Review

High-speed atomic force microscopy
for nano-visualization of dynamic
biomolecular processes

Toshio Ando a,b,c,*, Takayuki Uchihashi a,b, Takeshi Fukuma c,d

a Department of Physics, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan
b CREST, JST, Sanbon-cho, Chiyoda-ku, Tokyo 102-0075, Japan
c Frontier Science Organization, Kanazawa University, Kakuma-machi, Kanazawa 920-1192, Japan
d PRESTO, JST, Sanbon-cho, Chiyoda-ku, Tokyo 102-0075, Japan

Commissioning Editor: J. Yoshinobu

Abstract

The atomic force microscope (AFM) has a unique capability of allowing the high-resolution imaging of biological samples on substratum surfaces in physiological solutions. Recent technological progress of AFM in biological research has resulted in remarkable improvements in both the imaging rate and the tip force acting on the sample. These improvements have enabled the direct visualization of dynamic structural changes and dynamic interactions occurring in individual biological macromolecules, which is currently not possible with other techniques. Therefore, high-speed AFM is expected to have a revolutionary impact on biological sciences. In addition, the recently achieved atomic-resolution in liquids will further expand the usefulness of AFM in biological research. In this article, we first describe the various capabilities required of AFM in biological sciences, which is followed by a detailed description of various devices and techniques developed for high-speed AFM and atomic-resolution in-liquid AFM. We then describe various imaging studies performed using our cutting-edge microscopes and their current capabilities as well as their limitations, and conclude by discussing the future prospects of AFM as an imaging tool in biological research.

© 2008 Elsevier Ltd. All rights reserved.
1. Introduction

The atomic force microscope (AFM) was invented in 1986 by Binnig et al. [1], four years after the invention of the scanning tunneling microscope (STM) [2]. Unlike STM or electron microscopy, AFM is unique in its ability to observe insulating objects, and hence, opened the door to the visualization of nanometer-scale objects in liquids. This unique capability was received with excitement by researchers of biological sciences as biomolecules only show vital activities in aqueous solutions. Before the AFM era, the high-resolution visualization of individual biopolymers (proteins, DNA) was only possible by electron microscopy in a vacuum environment. Many AFM imaging studies have been performed on various biological samples to explore the potential of this new microscope.
Through these studies, techniques for obtaining high-resolution images have been developed. However, AFMs unique capability, i.e., the high-resolution visualization of “active biomolecules” in solutions, does not seem to have contributed significantly to answering many biological questions.

One of the essential features of biological systems is “dynamics”. The functions of biological systems are produced through dynamic processes that occur in biopolymers, biosupramolecules, organelles, and cells. Therefore, what is required of AFM for biological sciences is the ability to rapidly acquire successive high-resolution images of individual biomolecules at work. This is solely because this type of imaging is impossible using other techniques. However, the imaging rate of conventional AFM is too slow to observe dynamic behavior of active biomolecules. Thus, endowing AFM with high-speed imaging capability is expected to have a revolutionary impact on biological sciences.

Over the past decade, various efforts have been directed toward increasing the AFM imaging rate. The most advanced high-speed AFM can now capture images at 30–60 ms/frame over a scan range of ~250 nm with ~100 scan lines. Importantly, the tip–sample interaction force has been greatly reduced without sacrificing the imaging rate, so that weak dynamic interactions between biological macromolecules are not disturbed significantly. Although the number of published reports is still limited, dynamic biomolecular processes have been successfully captured on video, some of which have revealed the functional mechanisms of proteins. As demonstrated by these studies, the newly acquired high-speed imaging capability has greatly heightened the value of AFM in biological sciences. High-speed bio-AFM will be established soon and commercially available in a few years, which is expected to increase the user population quickly. In this article, we attempt to provide the potential users and developers with comprehensive descriptions of high-speed AFM including various techniques involved in the instrumentation, applications to biological studies, current capabilities and limitations, and future prospects. In addition, we describe the recent progress in increasing the spatial resolution to the atomic level for frequency-modulation in-liquid AFM. Concise reviews on high-speed AFM were previously presented [3,4]. AFM movies placed at http://www.s.kanazawa-u.ac.jp/phys/biophys/roadmap.htm and at the publisher’s web site will give readers an indication of the power of this state-of-the-art microscope.

2. Basic principle of AFM and various imaging modes

A typical setup of tapping-mode AFM is depicted in Fig. 1. The AFM is a sort of a “palpation” microscope. It forms an image by touching the sample surface with a sharp tip attached to the free end of a soft cantilever while the sample stage is scanned horizontally in 2D. Upon touching the sample, the cantilever deflects. Among several methods of sensing this deflection, optical beam deflection (OBD) sensing is often used because of its simplicity; a collimated laser beam is focused onto the cantilever and reflected back into closely spaced photodiodes [a position-sensitive photodetector (PSPD)] whose photocurrents are fed into a differential amplifier. The output of the differential amplifier is proportional to the cantilever deflection. During the raster scan of the sample stage, the detected deflection is compared with the target value (set point deflection), and then the stage is moved in the z-direction to minimize the error signal (the difference between the detected and set point deflections). This closed-loop feedback operation can maintain the cantilever deflection (hence, the tip–sample interaction force) at the set point value. The resulting 3D
movement of the sample stage approximately traces the sample surface, and hence, a topographic image can be constructed using a computer, usually from the electric signals that are used to drive the sample stage scanner in the $z$-direction. Sometimes, the topographic image is constructed using values obtained by summing the electric signals used for driving the $z$-scanner and the error signals with an appropriate weight function. This method can give a more accurate topographic image than the former method. In the operation mode (constant-force mode; one of DC modes or contact-modes) described above, the cantilever tip, which is always in contact with the sample, exerts relatively large lateral forces to the sample because the spring constant of the cantilever is large in the lateral direction.

To avoid this problem, tapping-mode AFM (one of dynamic-modes) was invented [5], in which the cantilever is oscillated in the $z$-direction at (or near) its resonant frequency. The oscillation amplitude is reduced by the repulsive interaction between the tip and the sample. Therefore, this mode is also called the amplitude-modulation (AM) mode. The amplitude signal is usually generated by an RMS-to-DC converter and is maintained at a constant level (set point amplitude) by feedback operation.

In AM–AFM, the cantilever oscillation amplitude decreases not only by the energy dissipation due to the tip–sample interaction but also by a shift in the cantilever resonant frequency caused by the interaction [6–8]. As the excitation frequency is fixed at (or near) the resonant frequency, this frequency shift produces a phase shift of the cantilever oscillation relative to the excitation signal. When this phase shift is maintained by feedback operation.
and an image is constructed from the electric signals used for driving the z-scanner, this imaging mode is called the phase-modulation (PM) mode. Alternatively, we can construct a phase-contrast image from the phase signal, while maintaining the amplitude at a constant level by feedback operation. More details of the PM mode and phase-contrast imaging are given in Chapters 7 and 11. Instead of using a fixed frequency, it is possible to set the excitation frequency automatically to the varying resonant frequency of the cantilever using a self-oscillation circuit [9,10]. In this case, the phase of the cantilever oscillation relative to the excitation signal is always maintained at $\pm 90^\circ$, and the resonant frequency shift is maintained at a constant level by feedback operation. This mode is called the frequency-modulation (FM) mode and is described in Chapters 8 and 12.

3. History of AFM studies on biomolecular processes

In this chapter, we briefly describe the history of bio-AFM, focusing on studies on biological processes without covering a wide range of bio-AFM studies (more comprehensive descriptions on the bio-AFM history are given in a recent article [11]). The history will show that the observation of dynamic biomolecular processes started soon after the invention of AFM, whereas studies with the aim of realizing the fast imaging capability were left until later.

In 1987, the in-liquid observation ability of AFM was demonstrated [12]. Interestingly, the liquid used was not water but paraffin oil, as the surface of sodium chloride crystal was observed. Around 1988, cantilevers manufactured using microfabrication techniques became available [13], and the OBD method for detecting cantilever deflection was introduced [14]; these devices promoted the AFM imaging of biological samples such as amino acid crystals [15], lipid membranes [16], biominerals [17], and IgG [18]. Even at this very early stage, Hansma and colleagues attempted to observe the dynamic behavior of biological samples in action. For example, they observed the fibrin clotting process initiated by the digestion of fibrinogen with thrombin at $\pm 1$ min intervals [19]. Some trial observations of dynamic biological processes were also performed on the viral infection of isolated cells [20] and antibody binding to an S-layer protein [21]. We can imagine that it must have been difficult at this early stage to observe these dynamic processes, as only the contact-mode was available (tapping-mode was invented in 1993 [5]); In the constant-force mode, biomolecules weakly attached to a surface are easily dislodged by the scanning tip. At this time, more effort was directed toward attaining suitable conditions under which high-spatial-resolution images could be obtained [22–34]. Using the constant-force mode, Engel and colleagues continuously obtained very beautiful high-resolution images of membrane protein systems such as gap junctions [22], *E. coli* OmpF porin [30,31], aquaporin-1 in red blood cells [32], and bacteriorhodopsin [33,34].

Soon after the tapping-mode was invented, this mode was shown to be operational in liquid-environment [35,36]. The acoustic method, which is now often used to excite cantilevers, was introduced [35]. Later, it was shown that this mode also produces high-resolution images of membrane proteins [37]. The tapping-mode enabled the imaging of biological samples weakly attached to a substratum, which led to a moderate revival of research activity on the exploration of biological processes, although the imaging rate was as low as before. For example, in 1994, Bustamante and colleagues imaged DNA diffusion on a mica surface [38] and DNA bending upon binding to $\lambda$ Cro protein [39], and Hansma and colleagues imaged DNA digestion with DNase [40] and the DNA–RNA
polymerase binding process [41]. These two groups continued these studies and obtained time-lapse images (~30 s intervals) of the RNA transcription reaction between DNA and RNA polymerase [42] and of the 1D diffusion of RNA polymerase along a DNA strand [43]. Other examples of dynamic processes that have been imaged are the proteolytic cleavage of collagen I by collagenase [44] and nuclear-pore closing by exposure to CO₂ [45].

Attempts to increase the scan speed of AFM were initiated by Quate and colleagues (e.g., [46–48]). Their aim was to increase the speed of lithographic processing and the evaluation of a wide surface area of hard materials. For this purpose, they developed cantilevers with integrated sensors and/or actuators, and cantilever arrays with self-sensing and self-actuation capabilities. It was only possible to fabricate these sophisticated cantilevers with relatively large dimensions, thus the resonant frequency was not enhanced markedly and the spring constant was large. The insulation coating of the integrated cantilevers that allows their use in liquids further lowered the resonant frequency. The approach they employed was adequate for their purposes but unsuitable for the use of AFM in biological research, as the required conditions for high-speed AFM in the two different fields are often considerably different. Therefore, their line of studies did not result in the realization of high-speed AFM for biological research. However, note that their approach will also be useful for high-speed bio-AFM if insulated and integrated small cantilevers with a small spring constant can be fabricated in the future.

In 1993, the scan speed limit of contact-mode AFM was theoretically analyzed [49], focusing on the relationship between the cantilever’s mechanical properties and the scan speed. Some efforts aimed at increasing the bio-AFM scan speed were initiated shortly before 1995. In fact, we started to develop high-speed scanners in 1994 and small cantilevers in 1997. Hansma’s group also started to develop devices for high-speed bio-AFM around 1995. They presented the first report on short cantilevers (23 μm by 12 μm) in 1996 [50], and subsequently a report on fast imaging in 1999, in which small cantilevers and an optical deflection detector [51] designed for the small cantilevers were used to obtain an image of DNA in 1.7 s [52]. The following year, they imaged the formation and dissociation of GroES–GroEL complexes [53]. However, because of the limited feedback bandwidth, this molecular process was traced by scanning the sample stage only in the x- and z-directions. We reported a more complete high-speed AFM system in 2001 [54] and 2002 [55]. In this study, we developed a high-speed scanner, fast electronics, small cantilevers (resonant frequency, ~600 kHz in water; spring constant, 0.1 N/m) and an OBD detector for the small cantilevers. An imaging rate of 12.5 frames/s was achieved, and the swinging lever-arm-like motion of myosin V molecules was filmed as successive images over a scan range of 240 nm. However, this was only the first step in the development of truly useful high-speed AFM for biological sciences.

4. Requirements for high-speed AFM in biological research

Biological macromolecules are highly dynamic. Their functions result from dynamic structural changes and dynamic interactions with other molecules. Motor proteins transport cargo to their destinations by ‘walking’ along their filamentous protein tracks [56]. Cytoskeletons undergo polymerization and depolymerization cycles under the action by regulatory factors [57,58]. Tightly wound chromosomes are unraveled and the exposed DNA double strands are separated by helicase proteins into single strands for replication and transcription [59]. The winding and unwinding of DNA produces tension, which
results in the formation of knots. The knots can be relaxed by the action of topoisomerases [60]; the tense helical strand is cut and thereby freely spins to relieve the tension, and then the broken strands are reconnected. A newly synthesized polypeptide is trapped in the cavity of a molecular chaperon, folds into a functional 3D entity, and then detaches into the solution [61]. Outlined pictures of these dynamic biological processes have been depicted through many indirect measurements from various angles. However, it is still difficult to obtain detailed pictures of many systems. There are many biomolecular systems remaining for which even outlines of their dynamic processes have never been obtained.

Dynamic biological processes generally occur on a millisecond timescale. Therefore, firstly, biological sciences require AFM to have the ability of filming the dynamic behavior of a purified protein weakly attached to a substratum in a physiological solution. The imaging rate required is at least a few frames/s, and ideally speaking, a few hundred frames/s. Physiological functions are often produced by the interaction among a few species of molecules. If all the molecules are attached to a substratum, they have almost no chance of interacting with each other. Therefore, the selective attachment of one species of molecule to a surface is required. “Dynamic interaction” implies that the force involved in the interaction is weak. The force acting in protein–protein interactions approximately ranges from 1 pN to 100 pN. Even the single “rigor” complex of a muscle-myosin head and an actin filament, which hardly dissociates in equilibrium, is ruptured quickly by a pulling force of \( \sim 15 \text{ pN} \) [62]. The force produced by motor proteins during ATP hydrolysis is generally a few piconewtons (e.g., see [63]). Therefore, it is further required that the tip–sample interaction force can be maintained at a very small level during imaging. However, we should note that the mechanical quantity which affects the sample is not the force itself but force impulse, i.e., the product of force and the time over which the force acts. In tapping-mode high-speed AFM, the time of force action is short, and therefore, a relatively large peak force (<20 pN) would not affect the sample significantly.

When a multicomponent system contains different species of proteins with a similar shape and size, we need a means of distinguishing them. They may be distinguished by very high-resolution imaging. However, in a sample whose dynamic processes are to be observed, the movement of protein molecules is caused not only by physiological reactions but also by thermal agitation. Thus, it is often difficult to realize very high-resolution for such a moving sample. We need a high-speed recognition imaging technique to place marks on a specific species of protein molecules while capturing the topographic images.

High-speed AFM will become more useful in biological sciences if it attains the capability of observing the fine structures on living cell membranes. A large number of membrane proteins play important roles in the functions of cells. However, little is known about their dynamic molecular processes. At present, AFM cannot be applied to the observation of fine structures on living cell membranes, as the membranes are extremely soft compared with available cantilevers. Thus, it is necessary for high-speed AFM to have the ability of noncontact imaging in liquids. Recent progress in the FM–AFM of liquids has enabled the high-resolution imaging of individual hydration layers on lipid membranes [64] (see Section 12.2). This successful imaging of individual hydration layers suggests that the tip–sample interaction force must have been very weak. In addition, the high-resolution was attained using cantilevers with relatively small quality factors. Therefore, it appears to be possible to realize high-speed quasi-noncontact AFM using the FM mode or high-speed noncontact AFM (nc-AFM) using completely different modes. We will discuss this issue in Section 13.1.
The spatial resolution of optical microscopy is not sufficient for directly observing the dynamic processes of intracellular organelles. A recently developed method of fluorescence microscopy, stimulated-emission-depletion (STED) fluorescence microscopy \cite{65,66}, has a spatial resolution of $\sim 20$ nm. This high-spatial-resolution has to be compromised to attain high temporal resolution as the number of photons collected is limited \cite{67}. The AFM has generally been considered a microscope for observing surfaces. However, it was recently demonstrated that an ultrasonic technique combined with AFM allows us to observe sub-surface structures \cite{68}. Hence, the realization of high-speed “diaphan-AFM”, which would enable observing the dynamic processes of organelles in living cells, is required for biological sciences.

