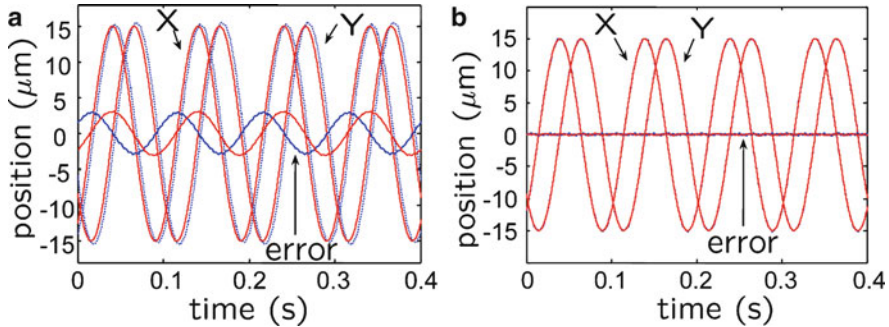
Fig. 5.15 Demonstration of better tracking with band-focused designs. Comparison of tracking of a 10 Hz sinusoidal reference signal (*dotted*) using a high-bandwidth controller in (**a**) and band-focused design in (**b**). The results show about 20 times reduction of error amplitudes by the band-focused design when compared to the high-bandwidth control design (© ASME 2008), reprinted with permission

feedback law can be obtained that makes S very small in this frequency range. This design was implemented on the positioning system shown in Fig. 5.8. The control design was obtained by implementing the optimal-control scheme described in Sect. 5.3.3, where the weighting functions W_P and W_T were chosen, respectively, as low-bandwidth band-reject and band-pass filters in the scanning frequency range. The comparison of experimental results from this design with the high-bandwidth robust controller designed in Sect. 5.3.3 are presented in Fig. 5.15. The reference $r = [r_x \, r_y]'$ is such that the positioning system tracks a circular motion with a 15 µm radius running at 10 Hz (the maximum velocities in x and y directions is about 0.95 mm/s). It is evident that the tracking error is significantly reduced. The maximum error decreased from 3 µm from the high-bandwidth controller to 150 nm for the band-focused controller.

5.3.5 Analysis and Design of 2DOF Control for Nanopositioning Systems

In this section, we show that the feasible space of performance specifications, which are constrained by the limitations described above in feedback-only configurations, can be extended by using a 2DOF design scheme. In contrast to the feedback-only scheme described in Sect. 5.3.1, where the controller acts only on the difference between the reference r and the position-measurement y_m, in the 2DOF scheme, the controller acts independently on them (see Fig. 5.16).

In the 1DOF scheme, the robustness to modeling uncertainties as well as resolution of the device are determined only by the feedback part of the controller, that is, the transfer function from d to y that characterizes robustness to modeling uncertainties is still determined by the sensitivity function $S = (1 + GK_{\mathrm{fb}})^{-1}$, and

Fig. 5.16 2DOF control
architecture: The
feedforward–feedback scheme
where the actuation signal
is $u = K_{ff}r + K_{fb}(r - y_m)$
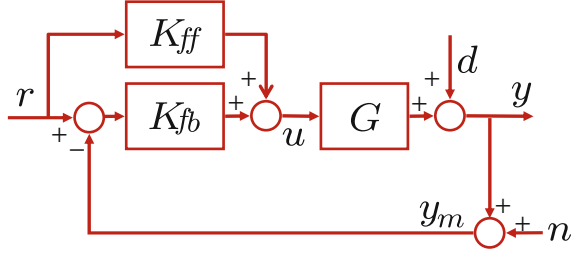
the transfer function from n to y that characterizes resolution is still determined by
the complementary sensitivity function $T = (1 + GK_{fb})^{-1}GK_{fb}$. The main difference
and advantage in 2DOF control design compared to the feedback-only design stems
from the fact that the transfer functions from r to y and from n to y are different
and can be designed independently. This difference gives greater independence in
designing for better trade-offs between different performance objectives. In this
setting, the relevant closed-loop signals are given by