5. Feedback bandwidth and imaging rate

In the development of high-speed AFM apparatuses, it is important to have practical guidelines that can quantitatively indicate how each device performance affects the scan speed and the imaging rate. An early theoretical consideration of the scan speed limit in contact-mode AFM was given in \cite{49}. Concerning tapping-mode AFM, the dependence of feedback bandwidth on various factors has been qualitatively described \cite{69}. Numerical simulations were also performed for this purpose, including the effect of the dynamics of the tip–sample interaction \cite{70}. However, they are not sufficient as practical guidelines. In this chapter, we derive the quantitative relationship between the feedback bandwidth and the various factors involved in AFM devices and the scanning conditions, based on an idea previously presented for the derivation \cite{71}.

5.1. Image acquisition time and feedback bandwidth

Supposing that an image is taken in time period $T$ over the scan range $W \times W$ with $N$ scan lines, then the scan velocity $V_s$ in the $x$-direction is given by $V_s = \frac{2WN}{T}$. Assuming that the sample has a sinusoidal shape with periodicity $\lambda$, the scan velocity $V_s$ requires a feedback operation at frequency $f = \frac{V_s}{\lambda}$ to maintain the tip–sample distance. The feedback bandwidth $f_B$ should be greater than or equal to $f$ and can therefore be expressed as

$$f_B \geq \frac{2WN}{\lambda T}.$$  \hspace{1cm} (1)

Eq. (1) gives the relationship between the image acquisition time $T$ and the feedback bandwidth $f_B$. For example, for $T = 30$ ms with $W = 240$ nm and $N = 100$, the scan velocity is 1.6 mm/s. When $\lambda$ is 10 nm, $f_B \geq 160$ kHz is required to obtain this scan velocity. Note that the maximum scan velocity achievable under a given feedback bandwidth depends on the spatial frequency contained in the sample topography.

5.2. Phase delays in open-loop and closed-loop

To determine how the open-loop phase delay is related to the closed-loop phase delay, here we consider a simple feedback loop in constant-force-mode AFM (Fig. 2). The sample height variation under the cantilever tip, $u_{in}(t)$, introduced by the $x$-scan of the sample stage is considered as the input signal to this system, and the $z$-scanner displacement, $u_{out}(t)$, is considered as the output signal. The time dependence of the closed-loop input–output relationship is represented by a transfer function $K(s)$ expressed as
where $T(s)$ is the open-loop transfer function given by $C(s)A(s)H(s)G(s)$ (see Fig. 2). The frequency dependence of $T(s)$ is given by $T_0(\omega) \exp[-i\phi(\omega)]$, where $T_0(\omega)$ and $\phi(\omega)$ are the gain and phase delay, respectively. Therefore, the frequency dependence of $K(s)$ is expressed as

$$K(i\omega) = \frac{-T_0 \cos(\omega) + T_0^2}{1 + 2T_0 \cos(\omega) + T_0^2} + iT_0 \sin(\omega).$$  \hspace{1cm} (3)$$

Thus, the closed-loop phase delay $\Phi(\omega)$ and the gain $K_0(\omega)$ are, respectively, given by

$$\tan\Phi(\omega) = \frac{-\sin(\omega)}{T_0(\omega) + \cos(\omega)},$$ \hspace{1cm} (4)$$

and

$$K_0(\omega) = T_0 \sqrt{1 + 2T_0 \cos(\omega) + T_0^2}.$$ \hspace{1cm} (5)$$

When feedback performance is satisfactory and the feedback gain $K_0(\omega)$ is maintained at $\sim 1$, the open-loop gain is approximately $T_0(\omega) = -1/[2\cos(\omega)]$. By substituting this relationship into Eq. (4), we obtain the relationship $\Phi(\omega) = \pi - 2\phi(\omega)$. Here, the phase difference of “$\pi$” appears because the direction of the $z$-scanner’s displacement is opposite that of the variations in the sample height. Thus, we can conclude that the closed-loop phase delay is approximately twice the open-loop phase delay, provided the feedback gain is maintained at $\sim 1$. This is also true in tapping-mode AFM.

5.3. Feedback bandwidth as a function of various factors

From the conclusion obtained above, the time delay in the closed-loop feedback control can be estimated by summing the time delays that are caused by the devices involved in the feedback loop. The closed-loop phase delay $\theta$ [\~2\phi(\omega)] is given by $\sim 2 \times 2\pi f/\Delta$, where $\Delta$ is the total time delay in the open-loop and $f$ is the feedback frequency. In tapping-mode AFM, the main delays are the reading time of the cantilever oscillation amplitude ($\tau_d$), the cantilever response time ($\tau_c$), the $z$-scanner response time ($\tau_s$), the integral time ($\tau_I$) of error signals in the feedback controller, and the parachuting time ($\tau_p$). Here, “parachuting” means that the cantilever tip completely detaches from the sample surface at a steeply
inclined region of the sample, and thereafter, time elapses until it lands on the surface again. It takes at least a time of \(1/(2f_c)\) to measure the amplitude of a cantilever that is oscillating at its resonant frequency \(f_c\). The response time of second-order resonant systems such as cantilevers and piezoactuators is expressed as \(Q/(\pi f_0)\), where \(Q\) and \(f_0\) are the quality factor and resonant frequency, respectively. The feedback bandwidth is usually defined by the feedback frequency that results in a phase delay of \(\pi/4\). On the basis of this definition, the feedback bandwidth \(f_B\) is approximately expressed as

\[
f_B = \frac{2f_c}{8} \left( 1 + \frac{2Q_c}{\pi} + \frac{2Q_s}{\pi f_s} + 2f_c (\tau_p + \tau_1 + \delta) \right),
\]

where \(f_s\) is the \(z\)-scanner’s resonant frequency; \(Q_c\) and \(Q_s\) are the quality factors of the cantilever and \(z\)-scanner, respectively. \(\delta\) represents the sum of other time delays and \(\alpha\) represents a factor related to the phase compensation effect given by the D component in the proportional–integral–derivative (PID) feedback controller or in an additional phase compensator. From Eqs. (1) and (6), we can estimate the highest possible imaging rate in a given tapping-mode AFM setup by examining the open-loop time delay \(\Delta\tau\). However, this estimation must be modified depending on the sample to be imaged, because the allowable maximum phase delay depends on the strength or fragility of the sample.

5.4. Parachuting time

Here, we determine the conditions that cause parachuting, and obtain a rough estimate of the parachuting time and its effect on the feedback bandwidth [71]. The theoretical results obtained here are compared with experimental data to refine the analytical expression for the parachuting time.

When a sample having a sinusoidal shape with periodicity \(\lambda\) and maximum height \(h_0\) is scanned at velocity \(V_s\) in the \(x\)-direction, the sample height \(S(t)\) under the cantilever tip varies as

\[
S(t) = \frac{h_0}{2} \sin(2\pi ft),
\]

where \(f = V_s/\lambda\). When no parachuting occurs, the \(z\)-scanner moves as

\[
Z(t) = -\frac{h_0}{2} \sin(2\pi ft - \theta).
\]

The feedback error ("residual topography", \(\Delta S\)) is thus expressed as

\[
\Delta S(t) = S(t) + Z(t) = h_0 \sin \frac{\theta}{2} \cos \left( 2\pi ft - \frac{\theta}{2} \right).
\]

The cantilever tip feels this residual topography (Fig. 3) in addition to a constant height of \(2A_0 (1 - r)\), where \(A_0\) is the free-oscillation amplitude of the cantilever and \(r\) is the dimensionless peak-to-peak amplitude set point. When the set point peak-to-peak amplitude is denoted as \(A_s\), \(r = A_s/(2A_0)\). The maximum extra force exerted onto the sample due to feedback error corresponds to a distance of \(h_0 \sin(\theta/2)\). Therefore, an allowable maximum phase delay \(\theta_a\), which depends on the sample strength, is determined by this distance. The amplitude set point \(r\) is usually determined by compromising two factors: (1) increase in tapping force with decreasing \(r\), and (2) decrease in the feedback bandwidth with increasing
owing to parachuting. Therefore, the allowable maximum extra force approximately corresponds to 
\[ C_2 A_0 \left( \frac{1}{C_0} \right) \sin \left( \frac{\theta}{2} \right) \frac{1}{C_2 A_0} \left( \frac{1}{C_0} \right) \],
which gives the relationship of 
\[ \sin \left( \frac{\theta}{2} \right) = \frac{h}{h_0} \frac{1}{C_2 A_0} \left( \frac{1}{C_0} \right) \].

When \( \Delta S(t) + 2A_0(1 - r) > 0 \), no parachuting occurs. Therefore, the maximum set point \( r_{\text{max}} \) for which parachuting does not occur is given by
\[ r_{\text{max}} = 1 - \frac{h_0}{2A_0} \sin \left( \frac{\theta}{2} \right). \] (10)

Eq. (10) indicates that \( r_{\text{max}} \) decreases linearly with \( h_0/2A_0 \) (Fig. 4a) and with phase delay in the feedback operation (Fig. 4b).

The parachuting time is a function of various parameters such as the sample height \( h_0 \), the free-oscillation amplitude \( A_0 \) of the cantilever, the set point \( r \), the phase delay \( \theta \), and the cantilever resonant frequency \( f_c \). Its analytical expression cannot be obtained exactly. As a first approximation, we assume that during parachuting, Eq. (9) holds and the \( z \)-position of the sample stage does not move. During parachuting, the average separation between the sample surface and the tip at the end of the bottom swing is given by
\[ h_d = \frac{1}{2} A_0 \left( \frac{1}{C_0} \right) \left[ \frac{1}{C_0} \right] + h_0 \sin \left( \frac{\theta}{2} \right) \frac{\cos \left( \frac{2\pi f t}{c} \right)}{C_0}, \]
where \( h_0 = \beta/2\pi f \) (see the text).

This integral results in
\[ h_d = 2A_0 \left( \frac{1}{C_0} \right) \left( \frac{\tan \beta}{\beta} \right) \left( \frac{1}{C_0} \right). \] (71)

The feedback gain is usually set to a level at which the separation distance of \( 2A_0(1 - r) \) decreases to approximately zero in a single period of the cantilever oscillation. Therefore, the parachuting time \( \tau_p \) is expressed as
\[ \tau_p = \left( \frac{\tan \beta}{\beta} - 1 \right) / f_c. \] (12)

However, the assumptions, under which the average separation during parachuting was derived, are different from the reality. As mentioned later (Section 5.6), the analytical expression for \( \tau_p \) should be modified in light of the experimentally obtained feedback bandwidth as a function of \( r \) and \( h_0/A_0 \).

5.5. Integral time in the PID feedback control

The main component of PID control is the integral operation. It is difficult to theoretically estimate the integral time constant \( \tau_I \) with which the optimum feedback control is attained. Intuitively, \( \tau_I \) should be longer when a larger phase delay exists in the feedback loop. In other words, when a larger phase delay exists, the gain parameters of the PID controller cannot be increased. Therefore, \( \tau_I \) must be proportional to the height of residual
topography relative to the free-oscillation amplitude of the cantilever. As the error signals fed into the PID controller are renewed every half cycle of the cantilever oscillation, $s_I$ must be inversely proportional to the resonant frequency of the cantilever. The feedback gain should be independent of parachuting, because the gain is maximized so that optimum feedback control is performed for a nonparachuting regime. Thus, $s_I$ is approximately expressed as $s_I = j h_0 \sin (h_0/2)/A_0 f_c$, where $j$ is a proportional coefficient.

5.6. Refinement of analytical expressions for $s_p$ and $s_I$

We experimentally measured the feedback bandwidth as a function of $2A_0/h_0$ and $r$ using a mock AFM system containing a mock cantilever and $z$-scanner [71] (Fig. 5). The mock cantilever and $z$-scanner are second-order low-pass filters whose resonant frequencies and quality factors are adjusted to have the corresponding values of a real cantilever and $z$-scanner. This mock AFM system is useful for conducting a rapid inspection...
of the feedback performance. The experimentally obtained feedback bandwidths are shown by the black lines in Fig. 6. Feedback bandwidths are theoretically calculated using Eq. (6), $\kappa$ and $\beta$ as variables (see Eqs. (11) and (12)) and known values of the other parameters. From this analysis, we obtained refined expressions for $\beta$ and $\tau_1$ as follows:

$$\beta = \cos^{-1}[A_0(1 - r)/5h_0 \sin(\theta/2)],$$

$$\tau_1 = 4h_0 \sin(\theta/2)/(A_0f_c).$$

Feedback bandwidths calculated using these refined expressions are shown by the gray lines in Fig. 6. They approximately coincide with the experimental data.

5.7. Summary of guidelines for developing high-speed AFM

The following are a summary of the guidelines for realizing AFM with a high-speed imaging capability.

![Diagram](image.png)  
**Fig. 5.** Circuit diagram of a mock AFM system. The disturbance signal fed into the input 2 simulates sample topography. The output simulates the oscillation of a cantilever tip interacting with a sample surface. The amplitude change caused by the interaction is given by the diode [71].

![Graph](image.png)  
**Fig. 6.** Feedback bandwidth as a function of the set point ($r$) and the ratio ($2A_0/h_0$) of the free-oscillation peak-to-peak amplitude to the sample height. The number attached to each curve indicates the ratio $2A_0/h_0$. The feedback bandwidths were obtained under following conditions: the cantilever’s resonant frequency, 1.2 MHz; $Q$ factor of the cantilever oscillation, 3; the resonant frequency of the z-scanner, 150 kHz; $Q$ factor of the z-scanner, 0.5. Black lines, experimentally obtained feedback bandwidths using a mock AFM; gray lines, theoretically derived feedback bandwidths.
(1) All time delay components involved in the feedback bandwidth must be similar. When even one component has a significant time delay compared with the others, the feedback bandwidth is governed by the slowest component.

(2) As the cantilever resonant frequency is involved in two time delay components, it is the most important device for achieving a high-speed scan capability.

(3) The quality factors of the cantilever and z-scanner have to be lowered.

(4) The resonant frequency of the z-scanner should be high (ideally, at a level similar to that of the cantilever).

(5) The derivative operation given by the D component of the PID controller or of an additional phase compensator can compensate for the feedback delay. To make this operation effective, the gain of the z-scanner resonant peaks at high frequencies have to be lowered. Otherwise, the derivative operation produces significant mechanical vibrations and therefore cannot be used.

(6) The free-oscillation peak-to-peak amplitude of a cantilever should be a few times larger than the maximum sample height. However, this condition has to be compromised to reduce the tapping force exerted from the oscillating tip on the sample.

(7) As the tip parachuting significantly lowers the feedback bandwidth, we have to develop methods that can shorten the parachuting time or avoid parachuting. We can avoid parachuting by using a small set point amplitude. However, as this increases the tip–sample interaction force, we have to find an alternative to a small set point amplitude.

(8) All electronics used should have bandwidths as high as possible.

(9) We have to bear in mind that techniques for control operations have a minor role in the improvement of the scan speed. The highest priority has to be assigned to the improvement of the scanner and cantilevers over the consideration of sophisticated control techniques. Then, we should resort to control techniques to alleviate, to some extent, the limitation imposed by the well-optimized hardware devices.

6. Optimization of devices for high-speed AFM

6.1. Cantilevers

The feedback delays related to the cantilever are the amplitude detection time and the cantilever’s response time, both of which decrease in inverse proportion to the resonant frequency. The resonant frequency $f_c$ and the spring constant $k_c$ of a rectangular cantilever with thickness $d$, width $w$, and length $L$ are expressed as

$$f_c = \frac{0.56}{L^2} \sqrt{\frac{E}{12\rho}},$$

and

$$k_c = \frac{wd^3}{4L^5} E,$$

where $E$ and $\rho$ are Young’s modulus and the density of the material used, respectively. Young’s modulus and the density of silicon nitride (Si$_3$N$_4$), which is often used as a material for soft cantilevers, are $E = 1.46 \times 10^{11}$ N/m$^2$ and $\rho = 2.087 \times 10^3$ kg/m$^3$, respectively. To
attain a high resonant frequency and a small spring constant simultaneously, cantilevers with small dimensions must be fabricated.

In addition to the advantage in achieving a high imaging rate, small cantilevers have other advantages. For a given spring constant, the resonant frequency increases with decreasing mass of the cantilever. The total thermal noise depends only on the spring constant and the temperature and is given by $\sqrt{k_B T/k_c}$ [9], where $k_B$ is Boltzmann’s constant and $T$ is the temperature in Kelvin. Therefore, a cantilever with a higher resonant frequency has a lower noise density. In the tapping-mode, the frequency region used for imaging is approximately the imaging bandwidth (its maximum is the feedback frequency) centered on the resonant frequency. Thus, a cantilever with a higher resonant frequency is less affected by thermal noise. In addition, shorter cantilevers have higher OBD detection sensitivity, because the sensitivity follows $\Delta \phi / \Delta z = 3/2L$, where $\Delta z$ is the displacement and $\Delta \phi$ is the change in the angle of a cantilever free-end. A high resonant frequency and a small spring constant result in a large ratio ($f_c/k_c$), which gives the cantilever high sensitivity to the gradient ($k$) of the force exerted between the tip and the sample. The gradient of the force shifts the cantilever resonant frequency by approximately $-0.5 kf_c/k_c$. Therefore, small cantilevers with large values of $f_c/k_c$ are useful for phase-contrast imaging and FM–AFM. The practice of phase-contrast imaging using small cantilevers is described in Chapter 7. The usefulness and limitation of small cantilevers having a large $f_c/k_c$ in FM–AFM is described in Section 8.2.

The small cantilevers recently developed by Olympus are made of silicon nitride and are coated with gold of $\sim$20 nm thickness (Fig. 7). They have a length of 6–7 $\mu$m, a width of 2 $\mu$m and a thickness of $\sim$90 nm, which results in the resonant frequencies of $\sim$3.5 MHz in air and $\sim$1.2 MHz in water, a spring constant of $\sim$0.2 N/m, and $Q \sim$ 2.5 in water. We are currently using this type of cantilever, although it is not yet commercially available. It is possible to manufacture smaller cantilevers by microfabrication techniques to attain a higher resonant frequency as well as a small spring constant. Considering the balance between their practical use and desirable mechanical properties, we cannot expect a resonant frequency in water of much higher than 1.2 MHz. It is not practical to use a cantilever with $w < 2 \mu$m, considering the diffraction limit of the optics in the OBD detector. To keep $w$ and the spring constant unchanged, $d/L$ should be unchanged (see Eq. (16)).

Fig. 7. Electron micrograph of a small cantilever developed by Olympus. Scale bar, 1 $\mu$m.
To double the resonant frequency under this condition, both \( d \) and \( L \) should be halved (see Eq. (15)). With such a short cantilever (\( \sim 3-4 \) \( \mu \)m long), the incident laser beam used in the OBD detector tends to be eclipsed by the cantilever supporting base. In addition, the allowable tilt range of the supporting base relative to the sample substratum surface becomes narrowed. Thus, the practical upper limit of the attainable resonant frequency in water is at the very most \( \sim 2 \) MHz.

6.2. Cantilever tip

The tip apex radius of the small cantilevers developed by Olympus is \( \sim 17 \) nm\,[72], which is not sufficiently small for the high-resolution imaging of biological samples. We usually attach a sharp tip on the original tip by electron-beam deposition (EBD) in phenol gas. A piece of phenol crystal (sublimate) is placed in a small container with small holes (\( \sim 0.1 \) mm diameter) in the lid. The container is placed in a scanning electron microscope (SEM) chamber and cantilevers are placed immediately above the holes. A spot-mode electron beam is irradiated onto the cantilever tip, which produces a needle on the original tip at a growth rate of \( \sim 50 \) nm/s. The newly formed tip has an apex radius of \( \sim 25 \) nm (Fig. 8a) and is sharpened by plasma etching in argon or oxygen gas, which decreases an apex radius to \( \sim 4 \) nm (Fig. 8b). The mechanical durability of this sharp tip is not high but is still sufficient to be used to capture many images.

This piece-by-piece attachment of the tip is time-consuming. Batch procedures for attaching a sharp tip to each cantilever have been attempted by the direct growth of either a single carbon nanofiber (CNF)\,[73,74] or a carbon nanotube (CNT)\,[75] at the cantilever tip. Tanemura found that Ar ion beam-irradiation onto a carbon-coated cantilever produces a single CNF only at the apex of the original tip\,[73]. In this method, the growth orientation is easily controlled by adjusting the direction of ion-beam irradiation relative to the cantilever plane. However, at present, this method requires a carbon coating on the cantilever and cannot produce CNFs with a radius less than 10 nm. Very recently, we attempted to grow a single CNT on an original cantilever tip by chemical vapor deposition (CVD) using ethanol as a carbon source and Co as a catalyst. Although the success rate is 10% at present, we have grown a CNT at the tip with the desired orientation (Fig. 9).

![Fig. 8. Electron micrographs of an EBD tip grown on an original cantilever tip. (a) Before and (b) after sharpening by plasma etching in argon gas.](image-url)
Carbon tips probably absorb red-laser light used for OBD sensing, as the laser light is tightly focused onto a free-end region of the cantilever and passes through it to some extent. This light absorption certainly elevates the temperature at the tip. In addition, heating also occurs by red-laser light absorption at the gold coat, although the absorption rate is very small. It remains to be examined how high the temperature increases using samples which exhibit transition phenomena at temperature moderately higher than ambient temperature.

6.3. Optical beam deflection detector for small cantilevers

Schäffer et al. designed an OBD detector for small cantilevers [51]; a laser beam reflected back from the rear side of a cantilever is collected and collimated using the same lenses as those used for focusing the incident laser beam onto the cantilever. We use the same method but instead of single lenses, an objective lens with a long working distance of 8 mm (CFI Plan FluorELWD20×C, NA, 0.45, Nikon) is used [54]. The focused spot is 3–4 \( \mu \text{m} \) in diameter. The incident and reflected laser beams are separated using a quarter-wavelength plate and a polarization splitter (Fig. 10). Our recent high-speed AFM is integrated with a laboratory-made inverted optical microscope with robust mechanics. The focusing objective lens is also used to view the cantilever and the focused laser spot with the optical microscope. The laser driver is equipped with a radio-frequency (RF) power modulator to reduce noise originating in the optics [76]. Its details are described in Section 8.4. The photosensor consists of a 4-segment Si PIN photodiode (3 pF, 40 MHz) and a custom-made fast amplifier/signal conditioner (∼20 MHz).
6.4. Tip–sample interaction detection methods

Tip–sample interactions change the amplitude, phase, and resonant frequency of the oscillating cantilever. They also produce higher-harmonic oscillations. In this section, we describe methods for detecting the amplitude and the interaction force. Methods for detecting shifts in the phase and resonant frequency are described in Chapter 7 and Section 8.1, respectively.

6.4.1. Amplitude detectors

Conventional RMS-to-DC converters use a rectifier circuit and a low-pass filter, and consequently require at least several oscillation cycles to output an accurate RMS value. To detect the cantilever oscillation amplitude at the periodicity of half the oscillation cycle, we developed a peak-hold method; the peak and bottom voltages are captured and then their difference is output as the amplitude (Fig. 11) [54]. The sample/hold timing signals are usually made from the input signals (i) (i.e., sensor output signals) themselves. Alternatively, external signals (ii) that are synchronized with the cantilever excitation signals can be used to produce the timing signals. This is sometimes useful for maximizing the detection sensitivity of the tip–sample interaction because the detected signal is affected by both the amplitude change and the phase shift. This is the fastest amplitude detector and the phase delay has a minimum value of $\pi$, resulting in a bandwidth of $f_c/4$. A drawback of this amplitude detector seems to be the detection of noise as the sample/hold circuits capture the sensor signal only at two timing positions. However, the electric noise picked up in this peak-hold method is less than that produced by the thermal fluctuations of the cantilever oscillation amplitude.

A different type of amplitude (plus phase) detector can be simply constructed using an analog multiplier and a low-pass filter. The sensor signal $s(t) \sim A_m(t) \sin(\omega_0 t + \phi(t))$ is
multiplied by a reference signal \[2\sin(\omega_0 t + \phi)\] that is synchronized with the excitation signal. This multiplication produces a signal given by \[A_m(t)\cos(\phi(t) - \phi) - \cos(2\omega_0 + \varphi(t) + \phi)\]. By adjusting the phase of the reference signal and placing a low-pass filter after the multiplier output, we can obtain a DC signal of \[\sim A_m(t)\cos(\Delta \varphi(t))\], where \(\Delta \varphi(t)\) is a phase shift produced by the tip–sample interaction. In this method, the delay in the amplitude detection is determined mostly by the low-pass filter. In addition, electric noise is effectively removed by the low-pass filter.

A different method (Fourier method) for generating the amplitude signal at the periodicity of a single oscillation cycle has been proposed [77]. In this method, the Fourier sine and cosine coefficients \((A\text{ and } B)\) are calculated for the fundamental frequency from the deflection signal to produce \(\sqrt{A^2 + B^2}\). The maximum bandwidth of this detector is \(f_c/8\), half that of the peak-hold method. The electric noise level in the Fourier method was similar to that in the peak-hold method (Fig. 12a). However, regarding the accuracy of amplitude detection, the performance of the Fourier method is better because the cantilever’s thermal deflection fluctuations can be averaged in this method. Thus, the Fourier method is less susceptible to the thermal effect than the peak-hold method, and consequently, the detected amplitude variation of cantilever oscillation under a constant excitation power were less than that detected by the peak-hold method (Fig. 12b).

6.4.2. Force detectors

The nonlinear impulsive tip–sample interaction induces higher-harmonic vibrations of the cantilever. In the amplitude detection described in the previous section, these vibrations are nearly neglected. The effect of the impulse \((\sim\text{peak force }\times\text{ interaction time})\) on the cantilever motion is distributed over harmonic frequencies (integral multiples of the fundamental frequency). When the amplitude of one of the higher-harmonic vibrations is used for image formation, it can result in high-contrast images containing maps of material properties extracted by the mechanical tip–sample interaction [78–81]. The images depend on the detected harmonic frequency.

Another attempt to use higher-harmonics has been made with the aim of increasing the detection sensitivity of the tip–sample interaction. Since the impulsive force is exerted transiently in a short time, its peak force is relatively large. This means that the peak force must be a highly sensitive quantity. The force \(F(t)\) cannot be detected directly because
the cantilever’s flexural oscillation gain is lower at higher-harmonic frequencies. $F(t)$ can be calculated from the cantilever’s oscillation wave $z(t)$ by substituting $z(t)$ into the equation of cantilever motion and then subtracting the excitation signal (i.e., inverse determination problem) [81–83]. Here, we do not need to resort to the differential of $z(t)$, which is a process that yields noisy signals. An operation in which the phase of each Fourier decomposed harmonic signal is shifted by $\pi/2$ and then multiplied by an appropriate gain is identical to using the differential. To ensure that this method is effective, the cantilever oscillation signal with a wide bandwidth (at least up to $4 \times f_c$) must be detected and fast analog or digital calculation systems are necessary for converting $z(t)$ to $F(t)$. In addition, a fast peak-hold system is necessary to capture the peak force.

By neglecting the friction force, $F(t)$ can be calculated roughly by

$$F(t) = \sum_{n=2}^{m} F_n(t) = \sum_{n=2}^{m} (1 - n^2)(A_n \cos n \omega_c t + B_n \sin n \omega_c t),$$  \hspace{1cm} (17)$$

where $F_n(t)$ represents the force component with a harmonic frequency of $n \times f_c$, $A_n$ and $B_n$ are the Fourier cosine and sine coefficients of the $n$th harmonic component, respectively,
\( \omega_c \) is the fundamental resonant angular frequency, and \( m \) indicates the upper limit of the series terms to be included. Because the excitation force and the fundamental frequency component of the interaction force cannot be separated, the term with \( n = 1 \) must be neglected. As the quality factor of small cantilevers is 2–3, the friction force does not contribute significantly to the total force compared with the other forces. Fig. 13b and c shows force signals \( F(t) \) that were obtained by off-line calculation using an oscillation signal of a small cantilever weakly interacting with a mica surface.

Since the time width of the impulsive force is narrow, it appears to be difficult to capture the peak force using a sample/hold circuit. Instead of capturing the peak force, we can calculate it using the time \( t_0 \) when the cantilever oscillation reaches the bottom. For the first harmonic component of cantilever oscillation, \( Z_1(t) = A_1 \cos \omega_0 t + B_1 \sin \omega_0 t = \sqrt{A_1^2 + B_1^2} \cos(\omega_c t - \varphi) \), the time \( t_0 \) is given by \( t_0 = (\pi + 2k\pi + \varphi)/\omega_c \), where \( k \) is an integer and \( \varphi \) is the phase delay given by \( \varphi = \tan^{-1}(B_1/A_1) \). Thus, the peak force is calculated as \( F(t_0) = \sum_{n=2}^{m} F_n(t_0) \). Although the timing of the peak-force may deviate from the timing when the cantilever reaches the end of the bottom swing, we can adjust \( t_0 \) so that the maximum force signal can be attained. By including more terms in the series given by Eq. (17), we can obtain a larger force signal, as shown in Fig. 13c. However, this probably increases noise, because noise contained in the Fourier coefficients is amplified by a factor of \( (1 - n^2) \). In addition, the Fourier coefficients in the higher-harmonic components become smaller with increasing \( n \). Therefore, the maximum term to be included in the series is determined by considering the total noise contained in the calculated peak force. For the real-time calculation of the force, a fast DSP system or a fast FPGA system is required. We are now attempting to build a peak-force detection circuit with a real-time calculation capability.

Recently, a method of directly detecting the impulsive force was presented [84,85]. The torsional vibrations of a cantilever have a higher fundamental resonant frequency \( (f_t) \) than

![Fig. 13. Force signals calculated from the oscillation signal of a cantilever interacting with a mica surface in water. (a) Cantilever oscillation signal; (b) and (c) force signals calculated taking the higher-harmonics up to the fifth component (b) or up to the 8th component (c).](image-url)
that of flexural oscillations \( f_c \). Therefore, the gain of torsional vibrations excited by impulsive tip–sample interaction is maintained at \( \sim 1 \) over frequencies higher than \( f_c \). Here, we assume that flexural oscillations are excited at a frequency of \( \sim f_c \). To excite torsional vibrations effectively, “torsional harmonic cantilevers” with an off-axis tip have been introduced [84,85]. Oscilloscope traces of torsional vibration signals indicated a time-resolved tip–sample force. We recently observed similar force signals using our small cantilevers with an EBD tip at an off-axis position near the free beam end. After filtering out the \( f_c \) component from the sensor output, periodic force signals appeared clearly (Fig. 14). To use the sensitive force signals for high-speed imaging, we again need a means of capturing the peak force or a real-time calculation system to obtain the peak force. As we do not need time-resolved force signals for imaging purposes, some simplification for the calculation can be used to increase the calculation speed.

6.5. High-speed scanners

The high-speed driving of mechanical devices with macroscopic dimensions tends to produce unwanted vibrations. Therefore, among the devices used in high-speed AFM, the scanner is the most difficult to optimize for high-speed scanning. Several conditions are required to realize high-speed scanners: (a) high resonant frequencies, (b) a small number of resonant peaks in a narrow frequency region, (c) sufficient maximum displacements, (d) small crosstalk between the three axes, (e) low quality factors. The following sections (Sections 6.5.1–6.5.3) describe several techniques for satisfying these conditions simultaneously. Active damping techniques are described in Section 6.6. The performance of our developed scanners is described in Section 6.6.4.