$$y = T_{yr}r - Tn + Sd, \quad e = S_{er}r + Tn - Sd,$$

$$u = S(K_{ff} + K_{fb})r - SK_{fb}n - SK_{fb}d, \tag{5.9}$$

where T_{yr} and S_{er} denote the transfer function from r to y and from r to e, respec-
tively, that is, $S_{er} = S(1 - GK_{ff})$ and $T_{yr} = SG(K_{ff} + K_{fb})$. The control objectives
translate to small roll-off frequency as well as high roll-off rates for T to have good
resolution, long range of frequencies for which S_{er} is small to achieve large band-
width and low (near 1) values of the peak in the magnitude plot of $S(j\omega)$ for robust-
ness to modeling uncertainties. Even though the 2DOF control design has greater
flexibility than the feedback-only design, the main challenges to design still arise
from implementation and algebraic (albeit less severe) limitations. The constraints
on hardware implementation in terms of sampling frequencies as well as saturation
limits of actuation signals limit the scope of this design. Similarly, the algebraic
limitations constrain the control design in this setting too; for instance, the constraint
$S(j\omega) + T(j\omega) \equiv 1$ has the same ramifications on the trade-off between the
resolution and robustness to modeling uncertainties as in the feedback-only design.

A control synthesis scheme based on the optimal-control framework is discussed
in [15]. In this scheme, both the feedforward and the feedback control laws are
solved in an optimal control setting. Following the same rationale as in the 1DOF
design, the regulated output z was chosen as $[W_S e \ W_T y \ W_u u]'$ for the optimal control
problem, and Φ, the transfer function from weighted $w = [r \ n]'$ to z is given by

$$\begin{bmatrix} z_s \\ z_t \\ z_u \end{bmatrix} = \underbrace{\begin{bmatrix} W_s S_{er} W_r & -W_s S W_n \\ W_t T_{yr} W_r & -W_t T W_n \\ W_u S(K_{ff} + K_{fb})W_r & -W_u SK_{fb} W_n \end{bmatrix}}_{= \Phi(K)} \begin{bmatrix} r \\ n \end{bmatrix}, \tag{5.10}$$
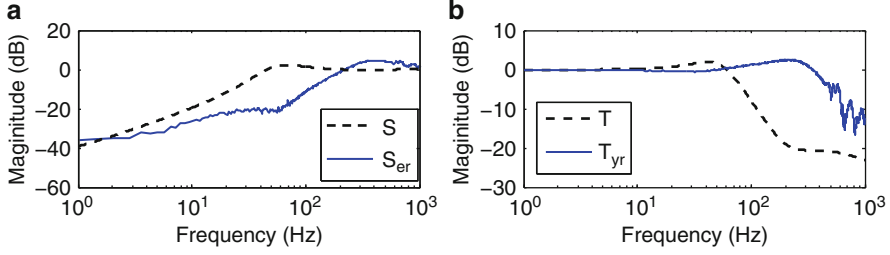
Fig. 5.17 Magnitude of $S_{er}(s)$ in (**a**) and $T_{yr}(s)$ (*solid*) in (**b**) obtained from experiments and compared to S and T(*dashed*). The 2DOF design achieves over 290% improvement in the bandwidth over the optimal 1DOF design for the same resolution and robustness specifications (© IOP 2009), reprinted with permission

where the weights W_r and W_n, which reflect the frequency content of the reference and the noise signals, provide greater independence in specifying trade-offs to the optimization problem. Accordingly, the $\mathcal{H}_\infty$ optimal control problem that we pose is $\min_K \|\Phi(K)\|_\infty$. The minimization of z_s reflects the tracking-bandwidth requirement. If we design the weight function $W_s(j\omega)$ to be large in a frequency range of $[0\ \omega_{BW}]$ and ensure that z_s is small over the entire frequency range (through the above optimization problem), then the tracking-error e will be small in the frequency range $[0\ \omega_{BW}]$; that is the closed-loop positioning device has a bandwidth of ω_{BW}. Alternatively, note that the transfer function from r to z_s is $W_s S_{er} W_r$. The optimization problem along with our choice of W_s and W_r ensures that the transfer function S_{er} is small in the frequency range $[0\ \omega_{BW}]$. Similarly the transfer function from n to z_t is the weighted complementary sensitivity function $W_t T W_n$, whose minimization ensures better resolution as it forces low control gains at high frequencies, and the transfer function from r to z_u is $W_u S(K_{ff} + K_{fb}) W_r$, which measures the control effort. Its minimization reflects imposing the limitation that the control signals be within the saturation limits. Figure 5.17a shows the experimentally obtained transfer function from reference to error, i.e. $S_{er}(s)$ which represents the tracking performance ($\omega_{BW} = 148.2$ Hz) and the transfer function from reference to output, i.e. $T_{yr}(s)$ is shown in Fig. 5.17b. There was an improvement of 290% in bandwidth for the same values of the resolution and robustness when compared to the feedback-only design. Similar improvement in other performance objectives (resolution and robustness) can be obtained by appropriately designed weight functions.

An analysis of the resulting 2DOF closed-loop system demarcates the roles of the feedforward and the feedback control. Since $S_{er} = S(1 - GK_{ff})$, and S cannot be made small over the entire bandwidth range (in order to allow for noise attenuation), the feedforward control is "active" in making S_{er} small beyond the frequency where $|S|$ is not small (say greater than $1/\sqrt{2}$). Also since $S = (1 + GK_{fb})^{-1}$ is completely determined by K_{fb}, the feedback component is dominant in frequencies where S is small. Therefore the main contribution of the feedforward component is in the frequency range where S is no longer small. However, this frequency range is

limited. Typically nanopositioning systems have very low gains beyond their flexure resonance frequencies. Therefore very high control inputs are needed to make the positioning systems practical beyond their flexure resonances. The saturation limits on control signals form the main constraints on attaining bandwidths beyond flexure resonances. Thus the feedforward components provide performance *enhancements* over feedback-only designs in the frequency range from the corner frequency of S to the flexure resonant frequency. This separation of roles becomes evident from the optimal feedback law, where the resulting design is such that $K_{ff} \approx 0$ when K_{fb} is large, even when the optimization cost function in the problem formulation does not discriminate between feedforward and feedback components.

Another important benefit of the 2DOF design is that it does yield performance specifications that are impossible for 1DOF designs. For instance, the experimental results in Fig. 5.17 show that the tracking bandwidth ω_{BW} of the closed-loop device can be made larger than the roll off frequency ω_T, which determines the resolution. The corner frequency ω_{BW} can never be made larger than ω_T in feedback-only (1DOF) designs, which suffer from a stricter trade-off between the resolution and the bandwidth. The 2DOF control based on a $\mathscr{H}_\infty$ controller has the bandwidth ω_{BW} of 148.2 Hz and ω_T of 60.1 Hz while the 1DOF optimal-controller yielded bandwidth of 49.4 Hz and the same roll-off frequency.

In this section, we have presented the demands, challenges, and control designs for nanopositioning devices. It has been shown that they fall naturally into the modern control theory paradigm. This theory provides an appropriate approach to quantify, incorporate, and achieve both the performance and robustness objectives. Using this approach, we have obtained significant improvements in bandwidth and repeatable sub-nanometer resolution for these devices. Even though the experimental results presented here have focused on bandwidth enhancements in existing positioning systems, the control theoretic framework allows for extreme specifications on resolution or robustness by making appropriate trade-offs. For instance the extreme specification of sub-nanometer resolution is achieved by a large sacrifice of tracking bandwidth in Sect. 5.3.4. Similar trade-offs can be imposed to achieve extreme specifications on robustness. The control design here is focussed on improving performance of single nanopositioning systems. There is a large scope for control systems theory in extending these design methods to an array of positioning systems specially developed for high-throughput applications.

5.4 Dynamic Mode Operation in AFM

In the dynamic mode operation of AFM, the cantilever is forced externally by a known signal. Typically the forcing is induced by moving the support of the cantilever using a dither piezo. Traditionally the support is forced sinusoidally at or near the first resonant frequency of the cantilever. However, recently, methods that force the cantilever with a superposition of sinusoids are being proposed.
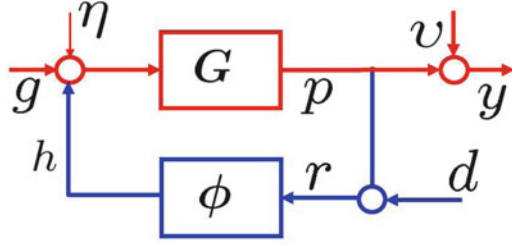
Fig. 5.18 The figure shows a feedback interconnection of a linear time invariant system G that models the cantilever with a memoryless non-linearity ϕ. Here g, η, v, and d model dither forcing, measurement noise, and sample topography, respectively

For the dynamic mode, the cantilever dynamics in (5.1) can be rewritten as

$$\ddot{p} + \frac{\omega_0}{Q}\dot{p} + \omega_0^2 p = \eta + g + h, \tag{5.11}$$

where the total force at the cantilever tip is the sum of the thermal noise force η, the external dither forcing g, and the force h on the tip due to the interaction between the tip atoms and the atoms on the sample. Note that the tip-sample interaction forces depend on the tip-deflection and possibly the tip velocity. Assuming a static dependence of the tip-sample interaction force on the tip-sample separation, r a system description of the dynamic mode AFM is provided by the feedback interconnection of the cantilever transfer function and the static nonlinearity ϕ as shown in Fig. 5.18. In this figure, the sample topography (or the initial offset between the sample and the cantilever) appears as d and the tip-sample separation r is modeled as the sum of the tip-deflection p and the tip-sample offset or the topography d. This viewpoint, where the cantilever tip-sample dynamics is posed as a feedback interconnection of a linear filter (representing cantilever dynamics) and static nonlinearity (representing tip-sample interaction), enables significant analysis and design in AFM. Such a viewpoint was first introduced in [27, 28].