6.5.1. Counterbalance

The quick displacements of a piezoelectric actuator exert impulsive forces onto the supporting base, which cause vibrations of the base and the surrounding framework and in

---

Fig. 14. Force signal directly obtained from the torsional signal of a small cantilever with an off-axis tip. The cantilever was excited at its first flexural resonant frequency (\( \sim 1 \) MHz) in water. The off-axis tip was intermittently contacted with a mica surface in water. Upper trace, torsional vibrations of the cantilever; lower trace, force signal obtained by filtering the torsional signal using a low-pass filter to remove the carrier wave (1 MHz). The torsional signal appears even under free-oscillation due to cross talk between flexural and torsional vibrations.
turn, of the actuator itself. To alleviate the vibrations, a counterbalance method was introduced for the $z$-scanner \[54\]: impulsive forces are countered by the simultaneous displacements of two $z$-piezoelectric actuators of the same length in the counter direction (Fig. 15a). In this arrangement, the counterbalance works effectively below the first resonant frequency of the actuators but does not work satisfactorily around the resonant frequencies. The vibration phase changes sharply around the resonant frequencies, and therefore, a slight difference in the mechanical properties of the two actuators disturbs the counterbalance.

It is possible to make different $z$-scanner designs that can counterbalance the impulsive forces more efficiently. An alternative design employs a piezoelectric actuator sandwiched between two flexures in the displacement directions (Fig. 15b). As a single actuator is used, its center of mass as well as the entire mechanism barely moves when the mechanical properties of the two flexures are similar to each other. This is true even at the resonant frequencies. This method can be used for both the $x$- and $z$-scanners. To ensure that this method is effective when used for the $z$-scanner, the flexure resonant frequency must be increased, which requires a large spring constant of the flexures, which reduces the maximum displacement of the scanner. A second alternative is that a piezoelectric actuator is held only at the rims or corners of a plane perpendicular to the displacement direction (Fig. 15c). This allows the actuator to be displaced almost freely in both directions, and consequently, the center of mass barely moves. The last alternative is that a piezoelectric actuator is glued to (or pushed into) a circular hole of a solid base so that the side rims parallel to the displacement direction are held (Fig. 15d). Even held in this way, the actuator can be displaced almost up to the maximum length attained under the load-free condition. We recently employed this method for the $z$-scanner (see Section 6.6.4).

A sample stage has to be attached to the $z$-piezoelectric actuator. To balance the weight, a dummy sample stage is attached to the opposite side of the piezoelectric actuator (in the configuration shown in Fig. 15a, it is attached to the second $z$-actuator). The weight of the dummy sample stage is chosen by considering the hydrodynamic pressure produced against the quick displacement of the sample stage. The sample-stage mass ($m_s$) decreases the lowest resonant frequency of the $z$-piezoelectric actuator (mass, $m_z$) by a factor of $\sqrt{1 + 2m_s/3m_z}$ when the configurations depicted in Fig. 15b–d are used. For the configuration shown in Fig. 15a, the factor is $\sqrt{1 + m_s/3m_z}$.

---

**Fig. 15.** Various configurations of holding piezoelectric actuator for suppressing unwanted vibrations. The piezoelectric actuators are shown in green, and the holders are shown in blue. (a) Two actuators are attached to the base, (b) the two ends of an actuator are held with flexures in the displacement direction, (c) an actuator is held only at the rims or corners of a plane perpendicular to the displacement direction, (d) an actuator is glued to solid bases at the rims parallel to the displacement direction. (i) Top view, (ii) side view. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
6.5.2. Mechanical scanner design

Tube scanners that have been often used for conventional AFM are inadequate for high-speed scanning, as their long and thin structure lowers the resonant frequencies in the \(x\), \(y\), and \(z\)-directions. The structural resonant frequency can be enhanced by adopting a compact structure and a material that has a large Young’s modulus to density ratio. However, a compact structure tends to produce interference (crosstalk) between the three scan axes. A ball-guide stage [54] is one choice for avoiding such interference. An alternative method is to use flexures (blade springs) that are sufficiently flexible to be displaced but sufficiently rigid in the directions perpendicular to the displacement axis [86,87]. The flexure-based stage is more stable and more easily built than the ball-guide stage. Note that the scanner mechanism, except for the piezoelectric actuators, must be produced by monolithic processing to minimize the number of resonant elements.

When we need a scan speed as high as possible, it is best to have asymmetric structures in the \(x\)- and \(y\)-directions; the slowest \(y\)-scanner displaces the \(x\)-scanner, and the \(x\)-scanner displaces the \(z\)-scanner, as in our currently used scanner (Fig. 16). Using this scanner, the maximum displacements (at 100 V) of the \(x\)- and \(y\)-scanners are 1 \(\mu\)m and 3 \(\mu\)m, respectively. Two \(z\)-piezoelectric actuators (maximum displacement, 2 \(\mu\)m at 100 V; self-resonant frequency, 360 kHz) are used in the configuration shown in Fig. 15a. The gaps in the scanner are filled with an elastomer to passively damp the vibrations. This passive damping is effective in suppressing low-frequency vibrations.

As developed at Hansma’s lab, a symmetrical \(x-y\) configuration has an advantage of being capable of rotating the scan direction [86]. Aluminum or duralumin is often used as a material for the scanner. Magnesium and magnesium alloys appear to be more suitable materials because of their larger mechanical damping coefficients and larger ratios of Young’s modulus to density. However, from our experiences, only a slight improvement is attained using these materials.
An alternative design for the $xy$-scanner has been used in a high-speed scanning tunneling microscope (STM) [88]. Two shear-mode piezoelectric actuators are simply stacked to produce a $xy$-scanner, on top of which a $z$-piezoelectric actuator (stack actuator) is attached. This type of scanner is commercially available from PI (Physik Instrumente) GmbH (Germany). This configuration results in a very compact structure. However, in shear-mode piezoelectric actuators, the polarization axis is perpendicular to the external electric field, and hence, a large electric field cannot be applied. The displacement rate is smaller than one-tenth that of stack-mode piezoelectric actuators. To attain sufficient displacements, a high voltage has to be applied to relatively thick piezoceramic plates. As the attachment of a $z$-piezoelectric actuator onto the stacked $xy$-scanner results in the formation of a long resonator with a low resonant frequency, we need a specially designed attachment to avoid this problem.

6.5.3. Caution regarding hydrodynamic effects

When the $z$-scanner is displaced quickly, hydrodynamic pressure is generated from the reaction of the sample solution against the cantilever base (Fig. 17). We examined this effect by quickly moving the sample stage toward the cantilever, while keeping the tip in contact with the sample stage [55]. As shown in Fig. 17a, when the sample stage dimensions were $5 \times 5$ mm$^2$, the cantilever vibrated at $\sim 20$ kHz due to the hydrodynamic force. When the size of the sample stage was reduced to $1$ mm in diameter, almost no vibrations appeared (Fig. 17b). Therefore, we have been using glass sample stages with a circular-trapezoid shape and a small top surface of $1$ mm diameter. When the cantilever was oscillated at its resonant frequency and the tip was intermittently in contact with the sample stage, no significant vibrations appeared, even when using a larger sample stage. However, we later recognized that even when no vibrations appeared, the hydrodynamic pressure decreased the feedback bandwidth in tapping-mode operation. After examining this effect by changing the cantilever position relative to the sample stage, we found that this decrease in feedback bandwidth was due to the slight and slow displacement of the cantilever base. We were able to solve this problem by tightly fixing the cantilever base using a stiff blade spring or by fixing the cantilever base at a position near the cantilever beam.

![Fig. 17](image)

Fig. 17. Disturbance on cantilever by hydrodynamic force and its dependence on the sample stage shape. The sample stage is quickly moved toward the cantilever whose tip is in contact with a mica surface in water. The lower signals represent the voltage applied to the $z$-piezoelectric actuator to move the sample stage. The upper signals represent the cantilever deflection. (a) The sample stage is a slide glass cut to $5 \times 5$ mm$^2$. (b) The sample stage is a glass of circular-trapezoid shape with a small top surface of $1$ mm diameter [55].
As a solution to this problem, it seems best to alter the design of the cantilever base so that it catches much less hydrodynamic pressure. In fact, after a cantilever base was partially cracked at the edge close to the cantilever beam, the base did not move by the exerted hydrodynamic pressure even when the base was not tightly fixed.

6.6. Active damping

It is impossible to manufacture a scanner that exhibits no resonance around the harmonic frequencies contained in its driving signals. Therefore, the elimination of unwanted vibrations from the scanner is a key to realizing high-speed AFM. Furthermore, the bandwidth of any mechanical system is limited by its dimensions and materials, and hence, some manipulation techniques are also required to overcome this limitation and expand the scanner bandwidth. In this section, we describe several techniques related to these issues. In high-speed AFM, z-scanner driving may be too fast to control even using advanced digital mode controllers. Therefore, it is assumed below that active damping of the z-scanner is performed using analog mode controllers. Vibration control is basically classified into three types: notch filtering, feedback control, and feed-forward control (Fig. 18). In the following sections, these three types of vibration control are explained in detail.

6.6.1. Notch filtering

Notch filtering (Fig. 18a) is convenient for eliminating resonance with a single peak. The transfer function \( N(s) = \frac{s^2 + \omega_0 s/\omega^2_0 - \omega^2_1}{s^2 + \omega s/\omega^2_0 - \omega^2_1} \) of a standard two-zero, two-pole notch filter is given by

\[
N(s) = \frac{s^2 + \omega_0 s/\omega^2_0 - \omega^2_1}{s^2 + \omega s/\omega^2_0 - \omega^2_1}.
\]

At \( \omega = \omega_0 \), \( N(s) \) becomes \( Q/\omega^2_0 \) (hence, no phase change). This value gives the minimum transmission rate. To eliminate the resonance of a scanner characterized by a simple transfer function \( G(s) = \frac{\omega^2_1}{s^2 + \omega_1 s/\omega^2 + \omega^2} \), \( \omega_0 \) and \( \omega_1 \) must be tuned to \( \omega_1 \) and \( Q_1 \), respectively. This tuning results in the composite transfer function \( C(s) = \frac{\omega^2_1}{s^2 + \omega_1 s/\omega^2 + \omega^2} \). Therefore, the scanner vibrations are damped completely using \( Q = 0.5-0.7 \). However, some effort is required to tune \( q \) as well as \( Q \) using analog notch filters. When the scanner is characterized by a more complicated transfer function with multiple peaks, it is important to find whether the multiple peaks arise from multiple resonators connected in series or in parallel. From the observed phase spectrum of the scanner, we can judge the type of connection. When the resonators are connected in series, the phase of the transfer function

![Fig. 18. Three types of active damping methods for suppressing the unwanted vibrations of the scanner. (a) Notch filtering method, (b) feedback control method, (c) feed-forward control method. G(s) represents the transfer function of a scanner to be controlled. H(s) represents the transfer function of a controller for active damping.](image)
decreases monotonically and significantly with increasing frequency, while for a parallel connection, the phase shows an up-and-down feature. For a case of serial connection, we can eliminate the multiple peaks by using notch filters connected in series. For the case of a parallel connection, it is impossible to eliminate the peaks.

6.6.2. Feedback control for active damping

The feedback operator $H(s)$ (see Fig. 18b) converts the resonant system $G(s)$ to a target system $R(s)$ expressed as

$$R(s) = \frac{G(s)}{1 - G(s)H(s)}.$$  \hspace{1cm} (19)

In the first step, let us consider the simplest case where $G(s)$ consists of a single resonant element with resonant frequency $\omega_1$ and quality factor $Q_1$. The target system is expressed as a single resonator with resonant frequency $\omega_0$ and quality factor $Q_0$. For these systems, $H(s)$ is expressed as

$$H(s) = \frac{1}{G(s)} - \frac{1}{R(s)} = \left(\frac{s^2}{\omega_1^2} + \frac{s}{Q_1\omega_1} + 1\right) - \left(\frac{s^2}{\omega_0^2} + \frac{s}{Q_0\omega_0} + 1\right).$$  \hspace{1cm} (20)

To eliminate the second-order term from $H(s)$, $\omega_0$ should equal $\omega_1$, which results in

$$H(s) = -\left(\frac{1}{Q_0} - \frac{1}{Q_1}\right)\frac{s}{\omega_1}. \hspace{1cm} (21)$$

This $H(s)$ is identical to a derivative operator with gain $= -1$ at the frequency $\hat{\omega} = Q_0Q_1\omega_1/(Q_1 - Q_0)$ (Fig. 19a). By adjusting the gain parameter of the derivative operator, we can arbitrarily change the target quality factor $Q_0$. This method is simpler than notch filtering, as the tuning parameter is simply the gain. This method is known as “Q-control” and is often used for controlling the quality factor of cantilevers [89–91]. For the Q-control of cantilevers with constant frequency excitation, we can change the phase of feedback signal by $90^\circ$ using a phase shifter, instead of a derivative operator. However, this method does not guarantee Q-control of a cantilever whose tip is transiently contacting with the sample, as higher-harmonic components are produced by the contact.

If we do not eliminate the second-order term from $H(s)$, then $H(s)$ is equivalent to a system consisting of a derivative operator and a second-derivative operator connected in parallel (Fig. 19b). This second-derivative operator corresponds to “mass-control”. Using this second type of Q-controller, we can extend the scanner bandwidth when a second-derivative circuit and a piezoelectric actuator driver, both having high bandwidths, can be used. Here, note that when either Q-controller is applied to the z-scanner, we must

![Fig. 19. Diagram of feedback control for active vibration damping. (a) Simple Q-control system, (b) combined control system able to extend the scanner bandwidth.](image-url)
measure the displacement or velocity of the \( z \)-scanner. However, it is difficult to do so. This problem is solved by using a mock scanner \( M(s) \) (a second-order low-pass filter) characterized by the same transfer function as the \( z \)-scanner (Fig. 20) [92]. The practical use of this technique is described in Section 6.6.4.

Real scanners often exhibit multiple resonant peaks, and therefore, neither of the methods described above work well. When elemental resonators are connected in series, we have to use mock scanners, \( M_1(s), M_2(s), \ldots \), each of which is characterized by a transfer function representing each elemental resonator. Each mock scanner is controlled by a corresponding Q-controller. For example, when the scanner consists of two resonators connected in series, the composite transfer function \( C(s) \) of the total system is expressed as

\[
C(s) = \frac{1}{1 - M_1(s)H_1(s)} \times \frac{1}{1 - M_2(s)H_2(s)} \times G_1(s)G_2(s). \tag{22}
\]

Since \( M_1(s) \) and \( M_2(s) \) are the same as (or similar to) \( G_1(s) \) and \( G_2(s) \), respectively, Eq. (22) represents a target system consisting of damped resonators connected in series (see Eq. (19)). Note that we can use \( M_1(s) \) and \( M_2(s) \) separately, unlike \( G_1(s) \) and \( G_2(s) \).

When the scanner consists of elemental resonators connected in parallel, active damping becomes more difficult. Here, we only consider the case in which two elemental resonators are connected in parallel. The controller \( H(s) \) is expressed as

\[
H(s) = \frac{1}{G_1(s) + G_2(s)} - \frac{1}{R(s)} = \frac{1}{\frac{a\omega_1^2}{s^2 + \omega_1^2 + \omega_0^2} + \frac{b\omega_2^2}{s^2 + \omega_2^2 + \omega_0^2}} - \left( \frac{s^2}{Q_0\omega_0^2} + \frac{s}{Q_0\omega_0} + 1 \right), \tag{23}
\]

where \( a \) and \( b \) are the weight factors of the two resonators (hence, \( a + b = 1 \)). To eliminate the highest-order term from \( H(s) \), \( \omega_0^2 \) should equal \( a\omega_1^2 + b\omega_2^2 \), which approximately results in (see Fig. 21)

\[
H(s) \sim [(\text{band-pass filter}) \times g - 1] \times s/\omega. \tag{24}
\]
The band-pass filter BP(s), \( \tilde{\omega} \), and the gain parameter \( g \) are, respectively, expressed as

\[
BP(s) = \frac{(\frac{a}{Q_2\omega_2} + \frac{b}{Q_1\omega_1})s}{s^2 + (\frac{a}{Q_2\omega_2} + \frac{b}{Q_1\omega_1})s + 1},
\]

(25)

\[
\tilde{\omega} = \frac{1}{\left(\frac{1}{Q_0\omega_0} - \frac{a}{Q_1\omega_1} - \frac{b}{Q_2\omega_2}\right)},
\]

(26)

\[
g = c\tilde{\omega} \left(\frac{a}{Q_2\omega_2} + \frac{b}{Q_1\omega_1}\right) + 1,
\]

(27)

where \( c \) is given by

\[
c = \frac{1}{\omega_0^2} + \frac{1}{\omega_1^2} - \frac{\omega_0^2}{\omega_1^2\omega_2^2} + \frac{1}{\omega_0\omega_1\omega_2} \left[ \frac{\omega_0}{Q_1Q_2} - \frac{\omega_1\omega_2}{\omega_0} - \frac{1}{Q_0} \left( \frac{a\omega_1}{Q_2} + \frac{b\omega_2}{Q_1} \right) \right].
\]

(28)

The effect of this controller \( H(s) \) is depicted in Fig. 22.

When the scanner’s transfer function is given by a combination of serial and parallel connections of elemental resonators, it becomes difficult to design a controller. It is best to manufacture a scanner with a simple structure so that the number of resonators is minimized.

6.6.3. Feed-forward control for active damping

The feed-forward control type of active damping (Fig. 18c) is based on inverse compensation (i.e., \( H(s) \sim 1/G(s) \)). Generally, inverse compensation-based damping has an advantage in that we can extend the scanner bandwidth. This damping method is much easier to apply for the \( x \)-scanner than for the \( z \)-scanner, because for the former, the scan waves are known beforehand and are periodic (hence, the frequencies used are discrete; integral multiples of the fundamental frequency) [93–95]. Here, we first consider the damping of

Fig. 22. Effect of feedback control for active damping on the resonant system that consists of two elemental resonators connected in parallel. This effect was theoretically obtained using the feedback Q-controller shown in Fig. 21. (a) Bode plots of the resonant system before damping. (b) Bode plots of the resonant system after damping.
the $x$-scanner vibrations. The waveforms of the $x$-scan are isosceles triangles characterized by amplitude $X_0$ and fundamental angular frequency $\omega_0$. Their Fourier transform is given by

$$F(\omega) = 2\pi X_0 \left[ \frac{1}{2} \delta(\omega) - \frac{2}{\pi^2} \sum_{k=-\infty}^{+\infty} \frac{1}{k^2} \delta(\omega - k\omega_0) \right] \quad (k : \text{odd}). \quad (29)$$

To move the $x$-scanner in the isosceles triangle waveforms, the input signal $X(t)$ sent to the $x$-scanner characterized by the transfer function $G(s)$ is given by the inverse Fourier transform of $F(\omega)/G(\omega)$, which is expressed as

$$X(t) = \frac{X_0}{2} - \frac{4X_0}{\pi^2} \sum_{k=1}^{+\infty} \frac{1}{k^2} \frac{1}{G(ik\omega_0)} \cos(k\omega_0 t) \quad (k : \text{odd}). \quad (30)$$

In practice, the sum of the first $\sim 10$ terms in the series of Eq. (30) is sufficient. We can calculate Eq. (30) in advance to obtain numerical values of $X(t)$ and output them in succession from a computer through a D/A converter. This method is convenient and works well. However, applying this method to the $y$-scanner is not practical because the period of "$y$-return" (scanning towards the origin after acquiring a frame) is much shorter than the total period of the $y$-scan (i.e., image acquisition time), and hence, we have to use too many terms in the series expansion of the $y$-scan waveform. We can simply use a round waveform for the $y$-return. Rounding the vertexes of the triangle waveforms is also useful for the $x$-scanner, although the deviation of the scan trajectory from the triangle waveform becomes large.

The transfer function $G(s)$ usually changes after the repeated use of the $x$-scanner, and therefore, more robust inverse-based control is required to avoid periodic examination of $G(s)$. To this end, feedback control can be applied, in addition to the feed-forward control mentioned above, in order to compensate for the change in $G(s)$ (and hence for the insufficient feed-forward control) [96,97]. The feedback control reduces the deviation of the output trajectory from the desired trajectory, and consequently, can also reduce the hysteresis and creep effects of the $x$-piezoelectric actuator.

Applying the inversion-based feed-forward active damping to the $z$-scanner is not easy because its scan waves are not predictable, and it is difficult to design an electronic circuit that is characterized by the inverse transfer function of the $z$-scanner. It may be possible to fabricate such a circuit when the scanner exhibits a single resonant peak. There seem to be two possible designs for the circuit. One is the parallel connection of a second-derivative operator, a derivative operator, and a proportional gain (i.e., $s^2/\omega_0^2 + s/(Q\omega_0) + 1$). The other can be fabricated by using a serially connected LRC circuit driven by a current regulation amplifier (Fig. 23) [98]. However, it is not easy to make this circuit.

![Fig. 23. Current-regulated circuit for feed-forward active damping control (inverse compensation) in the single resonator system.](image-url)
It is desirable to develop a technique that enables the automatic production of an inverse transfer function (APITF) for a given transfer function. The transfer function $K(s)$ of the circuit shown in Fig. 24a is given by $K(s) = 1/[1 + g(M(s) - 1)]$; here, $M(s)$ is a mock scanner. In the case of $g = 1$, a complete inverse transfer function, $1/M(s)$, can be realized. However, delays occur in the electronic components such as the operational amplifiers used in the circuit. Therefore, the gain factor $g$ must be less than 1, and consequently, the complete inverse transfer function cannot be realized. In fact, $K(s)M(s)$ has a large resonant peak at a frequency higher than the peak frequencies of $M(s)$. However, this peak can be eliminated using a notch filter. When the disagreement between $K(s)$ and $1/M(s)$ is significant, we can reduce the difference by placing a phase compensator ($1 +$ derivative operation) immediately before or after the mock scanner and by using a multiple-loop circuit, as indicated in Fig. 24b. The effectiveness of this technique is shown in Fig. 25, which was obtained by numerical calculations of $K(s)$ for the case that $G(s)$ consists of two elemental resonators connected in parallel. The resonant peaks are eliminated and the bandwidth is significantly extended. Unfortunately, this method has a limitation; it does not work when $G(s)$ has more than three poles. We have

![Fig. 24. Circuit diagram for APITF. $G(s)$ represents a given transfer function for a real resonant system. $M(s)$ represents a transfer function of a mock resonator for $G(s)$. (a) Single loop circuit type. $K(s)$ approximately gives $1/M(s)$. (b) Nested circuit for improving the approximation [111].](image)

![Fig. 25. Effect of feed-forward control for active damping on the resonant system that consists of two elemental resonators connected in parallel. This effect was theoretically calculated using the 7-fold-nested APITF system. (a) Bode plots of the resonant system before damping. (b) Bode plots of the resonant system after damping.](image)
been using this APITF technique for compensating the slow response of the photothermal bending of a cantilever (see Section 6.10) and z-scanners (Section 6.6.4).

6.6.4. Practice of active damping of the scanner vibrations

In this section, we describe the practice of active damping applied to the scanners that we are currently using (one type is shown in Fig. 16) and to other scanners under development.

Here, we first show the effect of active damping applied to the x-scanner. The active damping was performed using an inverse compensation method. The x-scanner transfer function (amplitude and phase spectra) is shown in Fig. 26. When the line scan was

Fig. 26. Frequency spectra (bode plots) of mechanical response of the x-scanner.

Fig. 27. Effect of feed-forward active-damping control on the x-scanner vibrations. (a) x-Scanner displacement driven by triangle waves without damping. (b) x-Scanner displacement driven by triangle waves compensated by the feed-forward control. (c,d) Mica surface images acquired at 32 frames/s (c) without and (d) with damping.
performed at 3.3 kHz without damping, its displacement exhibited vibrations (Fig. 27a). When it was driven in a waveform calculated using Eq. (30) with the maximum term \( k = 17 \) in the series, the \( x \)-scanner moved approximately in a triangle waveform (Fig. 27b). This damping effect was also verified by imaging mica surfaces at 33 frames/s over a scan range of \( 240 \times 240 \) nm\(^2\) with 100 scan lines (this imaging speed corresponds to an \( x \)-scan frequency of 3.3 kHz). When no active damping was applied, striation patterns parallel to the \( y \)-axis appeared (Fig. 27c). The periodicity of the patterns indicates that the vibrations had a frequency of \( \approx 60 \) kHz, which coincided with the resonant peak frequency shown in Fig. 26. When active damping was applied, the striation patterns completely disappeared (Fig. 27d). Here, note that the resonant vibrations in the \( x \)-direction also produce vibrations in the \( z \)-direction.

In the \( z \)-scanner that we are currently using for imaging studies, the \( z \)-piezoelectric actuators have a resonant frequency of 410 kHz under free-oscillation. The counterbalance method employed is depicted in Fig. 15a. Using this method of counterbalancing, no structural resonances appear at frequencies lower than the first resonant frequency (171 kHz) of the two identical \( z \)-piezoelectric actuators (gray lines in Fig. 28). The resonant frequency of 171 kHz is derived from the fact that one end of each piezoelectric actuator is attached to the base (410/2 = 205 kHz) and the other end is attached to a sample stage (or a dummy stage). Higher resonant frequencies also appear at around 350 kHz.

Fig. 28. Frequency spectra of mechanical response of the \( z \)-scanner. (a) Gain spectra, (b) phase spectra. Gray lines represent the response without feedback Q-control, and black lines represents the response with feedback Q-control.
Judging from the phase spectrum (gray line in Fig. 28b), the two main resonant elements are connected in parallel. For the active damping of these resonant vibrations, we employ the simple Q-control method (Fig. 19a) described in Section 6.6.2. In addition, to compensate for a phase delay caused by active damping, a \((1 + \text{derivative})\) circuit is inserted between the Q-controller and the piezodriver. Although this method does not work perfectly, as mentioned in the Section 6.6.2, we use a mock z-scanner with two resonator circuits connected in parallel. This gives a better result compared with the use of a mock scanner with a single resonator. Using this method of Q-control, the first resonance completely disappears (black line in Fig. 28a) and the phase delay is not significantly deteriorated (black line in Fig. 28b). As the quality factor is reduced significantly, the z-scanner response time is markedly improved from 17.1 \(\mu\text{s}\) to 0.93 \(\mu\text{s}\) (Fig. 29). This improvement in the response speed of the z-scanner is more clearly confirmed by the cantilever deflection responded to the stepwise variation of the set point level of the PID controller (Fig. 30). The set point variation corresponds to a 100 nm height change. Without active damping, the response is very slow (Fig. 30a). Using the active damping and phase compensation by \((1 + \text{derivative})\) operation, the input rectangular signals are well restored in the cantilever deflection signal (Fig. 30b). The slight ringing behavior observed at the rising and falling edges are due to the remaining resonances at around 350 kHz. These resonances could be removed using a notch filter or a low-pass filter. We do not remove them, since such filtering increases the phase delay. In practice, these remaining resonances negligibly disturb

---

Fig. 29. Effect of active damping on the transient response of the z-scanner. In (a), the z-scanner has its natural \(Q = 9\), while in (b) the z-scanner is actively damped so that its \(Q\) is reduced to 0.5. The z-scanner began to be oscillated at \(t = 100 \mu\text{s}\) by ac signals (lines (i)) of 150 kHz. The amplitude of the ac input signals was attenuated when operated without active damping. The enveloping curves (gray lines) of the z-scanner’s displacements (lines (ii)) indicate the transient changes in the amplitude of the z-scanner’s oscillation.
the imaging of protein molecules, even at an imaging rate of 30 ms/frame over a scan range of 240 nm with 100 scan lines. This is understandable from the relationship \( f = V_s/s \) (see Section 5.1). Under these imaging conditions, the maximum feedback frequency is estimated to be about 200 kHz when the apparent feature size of the sample protein molecules is about 8 nm. Therefore, the \( z \)-scanner is not excited even at an imaging rate of 30 ms/frame.

We recently attempted to develop other types of scanner in which a \( z \)-piezoelectric actuator is held either at the four side rims parallel to the displacement direction [99] (method depicted in Fig. 15d) or at the rims and corners of a plane perpendicular to the displacement direction (method depicted in Fig. 15c). The gain and phase spectra of the mechanical response of the former \( z \)-scanner are shown with the blue lines in Fig. 31a and b. The \( z \)-scanner exhibited large resonant peaks at 440 kHz and 550 kHz. The resonant frequency of 440 kHz is very similar to that of the free-oscillation of the piezoelectric actuator. Judging from the phase spectrum, the two resonators are connected in series. We used two feedback active damping circuits connected in series (see Eq. (22)). The resulting gain and phase spectra are shown with the red lines in Fig. 31a and b, respectively. The peak at 440 kHz is almost completely removed and the frequency that gives a 90° phase delay reaches 250 kHz.

The gain and phase spectra of the mechanical response of the latter \( z \)-scanner (a \( z \)-piezoelectric actuator was held as depicted in Fig. 15c) are shown with blue lines in Fig. 31c and d. The \( z \)-scanner exhibited large resonant peaks at 370 kHz and 540 kHz. The reso-
nant frequency of 370 kHz is similar to that of the free-oscillation of the piezoelectric actuator (410 kHz). Judging from the phase spectrum, the two resonators are connected in parallel. Thus, we applied feed-forward active damping using a double-loop APITF circuit containing a mock scanner composed of two LRC circuits connected in parallel. The resulting gain and phase spectra are shown with the red lines in Fig. 31c and d, respectively. The peak at 370 kHz is completely removed and the frequency that gives a 90° phase delay is increased from 370 kHz to 500 kHz. A problem that we have not been able to resolve is the occurrence of small resonances around 50–80 kHz. This is likely to arise from that holding the \( z \)-piezoelectric actuator is not tight enough.

Among various methods of holding a \( z \)-piezoelectric actuator (Fig. 15), the method depicted in Fig. 15d [99] thus far resulted in the best performance regarding the high-frequency of the first resonance, mechanical stability and small crosstalk between the \( x \)- and \( z \)-axes. The maximum displacement (1 \( \mu m \)) is sufficient for imaging biological macromolecules. Since the first resonant frequency is sufficiently high, it will not be excited even at video-rate scanning. Therefore, we do not have to use complicated active damping methods to completely eliminate the resonant bands. However, it seems the best to remove

Fig. 31. Frequency spectra of mechanical response of two types of \( z \)-scanners. (a) and (b) are, respectively, the gain and phase spectra of a \( z \)-scanner whose piezoelectric actuator is held at the four rims parallel to the displacement direction (see Fig. 15d). (c) and (d) are, respectively, the gain and phase spectra of a \( z \)-scanner whose piezoelectric actuator is held at the four corners of a plane perpendicular to the displacement direction (see Fig. 15c). The blue lines represent the responses without feed-forward active damping, and the red lines represent the responses either with feedback Q-control ((a),(b)) or feed-forward active damping ((c),(d)). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
resonant peaks at high frequencies, so that the high gain margin of feedback control for the constant tip–sample distance is attained.

6.7. Dynamic PID control

To observe biological processes containing delicate protein–protein interactions, the tip–sample interaction force has to be minimized. The tapping force exerted on the sample by the oscillating cantilever tip can be reduced by using an amplitude set point \( r \) close to 1. As described in Section 5.4, this set point condition induces parachuting, during which the error signal saturates. The saturated error signal becomes smaller as \( r \) increases, and hence, the parachuting time \( \tau_p \) is prolonged. The feedback gain cannot be increased to shorten \( \tau_p \), since a larger gain induces overshoot at uphill regions of the sample, resulting in the instability of the feedback operation. In this section, we describe a new PID controller ("dynamic PID controller") developed to solve this problem [71].