For most practical implementations of control, it is necessary to identify the transfer function G from input–output experiments instead of deriving it from (5.1), since various latencies and dynamics of the AFM system components are not captured by (5.1). One means of obtaining such a description is to obtain the frequency response of the cantilever with the forcing at the dither as the input and the photodiode voltage signal as the output. It is to be remarked that in Fig. 5.18 all the forces η, h, and g see the same transfer function G. However, η is a distributed force, g is realized by forcing the base of the cantilever, whereas h appears at the tip end of the cantilever. The validity of the assumption that all these force inputs appear at the same point in the feedback interconnection and that they see the same transfer function G needs to be verified for the particular application and the dynamic method being analyzed.

In many applications, the identification of the function ϕ for a fixed lateral coordinate on the sample is the goal. In such cases, the sample position with respect to the cantilever is changed quasi-statically using the vertical piezo and the tip-deflection p is measured. That is, d is altered in a quasi-static manner by using the z piezo. Typically the dither forcing is of the form $g = A_f \sin \omega_0 t$, where ω_0 is the first resonant frequency of the cantilever. The cantilever deflection p, when the sample's influence is negligible, is sinusoidal with phase $\angle G(j\omega_0)$ and amplitude $|G(j\omega_0)|A_f$. Under the sample's influence, for a given offset d, the cantilever deflection is assumed to be periodic with period $2\pi/\omega_0$. This assumption implies a periodic force h on the cantilever tip. The sample force h on the cantilever tip can be reconstructed by realizing that the measured tip-deflection is

$$y = G(\eta + h + g) + \vartheta. \qquad (5.12)$$

Therefore, reconstruction of h requires inversion of the cantilever transfer function G and appropriate methods to address measurement noise ϑ and process noise η. Once h is estimated, a parametric model of ϕ can be assumed and the parameters of ϕ identified (see [28]), or, as introduced in the preliminary work [16], it is possible to obtain the force–separation relationship from the reconstructed h and deflection p data without assuming any model of the relationship.

When the topography image is being sought, $d = d(x,y)$ models the topography, where (x,y) represent the lateral coordinates of sample position being imaged. The lateral position of the sample is changed typically by a piezoactuated positioning system. Therefore, the sample-topography estimation problem is translated to reconstructing the sample profile d as a function of (x,y) from the measured deflection signal y.

The dynamical system depicted in Fig. 5.18 can lead to complex behavior (see [3, 14, 23]). An analysis of the amplitude and phase of the first harmonic of the cantilever oscillations when forced sinusoidally is provided in [25,26]. The study of complex behavior is significant in its own right. However, for imaging and imaging related analysis, it is evident that simplifying assumptions need to be made that render the models tractable while capturing the dynamics relevant to imaging.

One such simplifying assumption is to break the feedback interconnection of Fig. 5.18 and consider the tip sample force $h(t)$ to be a train of impulses that model the cantilever tapping the sample as it oscillates. This assumption is also well motivated as the cantilever oscillation amplitude is typically in the 25 nm regime and the tip-sample interaction forces are effective for separations less than 5–6 nm. Thus the cantilever tip spends only a small fraction of its trajectory under the sample's influence and therefore can be approximated as an impulse every oscillation cycle. Thus a reasonable approximation for h is given by

$$h(t) = \sum_n a_n \delta(t - nT), \qquad (5.13)$$

where δ denotes an impulse function, T is the time period of the dither forcing and a_n models the strength of the impulsive force during the nth cycle. The problem then
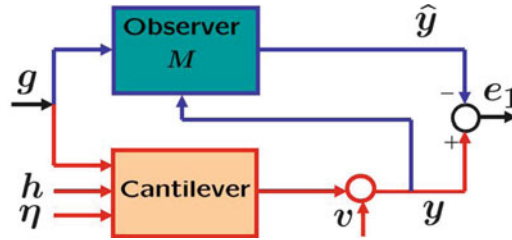
Fig. 5.19 Figure shows that an observer can be implemented on the cantilever model parameters. The observer mainly comprises a model of the cantilever and its role is to estimate the cantilever dynamics. The observer is provided with the dither forcing and the measured cantilever deflection. The mismatch in the cantilever deflection and the observer-based estimated signal is primarily due to the sample force ϕ, thermal noise η, and measurement noise ϑ

reduces to estimating the information source a_n. Such a modeling simplification has resulted in the high speed sample detection scheme described next, the transient force atomic force microscopy (TfAFM), which forms the first instance of the use of observers for AFM applications.