6.7.1. Dynamic PID controller

With conventional PID controllers, the gain parameters are constant, irrespective of whether scanning is underway at uphill regions or downhill regions of the sample. To avoid parachuting or to shorten the parachuting period, the gain parameters should be larger at downhill regions than at uphill regions. The amplitude signal is an indicator of the type of region that is being scanned. Generally, at uphill regions the amplitude is smaller than the set point amplitude, while it is larger at downhill regions. Thus, with the dynamic PID controller that we developed, the gain parameters are automatically altered depending on the cantilever oscillation amplitude. A threshold level \( A_{\text{upper}} \) is set between the set point peak-to-peak amplitude \( A_s \) and the free-oscillation peak-to-peak amplitude \( 2A_0 \) (or is set to \( A_s \)). We usually use a value of \( A_{\text{upper}} \) that is much closer to \( A_s \) than \( 2A_0 \), the reason for which is mentioned later. When the cantilever peak-to-peak oscillation amplitude \( A_{p-p} \) exceeds \( A_{\text{upper}} \), the gain parameters are increased. There are several methods of imple-

![Diagram of feedback loop with dynamic PID operator](image-url)
menting this automatic alteration in the gain parameters. We insert a circuit called a “dynamic operator” between the error signal output and the input of a conventional PID circuit (Fig. 32). This dynamic operator functions as follows (see Fig. 33). When \((A_{p-p} - A_{upper}) > 0\), the differential signal \((A_{p-p} - A_{upper})\) is amplified and added to the error signal. The error signal that contains this extra signal is fed to the conventional PID. This “false error signal”, whose value is larger than that of the “true error signal”, produces a quicker feedback response, and therefore, \(A_{p-p}\) quickly becomes smaller than \(A_{upper}\) and the feedback operation automatically returns to the normal mode. Thus, even when \(A_s\) is very close to \(2A_0\), the parachuting period is shortened markedly, and hence, practically no feedback saturation occurs. When \(A_{upper}\) is far from \(A_s\), the dynamic operator produces an abrupt jump in the error signal. This is the reason why we use a value of \(A_{upper}\) that is much closer to \(A_s\) than \(2A_0\). Noted that to ensure that this dynamic PID controller is effective even when the set point \(r\) is very close to 1, the amplitude signal should not contain noise greater than \(\sim 2A_0 (1 - r)/3\).

Similar manipulation of the error signal can also be performed when \(A_{p-p}\) is smaller than \(A_s\). In this case, a new threshold level \(A_{lower}\) is set with a value sufficiently lower than \(A_s\). When \(A_{p-p}\) becomes lower than \(A_{lower}\), the differential signal \((A_{p-p} - A_{lower})\) is amplified and then added to the error signal \((A_{p-p} - A_s)\). This can prevent the cantilever tip from pushing into the sample too strongly, particularly at steep uphill regions of the sample.

These manipulative operations are implemented using the circuit shown in Fig. 34. The circuit has three branches in the horizontal direction. The true error signal passes through the middle branch. A DC signal corresponding to \((A_{upper} - A_s)\) is fed to the input terminal of the upper branch and passes through a subtracter to produce \((A_{p-p} - A_{upper})\). The output is amplified (with a gain of \(G_{upper}\)), passed through a precision diode circuit, and finally added to the true error signal. The precision diode circuit compensates for an offset inherent in the diode chip. The final output, \(L[A_{p-p} - A_{upper}] \times G_{upper} + (A_{p-p} - A_s)\), is fed to the conventional PID input terminal. Here, \(L\) is an operator such that \(L[x] = x\) (if \(x > 0\)) or \(L[x] = 0\) (if \(x < 0\)). The signal \((A_{p-p} - A_{upper})\) can be constructed directly from

Fig. 33. Schematic showing the principle of the dynamic PID control. Solid line: an amplitude-distance curve; gray line: an error signal used in the conventional PID control; and broken line: an error signal used in the dynamic PID control [71].
$A_{\text{p-p}}$ and $A_{\text{upper}}$. However, in this case, $A_{\text{upper}}$ has to be tuned depending on the value of $A_s$, because $A_{\text{upper}}$ has to be larger than $A_s$. On the other hand, using the method described above, a positive DC signal corresponding to $(A_{\text{upper}} - A_s)$ is tuned independently of the value of $A_s$. The false error signal, $-L[A_{\text{lower}} - A_{\text{p-p}}] \times G_{\text{lower}} + (A_{\text{p-p}} - A_s)$, is similarly constructed in the lower branch.

Fig. 34. Circuit diagram of dynamic operator [71]. For details, see the text.

Fig. 35. Feedback bandwidth as a function of the set point $r$, measured using the mock AFM system. Solid curves: feedback bandwidths measured using the mock AFM system with a conventional PID controller; dotted curves: feedback bandwidths measured using the mock AFM system with the dynamic PID controller. The solid curves and the dotted curves are aligned from top to bottom according to the ratio $2A_0/h_0 = 5$, 2, 1, and 0.5 [71].
6.7.2. Performance of dynamic PID control

Dynamic PID control significantly enhances the feedback bandwidth, particularly when the set point \( r \) is close to 1 (see dotted curves in Fig. 35). The feedback bandwidth is independent of \( r \), provided \( r \) is less than \( \sim 0.9 \), indicating that parachuting does not occur. The superiority of the dynamic PID control is also clear from captured images of a mock sample with steep slopes (Fig. 36a). The images were obtained using a mock AFM. Herein, a mock cantilever with \( Q = 3 \) oscillating at its resonant frequency of 1.2 MHz is scanned over a mock sample surface (rectangular shapes with two different heights) from left to right at a scan speed of 1 mm/s (frame rate of 100 ms/frame). Here, the height of the taller rectangle is \( 2A_0 \), and \( A_s \) is set at \( 0.9 \times 2A_0 \). When using the conventional PID controller, the topographic image became blurred (Fig. 36b) due to significant parachuting at steep downhill regions. On the other hand, the use of the dynamic PID controller resulted in a clear image (Fig. 36c), indicating that almost no parachuting occurred. Dynamic PID control improves both the feedback bandwidth and the tapping force simultaneously, and is therefore indispensable in the high-speed AFM imaging of delicate samples. This is demonstrated by the high-speed imaging of fragile biological samples (Chapter 9).

6.8. Drift compensator

The drift in the cantilever excitation efficiency poses a problem, particularly when \( r \) is close to 1 (i.e., \( A_s \) is set close to \( 2A_0 \)). For example, as the efficiency is lowered, \( 2A_0 \) decreases and, concomitantly, \( A_{p-p} \) decreases. The feedback system interprets this decrease in \( A_{p-p} \) as an overly strong interaction of the tip with the sample and therefore withdraws the sample stage from the cantilever; this is, of course, an incorrect operation. This withdrawal tends to dissociate the cantilever tip completely from the sample surface, particularly when \( A_s \) is close to \( 2A_0 \). For example, when \( 2A_0 = 5 \) nm and \( A_s = 4.5 \) nm \((r = 0.9)\), their difference is only 0.5 nm. The satisfactory operation of dynamic PID control under such a condition requires high stability of the excitation efficiency. The stabilization of the

![Fig. 36. Pseudo-AFM images of a sample with rectangles with two different heights. (a) A mock AFM sample. The images were obtained using a conventional PID controller (b) or using the dynamic PID controller (c). These simulations with the mock AFM system were made under the following conditions: the cantilever resonant frequency, 1.2 MHz; quality factor of the cantilever oscillation, 3; the resonant frequency of the \( z \)-scanner, 150 kHz; and quality factor of the \( z \)-scanner, 0.5; the line scan speed, 1 mm/s; the line scan frequency, 1 kHz; the frame rate, 100 ms/frame, the ratio of \( 2A_0 \) to the total sample height = 1; and \( r = 0.9 \) [71].](image-url)
excitation efficiency was previously attempted by Schiener et al. by using the second-harmonic amplitude of the cantilever oscillation to detect drifts [100]. The second-harmonic amplitude is sensitive to the tip–sample interaction, and therefore, the drift in $A_0$ is reflected in the second-harmonic amplitude averaged over a period longer than the image acquisition time. Instead of controlling $A_0$, they controlled $A_s$ to maintain a constant value of the difference ($2A_0 - A_s$). However, this control method varies the tapping force and feedback bandwidth because the sample height increases relative to the oscillation amplitude. To compensate for drift in the cantilever excitation efficiency, we also used the second-harmonic amplitude of cantilever oscillation, but instead of controlling $A_s$, we controlled the output gain of a wave generator (WF-1946A, NF Corp., Osaka, Japan) connected to the excitation piezoelectric actuator (Fig. 37) [71]. We only used an I-controller whose time constant was adjusted to 1–2 s (about 10 times longer than the image acquisition time). The performance of this method of drift compensation is shown in Fig. 38. A sample of myosin V bound to actin filaments in solution was imaged successively for 3 min at 100 ms/frame. Very stable imaging was achieved, even with the small difference

![Fig. 37. Circuit diagram of compensator for drift in cantilever excitation efficiency. The peak frequency of the band-pass filter is adjusted at the second harmonic frequency. Instead of using the band-pass filter and RMS-to-DC converter, we usually use a lock-in amplifier.](image)

![Fig. 38. Successive imaging of myosin V attached to actin filaments using a compensator for drift in the cantilever-excitation efficiency. The imaging was successively made for 3 min at frame rate of 10 frames/s and with $A_0 = 2.5$ nm and $r = 0.92$. Only five images obtained at times indicated by arrows are shown. Black line: the output from the drift compensator; gray line: the second harmonic amplitude. At 3 min, the compensator was switched off [71].](image)
(2A₀ - Aₛ) = 0.4 nm. In addition, the fragile actin filaments did not disassemble during imaging. The output signal from the slow controller (I-controller) increased with time, indicating that the cantilever excitation efficiency decreased with time. On the other hand, the second-harmonic amplitude remained constant. When the slow controller output was disconnected at 3 min, no image was obtained, because of the complete detachment of the tip from the sample.

6.9. Line-by-line feed-forward control

As described in Sections 5.4–5.6, one of the limiting factors in the feedback bandwidth of AFM is the ratio of \((2A₀ - Aₛ)[=2A₀(1 - r)]\) to the maximum sample height \(h₀\). The difference \((2A₀ - Aₛ)\) between the peak-to-peak free-oscillation amplitude and the set point peak-to-peak amplitude should be increased to increase the feedback bandwidth. On the other hand, it is necessary to bring the set point amplitude \(Aₛ\) close to the free-oscillation amplitude \(2A₀\) to reduce the loading force of the tip acting on the sample. Thus, the two aspects of soft tapping and fast imaging are inherently in conflict. We have presented a solution to this problem by introducing the dynamic PID controller [71], as described in Section 6.7. An additional approach to solving this problem is to use a two-degrees-of-freedom (2-DOF) controller. This controller consists of two controllers; one is a conventional PID feedback controller (or dynamic PID controller) and the other is a feed-forward controller.

Fig. 39. Block diagram of two-degrees-of-freedom (2-DOF) control. Feedback and feed-forward controllers are combined [101].
open-loop controller that moves the sample stage according to the sample height variation recorded in the previous line scan [93,101]. The feedback controller thus only compensates for the height difference between the previous and current scan lines.

6.9.1. Feed-forward controller

Fig. 39 shows a block diagram of the 2-DOF control scheme that we developed. The feed-forward controller consists of 12 bit D/A and A/D converters, two memory sets and two high-speed analog switches. The feed-forward controller starts to record the input signal into memory 1, which is triggered by the start signal from the scan controller. This recording is synchronized with the pixel clock signals. After the clock number reaches the preset number (i.e., the number of pixel × integer in one scan line), the recorded signal is then output from memory 1, and the input signal is simultaneously recorded into memory 2. These processes are repeated until one frame is acquired. The signal fed to the feed-forward controller input is an appropriately weighted sum of the PID feedback output signal and the error signal. Note that the error signal is also used to generate the feed-forward signal. In aqueous environment, the $Q$-value of the cantilever is not large, and is typically less than 3. Therefore, the error signal reflects the deviation of the tip trajectory from the actual topography. The use of this error signal to generate the feed-forward signal is essential for compensating the feedback delay. The PID output to be fed to the feed-forward controller should be attenuated (the gain is usually set at ~0.5). This is because, when the surface geometry is significantly different between two adjacent lines, the feed-forward operation would deteriorate the feedback operation rather than alleviate the load. Regarding the error signal to be fed to the feed-forward controller, its input gain can only be adjusted by considering the relationship between the OBD sensor sensitivity and the piezoelectric displacement efficiency.

6.9.2. Performance of feed-forward compensation

To quantitatively evaluate the performance of the feed-forward compensation, we used the mock AFM circuit which is described in Section 5.6. For this evaluation, the oscillation frequency and amplitude of the mock cantilever were set at 800 kHz and 0.1 V, respectively. A sinusoidal wave with a frequency of 7.3 kHz and an amplitude of 0.1 V was used as the mock sample topography.

Fig. 40 shows a comparison of pseudo-AFM images produced by the mock AFM circuit with and without feed-forward compensation. Here, the amplitude set point $r$ was set at 0.98. The scan direction was from left to right. As shown in Fig. 40a and c, the image obtained without feed-forward compensation exhibits a distorted sinusoidal shape due to the parachuting of the tip. On the other hand, as shown in Fig. 40e and g, images obtained using feed-forward compensation have a sinusoidal shape without appreciable distortion. The height of the sinusoidal shape in the image obtained using the feed-forward compensation is 1.8 times larger than that obtained without feed-forward compensation. In addition, as shown in Fig. 40d and h, the feed-forward compensation reduces the error signal by 25%.

To confirm the applicability of the feed-forward compensation, we imaged soft biomolecules. We used microtubules as a test sample. Tubulins were purified from a pig brain and assembled into microtubules in a buffer solution by adding glycerin, guanosine-5'-triphosphate (GTP), and taxol at 37 °C. The microtubules were adsorbed on a mica surface and then the AFM experiment was performed in the same buffer solution containing taxol.
Fig. 41a and b shows topographic images taken with and without feed-forward compensation, respectively. These images were acquired at 1 s/frame, with an amplitude set point of 0.9 and a free peak-to-peak amplitude of 6 nm. The x-scan direction was from left to right on the images. In the image acquired using the feed-forward compensation (Fig. 41a), a single microtubule with a height of 25 nm as well as isolated protofilaments can be clearly observed. At the scanning position marked by an arrow in Fig. 41b, the feed-forward compensator was switched off and thereafter only the feedback control was used. Without the feed-forward compensation, the right-hand side of the microtubule shows tip parachuting and the image quality markedly deteriorates. Furthermore, the isolated protofilaments on the mica surface are not observed. This is because the ratio of the microtubule height of 25 nm to the peak-to-peak free-oscillation amplitude of 6 nm is too large compared with \( \frac{1 - 0.9}{0.9} \times 6 = 0.6 \) nm (see Eq. (10)), and therefore, the feedback bandwidth is not sufficient to precisely trace the topography. When the set point amplitude was reduced without feed-forward compensation to improve the image quality, the micro-

Fig. 41. AFM images of a microtubule on a mica surface obtained (a) with and (b) without the feed-forward compensation. In (b), the feed-forward control was cut at the position marked by an arrow. In (a) and (b), the amplitude set point \( r \) for feedback control was 0.9 and the peak-to-peak free-oscillation amplitude was 6 nm. In (c), the set point was increased to 0.95 without the feed-forward compensation. All images were acquired at 1 s/frame. The scan size was 200 \( \times \) 200 nm\(^2\) with 128 \( \times \) 128 pixels.
tubule was completely disassembled. In Fig. 41c, the amplitude set point \( r \) was increased to about 0.95 without the feed-forward compensation. It can be seen that the microtubule is disassembled at the positions where the feed-forward compensator was switched off (as indicated in Fig. 41b). From these results, feed-forward compensation is effective for improving the feedback bandwidth and reducing the loading force on the sample.

6.10. Photothermal actuation of cantilevers

There is a limit to the enhancement of the bandwidth of a piezoelectric actuator-based \( z \)-scanner due to the limited performance of available piezoelectric actuators. The upper limit seems to be \( \sim 500 \) kHz in practice, considering the required maximum displacement and the effect of the mass of the attached sample stage on the resonant frequency. The tip–sample distance can be controlled not only by the \( z \)-scanner but also by direct cantilever actuation. As small cantilevers have a high resonant frequency in water, the direct actuation of a small cantilever is expected to afford a high bandwidth for the control of tip–sample distance. At present, small self-actuation cantilevers cannot be manufactured. There have been several reports describing the direct actuation of cantilevers coated with or attached to ferromagnetic materials using a magnetic field [102,103]. Unfortunately, the magnetic coating stiffens the cantilevers and the attachment of magnetic materials lowers the cantilever resonant frequency. In addition, it is difficult to produce a high-frequency magnetic field generator with high bandwidths. Acoustic radiation pressure (ARP) generated by a focused acoustic transducer with a frequency of 100–300 MHz has recently been used for the direct actuation of a cantilever in liquids [104]. The force generated by the ARP can be focused on an area of a few microns in diameter, and therefore, can be used for the direct actuation of a small cantilever. However, the ARP also exerts forces on the sample, and therefore, a special arrangement is required for the implementation of the sample stage and the transducer [105,106]. Also, although it was not mentioned in these articles, it is likely that the sample solution temperature is increased by the ARP.