5.4.1 Observer-Based Dynamic Mode Methods

For h given by (5.13), the problem of estimating a_n that represent the magnitude of the sample interaction strength, becomes closely related to the problem of estimating the initial condition reset of the cantilever system state (represented by the transfer function $G(s)$) caused by the impulsive force during the nth cycle. The impulsive force contribution during the nth cycle changes the initial condition of the cantilever state and the initial condition change response appears at the measured output y. For many applications, the transfer function $G(s)$ has a large quality factor that translates to an impulse response that has a long duration. Thus, unless impulsive force inputs are separated by long intervals, the output suffers from the interference of the effects of two initial condition resets. Such an effect leads to inter-symbol interference, where the sample is considered the information source to be deciphered. An effective way of shortening the channel $G(s)$ response time is to implement an observer as shown in Fig. 5.19. The impulse response of the system viewed from the input h to the output e_1 in Fig. 5.19 can be considerably shortened (effectively from Q oscillation cycles to four cycles, where Q is the quality factor of the cantilever). Thus the channel $G(s)$ is effectively equalized using communications terminology and the source symbol sequence a_n can be deciphered considerably faster. This effectively translates to faster detection of the sample features. The details of this method can be found in [20, 21] (see Fig. 5.20).

Another challenge in dynamic mode AFM methods is the need for a measure of the fidelity of the data that is being interpreted as the image of the sample.
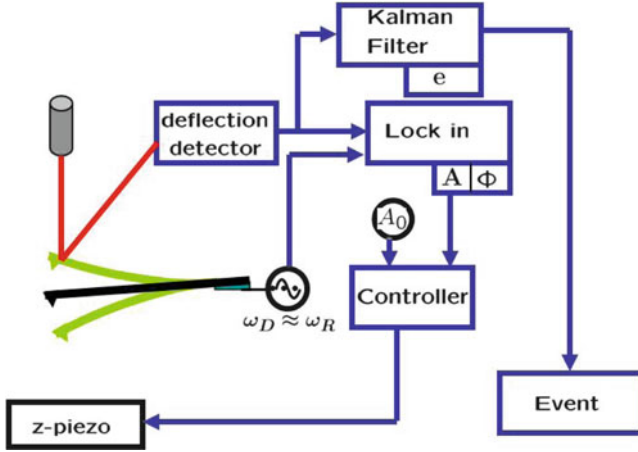
Fig. 5.20 The figure shows a schematic of transient force atomic force microscopy where in addition to the Kalman observer, a closed-loop system regulates a setpoint amplitude A_0 primarily to maintain the cantilever tip engaged with the sample surface. Every time the cantilever impacts a surface feature that is modeled as an impulsive force, an event is registered

For example, in amplitude-modulation AFM, it is often difficult to estimate in real time whether the cantilever tip is interacting with the sample or not. This issue is particularly acute for high-speed AFM applications and for samples with high aspect ratio features where the cantilever "parachutes" [2, 32] off into a valley in the sample's topography. In the absence of the sample, in steady state, after the initial condition mismatch between the cantilever state and the observer state has died out, the error e_1 is close to zero. The remnant mismatch in steady state is due to photodiode noise and the thermal noise that affect the measured cantilever deflection without affecting the observer circuit. In the presence of the sample, the cantilever behaves as a modified cantilever with changed quality factor and changed resonant frequency (see Fig. 5.21a). If the sample force is given by $h(p, \dot{p})$, the estimate of the equivalent cantilever frequency and the quality factor are given by [17]

$$\omega_0^{'2} = \omega_0^2 + \frac{2}{a} \frac{1}{2\pi} \int_0^{2\pi} h(a\cos\psi, -a\omega\sin\psi)\cos\psi\,d\psi \qquad (5.14)$$

and

$$\frac{\omega_0'}{Q'} = \frac{\omega_0}{Q} + \left(\frac{1}{a\omega} \frac{1}{\pi} \int_0^{2\pi} h(a\cos\psi, -a\omega\sin\psi)\sin\psi\,d\psi \right), \qquad (5.15)$$