Another technique for direct cantilever actuation is to use a photothermal bending effect. The irradiation of a laser beam actuates a gold-coated cantilever either by the pressure induced by the radiation or by photothermal expansion [107–110]. In liquids, the latter effect is dominant because of the low quality factor of cantilevers. We used a 980 nm IR laser for the actuation and a 675 nm red laser for detecting the deflection [111]; these laser beams were focused onto a small cantilever using an objective lens. The irradiation position of the IR laser onto the cantilever was adjusted so that the maximum cantilever deflection was attained. The DC deflection efficiency was about 1.1 nm/mW; the laser power was not that actually irradiated onto the cantilever but that measured at the exit of the objective lens. The power of the IR laser can be modulated from 0 to 200 mW with a frequency range of 0–10 MHz, and therefore, the maximum deflection of the cantilever should have been 220 nm. In practice, the deflection was 120 nm due to the partial eclipse of the incident laser light at the objective lens inlet. This deflection range is sufficient for imaging protein samples. We attempted to use a 405 nm violet laser for the actuation and obtained a larger deflection efficiency of 10 nm/mW. However, the biological samples were damaged by the irradiation.

As expected, the photothermal bending response of the cantilever to the laser power modulation was slow. The cantilever response to the modulation with a rectangular wave was well fitted by a double-exponential function with time constants of 8.4 ms and 123 ms.
This slow response is also apparent in the frequency response (curves marked with (b) in Fig. 42). As shown with the curves marked with (a) in Fig. 42, this slow response was successfully eliminated by inverse compensation, using an APITF circuit [111](Fig. 24b). We examined how a high feedback bandwidth can be achieved by controlling the tip–sample distance using the present photothermal bending method. Under a condition of $2A_0/h_0 = 5$, the feedback bandwidth of 100 kHz was obtained for a small cantilever having a resonant frequency of 1.2 MHz and a quality factor of $\sim 3$ in water [111]. This feedback bandwidth is 43% higher than the case when the tip–sample distance is controlled by our piezoelectric actuator-based $z$-scanner with a resonant frequency of 170 kHz and a quality factor of 0.5. Because of this high feedback bandwidth, myosin V molecules were clearly imaged at a rate of 32 frames/s over a scan range of $240 \times 240$ nm$^2$ with 100 scan lines [111].

Here, note that there is a drawback to direct cantilever actuation when it is used to control the tip–sample distance. The cantilever has to be deflected up to the maximum sample height. Therefore, the dynamic range in the OBD sensor has to be large, resulting in lower sensor sensitivity. This is independent of the means used for the direct cantilever actuation. When a piezoactuator-based $z$-scanner is used to control the tip–sample distance, the DC deflection of the cantilever is smaller than the maximum sample height.

6.11. Data acquisition system

High-speed data acquisition (DAQ) systems are commercially available. We have been using DA/AD boards (10 MHz, 12 bits) with a memory of $\sim 0.5$ MB/channel and direct memory access (DMA) controller chips, together with software that we have created. DMA allows high-speed data transfer without imposing a heavy load on the computer CPU. We can attain a minimum image acquisition time of $\sim 3$ ms using the DA/AD

![Fig. 42. Frequency spectra of the gain (solid lines) and phase (broken lines) for a small cantilever excited by the intensity-modulated IR laser (a) with and (b) without inverse compensation. Insert: The time–domain response of the cantilever displacement driven by the laser modulated with a rectangular wave, without inverse compensation. The response shown with dots is fitted by a double-exponential-function curve (solid line), $y(1 - \exp(-t/\tau_1)) + \beta(1 - \exp(-t/\tau_2))$ with time constants of $\tau_1 \sim 1.84 \mu$s and $\tau_2 \sim 123 \mu$s [111].](image)
boards. As DAQ systems with higher rates (100 MHz, 12–16 bits) are available, faster frame acquisition is possible. However, it takes time to display a bitmap image, even when data are transferred directly to a bitmap memory area. For example, it takes about 6–10 ms to display a 256 × 256 pixel image. To avoid a time lag between frames caused by this delay, we have to reduce the number of frames to be displayed. Large amounts of data stream out of high-speed imaging. The maximum number of acquirable successive frames depends on the size of the computer memory. For a typical memory size of 2 GB for a personal computer, approximately 20,000 successive images with 12-bit resolution and 256 × 256 pixels can be recorded (when memory size required by the operational system is ignored). For a frame rate of 30 ms/frame, we can record continuously for 10 min. Quantitative descriptions of the AFM image acquisition abilities of a given DAQ system are given by Fantner et al. [112,113].

7. High-speed phase detector

One of the unique features of tapping-mode AFM, which is the operation mode of our high-speed AFM, is the capability of simultaneously imaging the topography and compositional variations of a sample surface [6–8,114–117]. The distribution of material properties is measured by detecting the phase difference between the excitation signal and the cantilever oscillation. The phase difference is related to several material properties such as viscoelasticity, elasticity, adhesion, hydrophobicity/hydrophilicity [6], and surface charge [117]. The energy dissipation of an oscillating cantilever due to inelastic tip–sample interactions is considered to be the main mechanism in the generation of phase-contrast [7,118]. Combining the information on these properties with a topography with unprecedented temporal resolution would provide a deeper insight into functional mechanisms and dynamic processes of biomolecules. In this chapter, a novel method of realizing high-speed phase detection and some of its applications are described [119].

For high-speed phase-contrast imaging, the phase difference must be detected with a sufficiently higher bandwidth than the feedback bandwidth used for topographic imaging. A lock-in amplifier, which is usually used in the phase-contrast imaging with a conventional AFM, cannot be applied to high-speed imaging as its bandwidth is too low; it is usually limited to less than kHz even without the use of filters. To achieve fast phase detection, a novel phase detector [119] was developed on the basis of a design by Stark and Guckenberger [120]. Fig. 43 shows the operational principle of the fast phase detection. A two-channel wave form generator produces sinusoidal and sawtooth waves with a given phase difference. The sinusoidal wave is used to oscillate the cantilever. The cantilever oscillation signal is fed to a phase detector composed of a high-pass filter, a variable phase shifter, a zero-crossing comparator, a pulse generator, and a sample-and-hold (S/H) circuit. A pulse signal is generated at either the rising or falling edge of the output signal of the zero-crossing comparator. The pulse signal acts as a trigger for the S/H circuit. When the trigger is generated, the amplitude voltage of the sawtooth signal is captured and retained by the S/H circuit. Thus, the sawtooth signal acts as a phase-voltage converter. The variable phase shifter enables us to vary the timing of the trigger signal generated within the cantilever oscillation cycle. This function is essential for obtaining the optimum phase-contrast because the phase-contrast markedly depends on the trigger-timing, as described in Section 11.1.2.
Fig. 44 shows a marked difference in the phase detection speed between a conventional RF lock-in amplifier and the fast phase detector. The input signal is a sinusoidal wave with a frequency of 900 kHz and phase responses measured using (a) a lock-in amplifier and (b) the fast phase detector. The phase of the input signal is shifted stepwise from 0° to 90° at time zero. Rectangular waves depicted by broken lines in (b) indicate the trigger signals [119].
a frequency of 900 kHz. The phase of this sinusoidal wave was shifted stepwise from 0° to 90° at time zero. Using the lock-in amplifier, the input signal oscillates over 20 cycles until the output phase reaches 90°, as shown in Fig. 44a. The time constant $\tau$ measured from this transient response was evaluated to be 4.5 $\mu$s, which corresponds to a bandwidth of about 35 kHz. This bandwidth is only 1/2–1/3 of the feedback bandwidth of our high-speed AFM. On the other hand, the fast phase detector responded much rapidly as shown in Fig. 44b. The phase of the input signal was changed at the point indicated by “A”. The trigger signal for the S/H circuit is shown by broken lines. The time constant $\tau$ was evaluated to be 0.12 $\mu$s, which corresponds to a bandwidth of about 1.3 MHz. This is much higher than the feedback bandwidth of our high-speed AFM. The detected phase shift did not change stepwise. This is due to the limited bandwidth of the operational amplifiers used after the S/H circuit.

The main source of noise of the fast phase detector is jitter originating from voltage noise contained in the PSPD output. The RMS value of jitter $\Delta t_{RMS}$ for a sinusoidal wave containing broadband white noise $V_{RMS}$ is given by

$$\Delta t_{RMS} = \frac{V_{RMS}}{2\pi A f_{osc}}.$$

(31)

Here, $A$ and $f_{osc}$ are the amplitude and frequency of the sinusoidal wave, respectively. The jitter gives rise to timing fluctuations in the trigger pulse for the S/H circuit and therefore, the phase error $\Delta P_{RMS}$ (degrees) is given by

$$\Delta P_{RMS} = \frac{V_{RMS}}{2\pi A} \times 360^\circ = \frac{360^\circ}{2\pi(SNR)}.$$

(32)

where SNR is the signal-to-noise ratio of the sensor output. For the test signal shown in Fig. 44b, SNR is about 17 ($A = 0.45$ V and $V_{RMS} = 26$ mV). This SNR gives a phase noise of 3.3°, which is in good agreement with the magnitude of phase noise (3°) evaluated from the phase signal shown in Fig. 44b. In the actual experiment, SNR is less than 10. Therefore, the intrinsic phase noise is expected to be larger than 5.7° due to the thermal fluctuations of the cantilever under a low-$Q$ condition. However, since the bandwidth of 1.3 MHz is much higher than the imaging bandwidth (ca. 100 kHz), the phase noise can be reduced using a low-pass filter.

8. Devices for atomic-resolution in liquids

The understanding of biological phenomena at the molecular level requires the direct imaging of individual mobile molecules in real-time as well as in real space under physiological environment. Thus, for bio-AFM, improvement of the spatial resolution is also important. Recently, there has been a great progress in improving the spatial resolution for dynamic-mode in-liquid AFM [121,122]. In this chapter, we describe the key features of this novel technique [76,123]. Its applications to biological studies are described in Chapter 12.

8.1. Operation mode

For AFM applications in liquid, contact-mode AFM [1] and AM–AFM [124] have mainly been used. Contact-mode AFM has been used for the subnanometer-scale imaging...
of proteins [34] and lipid molecules [125] in model biological membranes, while AM–AFM has mainly been used for the nanometer-scale imaging of isolated biomolecules [126,127] and cells [35] attached to a substrate. However, the recent improvement in spatial resolution was brought by employing FM–AFM [9] that has mainly been used for atomic-scale imaging in ultra high vacuum (UHV) environments. In this section, we describe the basic principles and experimental setup for FM–AFM and its distinctive features compared with other operation modes.

Fig. 45a shows a typical experimental setup for FM–AFM. A cantilever is mechanically oscillated at its resonance frequency using a self-oscillation circuit. The cantilever vibration is induced by an actuator next to the cantilever. The cantilever deflection is typically detected with an OBD sensor. The deflection signal output from the pre-amplifier is fed into a phase shifter which adjusts the phase of the signal so that the phase difference between the deflection and excitation signals (\(\phi\)) is kept at \(-90^\circ\). The output from the phase shifter is amplified by an automatic gain control (AGC) circuit to keep the cantilever oscillation amplitude (\(A\)) constant. The output from the AGC circuit is routed back to the actuator to form a self-oscillation loop.

As the cantilever approaches the surface, the tip–sample interaction force (\(F_{ts}\)) induces a resonance frequency shift (\(\Delta f_c\)) given by [128]

\[
\Delta f_c(z) = \frac{f_c}{2k_c} \int_{-A}^{A} \frac{\partial F_{ts}}{\partial z}(z + \Delta z) \frac{\sqrt{A^2 - \Delta z^2}}{\pi A^2/2} d(\Delta z),
\]

where \(k_c\) is the cantilever spring constant and \(z\) is the vertical tip position. This behavior is schematically shown in Fig. 45b. Eq. (33) shows that \(\Delta f_c\) is proportional to the force gradient integrated over the cantilever oscillation range with a weight function. If \(A\) is small enough for the force gradient to be assumed constant, Eq. (33) can be simplified to [9]

\[
\Delta f_c(z) = -\frac{f_c}{2k_c} \frac{\partial F_{ts}}{\partial z}.
\]
As the amplitude vs. frequency curve is shifted, the phase vs. frequency curve is also shifted by $\Delta f_c$. Since $\phi$ is kept constant at $-90^\circ$ with the self-excitation circuit, the cantilever oscillation frequency ($f$) changes with the change in resonance frequency ($f_c$), as shown in Fig. 45b. Thus, $\Delta f_c$ is measured by directly measuring the frequency change of the cantilever vibration. In the FM–AFM setup, the deflection signal is fed into a frequency detector which outputs a voltage signal proportional to the frequency of the input. A phase-locked loop (PLL) circuit is most commonly used as the frequency detector. The $\Delta f_c$ signal is fed into a PI controller which controls the vertical position of the sample to regulate the tip–sample distance.

The advantages of using FM–AFM over contact-mode AFM include the precise control of the vertical tip position without instabilities especially in the short-range interaction regime, the capability of imaging isolated molecules weakly bound to a substrate, and enhanced sensitivity to the short-range interaction force [129] and hence a higher spatial resolution [130]. The advantages of FM–AFM over AM–AFM are still under discussion. One of the clear advantages of the former is the capability of the quantitative and independent measurement of conservative and dissipative forces [131]. Practically, the frequency signal in FM–AFM changes much more rapidly than the amplitude signal in AM–AFM, which facilitates the precise control of the tip position near the surface [121]. This does not necessarily mean that FM–AFM has a higher signal-to-noise ratio in the force measurement but does explain why FM–AFM has been used for reproducible atomic-scale imaging both in vacuum [10] and liquids [132].

8.2. Cantilever stiffness

In contact-mode AFM and AM–AFM for biological applications, relatively soft cantilevers (typically $k_c < 0.1 \text{ N/m}$) have been used while stiff cantilevers (typically $k_c > 10 \text{ N/m}$) have been used for high-resolution imaging with FM–AFM. In this section, we describe the necessity of using a stiff cantilever and its merits and demerits.

The RMS amplitude spectral density ($n_{z_B}$) of cantilever Brownian vibration induced by thermal energy is given by [9]

$$n_{z_B} = \frac{1}{\sqrt{\pi f_c k_c Q_c}} \frac{1}{\left[1 - (f/f_c)^2\right]^2 + \left[f/(f_c Q_c)^2\right]^2},$$

where $k_B$, $T$, and $Q_c$ are Boltzmann’s constant, the absolute temperature and the $Q$ factor of the cantilever, respectively. The total RMS vibration amplitude $z_B$ induced by the thermal energy is obtained by integrating $n_{z_B}$ over the whole frequency range or directly from the equipartition theorem and given by [9]

$$z_B = \sqrt{\frac{k_B T}{k_c}}.$$  \hfill (36)

For cantilevers with $k_c$ of 0.1 N/m and 10 N/m, $z_B$ is about 0.2 nm and 0.02 nm, respectively. This thermal vibration amplitude ultimately determines the accuracy of control over the vertical tip position. Note that the thermal vibration can be reduced once the cantilever is in contact with the surface, such as in contact-mode AFM. However, delicate true atomic-resolution imaging requires precise control of the vertical tip position in the tip–
sample distance range from near contact to atomic-scale contact. Therefore, the high stiffness of the cantilever is essential for true atomic-resolution imaging as well as for atomic-scale manipulation.