respectively, where a is the steady-state amplitude of the cantilever in the presence of the sample. Thus the cantilever behavior under the sample presence is given by

$$\ddot{p} + \frac{\omega_0'}{Q'}\dot{p} + \omega_0^{'2}p = g(t) + \eta(t). \qquad (5.16)$$
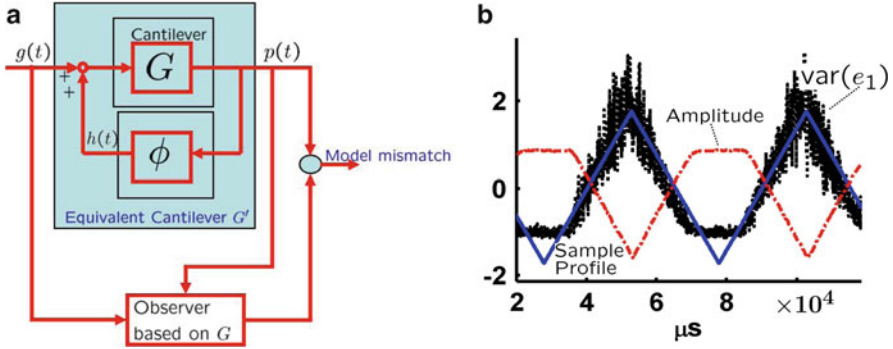
Fig. 5.21 (**a**) The cantilever-sample combination behaves as an equivalent cantilever with changed resonant frequency and changed damping factor. This combined system can be compared to an observer based on the parameters of the cantilever alone that provides a measure of the samples influence on the cantilever trajectory. (**b**) shows the experimentally obtained model mismatch signal (in black), which is the rms value of the signal e_1 when the sample first approaches the cantilever tip and then is retracted (shown by the blue triangular signal). The amplitude signal is shown in red dash-dot format

Note that the observer still behaves according to nominal cantilever parameters, thus the difference between the observer and the cantilever is due to the change in the model parameters, which do not decay with time. Thus the persistent error between the observer and the cantilever dynamics can be used to provide a signal that indicates the presence of the sample. This is unlike, and in contrast to, the change due to initial condition mismatch used for transiently detecting the sample's presence in [21]. Figure 5.21a shows the implementation architecture used, where the cantilever-sample system is viewed as an equivalent cantilever with modified quality factor and stiffness. The measured output and the observer estimated output are compared. Figure 5.21b shows the root mean square of the estimation error e_1 as the sample-cantilever separation is first decreased in the approach phase and then the separation is increased in the retract phase. As the sample approaches the cantilever, the effect of the nonlinearity ϕ on the cantilever is increased and thus the equivalent cantilever dynamics deviates more from the nominal cantilever model. The observer is based on the parameters of the nominal cantilever model and thus there is increased mismatch between the observer estimated cantilever deflection and the measured cantilever deflection. The trend of increased rms value of the signal e_1 is evident from the data shown in Fig. 5.21b. Note that this signal has a near linear relationship with the tip-sample separation. This experimentally observed linear behavior may be exploited in future research. In related work [19], real-time estimation of equivalent cantilever parameters using modified recursive least squares method is presented.

Figure 5.22 shows the AFM operated in a amplitude modulation AFM mode where the control signal (the vertical z signal to the piezo scanner) is used as an estimate of the sample height with the probe-loss architecture implemented. It is

Fig. 5.22 (Experimental data) The blue solid line depicts the sample topography, red dotted line shows the conventional signal used for imaging; the control signal and the black dotted points represent the model-mismatch (probe-loss) signal. It is evident that the probe-loss affected region of the sample can be identified in real-time
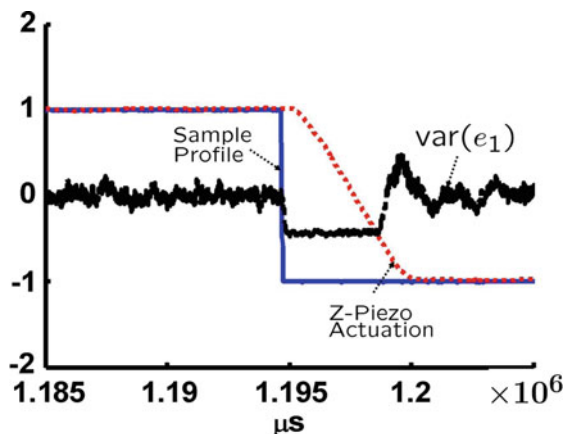
evident that as soon as a valley in the sample topography is encountered the probe-loss signal goes low. The controller actuates the z positioner to move the sample closer to the cantilever tip. Eventually the cantilever again engages with the sample and the probe-loss signal goes high. The probe-loss is detected within a couple of oscillations of the dither forcing. This remarkable speed is possible because the measurement noise in an AFM setup is quite small allowing for the observer gain to be high. It is also evident that the probe-loss signal can be used as an effective real-time means of estimating which parts of the image can be trusted and which cannot be trusted. The related observer-based framework thus provides for quantitative measurement of the image fidelity.

5.5   Conclusions

This chapter shows the impact of system-theoretic tools on nanotechnology, especially in scanning probe microscopy. Feedback formed an important and integral part of the scanning probes right in their original designs. The various applications presented in this article confirm that system-theoretic tools will continue to play a fundamental role in realizing major objectives of nanotechnology – in achieving practical control, manipulation, and investigation of matter at atomic scales.

This chapter focused on control systems theoretic modeling, analysis, and synthesis of new modes of operations that significantly expand the range of performance specifications and capabilities of scanning probe microscopes. A systems perspective was presented that enabled a study of fundamental limitations on the performance of these devices. For instance, this framework made it possible to study inherent trade-offs between resolution, tracking-bandwidth, and reliability specifications on positioning systems. In addition to determining fundamental limitations, this framework led to a better understanding of existing technology and

allowed us to exceed some technological hurdles that were previously thought to be fundamentally limiting. The system theoretic viewpoint leads to a new generation of techniques that can potentially enable probing material at the sub-angstrom level at significantly higher bandwidths. For instance, TfAFM, which achieves enormous detection bandwidth that is independent of the quality factor of the probe, and, therefore, independent of the resolution, or, ThNcAFM, that has made it possible to image with resolution as high as 0.25 Å in ambient conditions. The orders-of-magnitude improvements achieved in areas such as precision positioning, sample imaging, and sample detection rates emphasize the potential of systems tools in nanotechnology.

Many nanoscientific applications and studies, especially scanning probe microscopy, require analytical and design tools that facilitate parallelization, integration of multi-range multi-resolution actuators and sensors, real-time estimation of material features and properties from dynamic measurements, detection of artifacts, and validation of the interpreted data. These requirements can be viewed as new challenges for the areas of system identification, distributed control, robustness analysis, nonlinear estimation, and system verification and validation. In addition to extending existing methodologies in systems theory, new tools need to be developed that will specifically address the above challenges.

References

1. T. Ando, N. Kodera, E. Takai, D. Maruyama, K. Saito, and A. Toda. A high-speed atomic force microscope for studying bilological macromolecules. *Proceedings of National Academy of Science* **98**(22), 034,106 (2001)
2. T. Ando, T. Uchihashi, N. Kodera, A. Miyagi, R. Nakakita, H. Yamashita, and M. Sakashita. High-speed atomic force microscopy for studying the dynamic behavior of protein molecules at work. *Japanese Journal of Applied Physics Part 1 Regular Papers Short Notes And Review Papers* **45**(3B), 1897 (2006)
3. M. Ashhab, M.V. Salapaka, M. Dahleh, and I. Mezic. Melnikov-based dynamical analysis of microcantilevers in scanning probe microscopy. *Nonlinear Dynamics* (November 1999)
4. B. Bhushan. *Handbook of micro/nano tribology*, second edn. CRC Press (1999)
5. G. Binnig, C. Gerber, E. Stoll, T.R. Albrecht, and C.F. Quate. Atomic resolution with atomic force microscopy. *Europhys. Lett.* **3**, 1281 (1987)
6. G. Binnig, C.F. Quate, and C. Gerber. Atomic force microscope. *Physical Review Letters* **56**(9), 930–933 (1986). DOI 10.1103/PhysRevLett.56.930
7. G. Binnig and H. Rohrer. Scanning tunnelling microscopy. *Helv. Phys. Acta* **55**, 726 (1982)
8. H. Bode. *Network analysis and feedback amplifier design*. New York, Van Nostrand Reinhold (1945)
9. A. Daniele, T. Nakata, L. Giarre, M.V. Salapaka, and M. Dahleh. Robust identification and control of scanning probe microscope scanner. In: 2ND IFAC Symposium on Robust Control Design. Budapest, Hungary (1997)
10. J. Dong, S. Salapaka, and P. Ferreira. Robust control of a parallel kinematics nano-positioner. *Journal of Dynamic Systems, Measurement, and Control* **130**, 041,007(1—15) (2008)
11. J.C. Doyle, B.A. Francis, and A.R. Tannenbaum. *Feedback control theory*. New York, MacMillan (1992) URL citeseer.ist.psu.edu/doyle90feedback.html

12. R. Feynman. There's plenty of room at the bottom – an invitation to enter a new field of physics. *A talk given on December 29, 1959 at the annual meeting of APS at Caltech*

13. J.S. Freudenberg and D.P. Looze. Right half-plane poles and zeros and design tradeoffs in feedback systems. *IEEE Transactions on Automatic Control* **30**(6), 555—565 (1985)

14. S. Hu and A. Raman. Chaos in atomic force microscopy. *Physics Review Letters* pp. 036,107:1—3 (2006)

15. C. Lee and S. Salapaka. Robust broadband nanopositioning: fundamental tradeoffs, analysis and design in two degree of freedom control framework. *Nanotechnology* **20**, 035,501 (2009)

16. D. Materassi, M. Salapaka, M. Basso. Identification of interaction potentials in dynamic mode atomic force microscopy. Decision and control, *2006 45th IEEE Conference*, pp. 3702—3705 (2006). DOI 10.1109/CDC.2006.377702

17. Y.A. Mitropolskii and N. Van Dao. *Applied asymptotic methods in non-linear oscillations*. The Netherlands, Kluwer Academic Publishers (1997)

18. C. Mohtadi. Bode's integral theorem for discrete-time systems. Control theory and applications, *IEE Proceedings D* **137**, 57—66 (1990)

19. A. Pranav and M.V. Salapaka. Real time estimation of equivalent cantilever parameters in tapping mode atomic force micrscopy. *Applied Physics Letters* (2009)

20. D. Sahoo and M.V. Salapaka. Observer based imaging metods for atomic force microscopy. *Proceedings of the IEEE conference on Decision and Control*, p. Accepted for publication (2005)

21. D.R. Sahoo, A. Sebastian, and M.V. Salapaka. Transient-signal-based sample-detection in atomic force microscopy. *Applied Physics Letters* **83**(26), 5521 (2003)

22. M.V. Salapaka, H.S. Bergh, J. Lai, A. Majumdar, and E. McFarland. Multi-mode noise analysis of cantilevers for scanning probe microscopy. *Journal of Applied Physics* **81**(6), 2480—2487 (1997)

23. S. Salapaka, M. Dahleh, and I. Mezic. On the dynamics of a harmonic oscillator undergoing impacts with a vibrating platform. *Nonlinear Dynamics* **24**, 333—358 (2001)

24. S. Salapaka, A. Sebastian, J.P. Cleveland, and M.V. Salapaka. High bandwidth nano-positioner: A robust control approach. *Review of Scientific Instruments* **73**, 3232—3241 (2002)

25. A. Sebastian, A. Gannepalli, and M. Salapaka. A review of the systems approach to the analysis of dynamic-mode atomic force microscopy. Control Systems Technology, *IEEE Transactions*, **15**(5), 952—959 (2007) DOI 10.1109/TCST.2007.902959

26. A. Sebastian and M. Salapaka. Amplitude phase dynamics and fixed points in tapping-mode atomic force microscopy. *Proceedings of the American Control Conference*, Boston, Massachusetts pp. 2499—2504 (2004)

27. A. Sebastian, M.V. Salapaka, D. Chen, and J.P. Cleveland. Harmonic balance based analysis for tapping-mode AFM. *Proceedings of the American Control Conference*, San Diego (June 1999)

28. A. Sebastian, M.V. Salapaka, D. Chen, and J.P. Cleveland harmonic and power balance tools for tapping-mode atomic force microscope. *Journal of Applied Physics* **89 (11)**, 6473—6480 (June 2001)

29. A. Sebastian and S. Salapaka. Design methodologies for robust nano-positioning. *IEEE Transactions on Control Systems Technology* **13**(6), 868—876 (2005)

30. A. Shegaonkar and S. Salapaka. Making high resolution positioning independent of scan rates: A feedback approach. *Applied Physics Letters* **91**, 203,513 (2007)

31. S. Skogestad and I. Postlethwaite. *Multivariable feedback control, analysis and design*, 2nd edn. John Wiley and Sons (2005)

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

33. M.D. Yi Sang and M. Grant. Thermal effects on atomic friction. *Phys. Rev. Lett.* **87**(17), 174,301 (2001)