The RMS frequency of noise spectral density \( n_f \) arising from the thermal cantilever vibration is given by [76]

\[
n_f = \sqrt{\frac{k_B T f_c}{\pi k_c Q_c A^2}}.
\]

(37)

This equation shows that \( n_f \) decreases with increasing \( k_c \). From Eqs. (34) and (37), the minimum detectable force gradient \( |\frac{\partial F_{ts}}{\partial z}|_{\text{min}} \) for small amplitude approximation is given by [9]

\[
|\frac{\partial F_{ts}}{\partial z}|_{\text{min}} = \sqrt{\frac{4k_c k_B T B}{\pi f_c Q_c A^2}},
\]

(38)

where \( B \) is the measurement bandwidth. If the force gradient is approximately constant, the magnitude of the conservative force \( (F_c) \) is proportional to \( A \). Thus, the minimum detectable conservative force is given by [133]

\[
(F_c)_{\text{min}} = \sqrt{\frac{4k_c k_B T B}{\pi f_c Q_c}}.
\]

(39)

This equation shows that \( (F_c)_{\text{min}} \) increases proportionally to \( \sqrt{k_c} \) with increasing \( k_c \). Practically, a cantilever with a higher value of \( k_c \) tends to have higher values of \( f_c \) and \( Q_c \), which partially compensate for the reduction of force sensitivity due to the high value of \( k_c \).

Table 1 shows the parameters of typical soft and stiff cantilevers and those of a small cantilever designed for high-speed AFM (Section 6.1). This table shows that the practical difference in the force sensitivity between soft and stiff cantilevers is not significant. Furthermore, the high values of \( f_c \) and \( Q_c \) greatly facilitate finding a correct cantilever resonance and stabilize the self-oscillation in FM–AFM.

The small cantilever has a high value of \( f_c \) in spite of the small value of \( k_c \). Thus, it should provide better force sensitivity than the other cantilevers with parameters shown in Table 1, although this has not yet been experimentally confirmed for FM–AFM. It should also allow the tip to be scanned more than 10 times faster than that of the other cantilevers. Note that this cantilever still suffers from problems such as large thermal vibration amplitude (Eq. (36)) and the instabilities described below. Therefore, it should

<table>
<thead>
<tr>
<th>Table 1</th>
<th>Comparison of various parameters in water between typical soft and stiff cantilevers and a small cantilever designed for high-speed AFM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( k_c ) (N/m)</td>
</tr>
<tr>
<td>Typical soft cantilever</td>
<td>0.1</td>
</tr>
<tr>
<td>Typical stiff cantilever</td>
<td>30</td>
</tr>
<tr>
<td>Small cantilever</td>
<td>0.2</td>
</tr>
</tbody>
</table>

\( k_c \), spring constant; \( f_c \), first resonant frequency; \( Q \), quality factor; \( (F_c)_{\text{min}} \), the minimum detectable conservative force; \( n_{z-B}(f_c) \), RMS amplitude spectral density of cantilever Brownian vibration at \( f_c \).
be used for applications where ultimate spatial resolution is not necessary but high force sensitivity and high scanning speed are essential.

As a tip approaches a surface, the attractive force between the tip and the sample increases. Thus, the tip experiences an increase in the force gradient. In contact-mode AFM, the tip suddenly jumps and comes in contact with the surface at a distance where the force gradient exceeds \( k_c \). Therefore, the following condition must be satisfied for the measurement of force without tip adhesion [134]:

\[
 k_c > \frac{\partial F_{us}}{\partial z}. \tag{40}
\]

Such adhesion is referred to as “jump-to-contact” and hinders precise control of interaction between the tip front atom and the surface atom or molecules. Since the operation of contact-mode AFM requires soft cantilever, the only way to meet this condition is to adjust the solution conditions such as adding salts [135]. This practically limits the application range.

In dynamic-mode AFM, the tip adhesion can be prevented by increasing \( A \) to meet the following condition [136]:

\[
 k_c A > |F_{us}|. \tag{41}
\]

Thus, a soft cantilever can be used without limitations in solutions, provided \( A \) is sufficiently large. However, the use of large amplitude practically limits the spatial resolution as explained below. Namely, high-resolution AFM imaging requires small amplitude operation without instabilities, which in turn requires high stiffness of the cantilever.

### 8.3. Oscillation amplitude

In AM–AFM, \( A \) has been set at 5–50 nm while the novel FM–AFM technique utilizes an amplitude of 0.1–1 nm. In this section, we describe the necessity of using small oscillation amplitude and the practical requirements for realizing this condition.

\( F_{us} \) includes long-range forces (e.g., electrostatic force and van der Waals force) and short-range forces (e.g., chemical interaction). To achieve true (i.e., non-averaged) atomic-resolution in AFM, the short-range interaction between the tip front atom and the surface atom or molecule must predominantly contribute to the distance feedback signal (i.e., \( z \) in contact-mode AFM, \( A \) in AM–AFM, \( \Delta f_c \) in FM–AFM).

The reduction of \( A \) in dynamic-mode AFM allows the tip to stay in the short-range interaction regime (typically \( z < 1 \) nm) for a longer time in an oscillation period. In the short-range interaction regime, the effect of the long-range interaction force on the force gradient is negligible compared with that of the short-range interaction force. Therefore, the use of small amplitude enhances sensitivity to the short-range interactions and reduces sensitivity to the long-range interactions, leading to a higher spatial resolution [129,130].

Noise in the distance feedback control (\( \delta z \)) in FM–AFM is determined by the ratio between the frequency noise \( \delta f \) and the slope of the \( \Delta f_c \) vs. distance curve

\[
 \delta z = \frac{\delta f}{\partial(\Delta f_c)/\partial z}. \tag{42}
\]
From Eq. (37), $\delta f$ for bandwidth $B$ is given by

$$\delta f = \sqrt{\frac{k_B T f_c B}{\pi k_c Q_c A^2}}.$$  

(43)

This equation shows that $\delta f$ is proportional to $A^{-1}$, whereas $\partial (\Delta f_c)/\partial z$ is known to be constant for small amplitudes and proportional to $A^{-3/2}$ for large amplitudes [129]. From Eq. (42), $\delta z$ is proportional to $A^{-1}$ for small amplitudes and $A^{1/2}$ for large amplitudes. Therefore, there is an optimal amplitude with a minimum value of $\delta z$. This optimal amplitude has been reported to be on the order of the decay length of the force components [129]; on the order of 0.1 nm and 1 nm for short-range and long-range interaction forces, respectively. Therefore, small amplitude operation improves the sensitivity to the short-range interaction force.

8.4. Deflection sensor noise

The noise density of cantilever deflection sensors used for conventional AFMs is typically 100–1000 fm/$\sqrt{\text{Hz}}$ [137] while that for high-resolution FM–AFM imaging in liquid is less than 20 fm/$\sqrt{\text{Hz}}$ [76,123]. In this section, we describe the necessity of such a low-noise deflection sensor and technical details for realizing such an excellent noise performance.

A deflection signal contains two major noise components. One is from the thermal Brownian vibration of the cantilever and the other is from the deflection sensor itself. The RMS value of the deflection noise density due to the thermal noise ($n_{zB}$) is described by Eq. (35), while that from the deflection sensor $n_{zs}$ can be regarded as a white noise at a frequency range relevant to dynamic-mode AFM. For dynamic-mode AFM, noise components in the frequency range from $f_c - B$ to $f_c + B$ influence its performance. Therefore, in order to achieve the theoretically-limited noise performance given by Eqs. (38) and (39), $n_{zs}$ must be reduced to less than $n_{zB}$ in this frequency range.

From Eq. (35), $n_{zB}$, at $f = f_c$ is given by

$$n_{zB}(f_c) = \sqrt{\frac{2k_B T Q_c}{\pi f_c k_c}}. \tag{44}$$

For the typical soft and stiff cantilevers with parameters shown in Table 1, $n_{zB}(f_c)$ is 2.8 pm/$\sqrt{\text{Hz}}$ and 74 fm/$\sqrt{\text{Hz}}$, respectively. Eq. (44) and these values show that a stiffer cantilever has a lower thermal noise peak and hence requires a lower noise level of the deflection sensor to achieve the theoretically-limited noise performance. In order to suppress the influence of $n_{zs}$ to less than 10% of that of $n_{zB}$, $n_{zs}$ must be less than $1/3$ of $n_{zB}$. For the typical stiff cantilever with parameters given in Table 1, $n_{zs}$ must be less than 25 fm/$\sqrt{\text{Hz}}$, which is much lower than a typical value ($n_{zs} = 100–1000$ fm/$\sqrt{\text{Hz}}$) for a conventional deflection sensor.

The noise performance of a deflection sensor is also important for operating FM–AFM with small oscillation amplitude. FM–AFM utilizes a deflection signal for producing a cantilever excitation signal as well as for detecting $\Delta f_c$. In a self-oscillation circuit, noise components that affects the cantilever oscillation spectrum are in the range from $f_c - f_c/ (2Q_c)$ to $f_c + f_c/(2Q_c)$. Thus, in order to make the oscillation spectrum free from the influence from the sensor noise, the above-mentioned criterion becomes $1/\sqrt{2}$ times smaller, namely 18 fm/$\sqrt{\text{Hz}}$.
As discussed above, true atomic-resolution imaging in liquid with dynamic-mode AFM requires the use of a stiff cantilever and a small oscillation amplitude, which in turn requires a low-noise cantilever deflection sensor. Among various methods proposed for detecting cantilever deflection, the OBD method is the most widely used technique due to its simple setup and the ease of optical beam alignment. In the rest of this section, we describe how to improve the value of $n_z$ for an OBD sensor to less than 18 fm/√Hz.

Fig. 46 shows a typical setup of an OBD sensor used for liquid-environment. A focused laser beam is irradiated onto the backside of the cantilever and the reflected beam is detected using a PSPD. The cantilever deflection results in an angular deviation of the reflected laser beam, which in turn gives rise to a displacement of the laser spot on the PSPD. The PSPD produces a differential current signal proportional to the laser spot displacement. The current signal is converted to a voltage signal using an $I-V$ converter and is fed into a differential amplifier, generating a voltage signal proportional to the cantilever deflection.

Major noise sources in the OBD sensors include noise arising from the laser beam and the photodiode (PD) shot noise. Theoretically, the PD shot noise should determine the minimum noise obtained using the OBD sensor. However, the noise performance of practical OBD sensors is often limited by the noise related to the laser beam. Here we presents a way to reduce the laser-related noise to less than that of the PD shot noise, and a method of reducing the shot noise to achieve the theoretically-limited noise performance in FM–AFM.

A laser diode operated with a relatively small output power compared with its rating shows intensity fluctuations originating from spontaneous light emission. This noise is referred to as “quantum noise.” As the laser power increases, the quantum noise becomes negligible compared with the average laser power. Instead, the intensity fluctuation due to the hopping of the laser oscillation mode becomes evident in the high-output-power regime. This noise is referred to as “mode-hop noise.” The intensity fluctuation induced by the quantum noise is eliminated using a differential amplifier as a common-mode noise. In contrast, that from the mode-hop noise is magnified as a differential noise, which thus significantly increases the deflection noise.

The magnitude of the mode-hop noise does not show a monotonic dependence on the output laser power but exhibits a steep increase at a threshold value [76]. Thus, the laser power should not exceed the threshold for the onset of laser mode hopping. Since a high-
output laser power is desirable for reducing the noise from the PD shot noise, the laser power should be set at a value slightly below this threshold.

In addition to the intrinsic noises discussed above, there are two extrinsic noise sources related to the laser beam. The reflection and scattering of the laser beam occasionally cause part of the laser beam to return to the optical resonator of the laser diode. This optical feedback generates additional laser oscillation modes, leading to an increase in mode hopping. This noise is referred to as “optical feedback noise”. On the other hand, some of the reflected or scattered laser beam is incident to the PSPD. This laser beam interferes with other laser beams which travel through different optical paths. Such unexpected optical paths are often unstable and hence the laser spot pattern on the PSPD shows a fluctuation over a relatively long time scale. This noise is referred to as “optical interference noise”. These extrinsic laser noises are particularly evident in liquid-environment AFMs. In the setup for liquid-environment AFM, many interfaces exist in the optical path of the laser beam. Such interfaces enhance the reflection and scattering of the laser beam, leading to an increase in the optical feedback noise and optical interference noise [76].

The extrinsic laser noise can be suppressed by modulating the laser power with an RF signal whose frequency is typically 300–500 MHz [138,139]. The RF laser power modulation changes the laser oscillation mode from single-mode to multimode. Since the mode hopping takes place due to the competition among the possible laser oscillation modes in the optical resonator, the multimode laser oscillation is much less sensitive to the optical feedback than the single-mode laser oscillation. Consequently, RF modulation considerably reduces the mode hopping induced by the optical feedback. In addition, the multimode laser beam has lower coherence than the single-mode laser. Thus, RF modulation is also effective for suppressing the optical interference noise.

Fig. 47 shows a waveform of a $\Delta f$ signal measured in water before and after turning off the RF laser power modulation [76]. The magnitude of the frequency noise suddenly increased after turning off the RF modulation, which shows that the RF modulation is effective in reducing the extrinsic laser noises. Although RF modulation also reduces the intrinsic mode-hop noise, it is often difficult to completely suppress it. Thus, the average laser power should be set at a value below the threshold for the onset of the mode-hop noise even with the RF laser power modulation.

Once the effect of the laser beam noise is sufficiently reduced, the PD shot noise starts to predominantly affect the deflection noise. The RMS deflection noise density arising from the PD shot noise ($n_{zp}$) is given by [123]

![Waveform of the frequency shift signal measured in water before and after turning off the RF laser power modulation. $A = 5$ nm, $f_0 = 140$ kHz, $k = 42$ N/m [76].](image)
\[
\eta_{sp} = \frac{a_0 \ell_c}{6 \eta n_m \ell_f} \sqrt{\frac{2e}{\eta \alpha P_0}}
\]  

(45)

where \( \eta \) is the efficiency of photo-current conversion of the PSPD, \( \alpha \) is the laser power attenuation in the optical path from the laser source to the PSPD, \( e \) is an electron charge, and \( n_m \) is the refractive index of the medium of the operating environment. \( P_0 \) is the laser power measured at the output of the beam source. \( \ell_f, \ell_c, a_0 \) are the focal length, cantilever length, and the diameter of the collimated laser beam, respectively. \( \chi \) is the correction factor for a Gaussian laser beam. If we define \( a_0 \) as the diameter of the beam cross-section at 1/e (or 1/e^2) of the maximum laser intensity, \( \chi \) is 1.13 (or 1.60). The equation gives us a direct guideline for reducing the deflection noise. Namely, increasing \( \ell_f, \chi \) and \( P_0 \), and decreasing \( a_0 \) and \( \ell_c \) will improve the noise performance of the deflection sensor using the OBD method.

Fig. 48 shows a deflection noise density spectrum measured in water using a low-noise OBD sensor developed as discussed above [123]. The experimentally measured spectrum (solid line) shows good agreement with the thermal noise spectrum (dashed line) calculated with Eq. (35) around the resonance frequency. Such a low deflection noise density (\( \eta_{zs} = 16 \text{ fm/} \sqrt{\text{Hz}} \)) ensures the theoretically-limited noise performance in FM–AFM even with a stiff cantilever (\( k_c = 21 \text{ N/m} \)).

9. Imaging of dynamic biomolecular processes

Presently, the world wide number of high-speed AFM users is small. In addition, high-speed AFM has not yet been completely established as a tool for studying dynamic biomolecular processes, although the performance of high-speed AFM has been markedly improved in the last 3–4 years. In exploring the potential of high-speed AFM, the first step must be to demonstrate biological processes that have been predicted or are known to occur. As shown below, such demonstrations (including unpublished results) have been successfully presented very recently at our laboratory and other places. Exceeding the level
of just demonstrations, some of these studies were able to give insights into the functional mechanisms of proteins. In this chapter, we introduce these successful recent studies together with early pioneering studies performed to develop this unprecedented field of high-speed bio-AFM studies. We hope that readers will be encouraged and convinced of the use of this new microscopic tool to find the answers to numerous unsolved biological questions.

9.1. Early imaging studies

In this section, we describe early imaging studies using high-speed AFM performed at our laboratory between 2001 and 2004.

9.1.1. Initial attempts and lessons

In 2001, the AFM imaging of moving protein molecules at an imaging rate of 80 ms/frame was successfully achieved for the first time [54]. This encouraged us to image various biomolecular processes. However, we soon realized that the tip–sample interaction force was too strong to image fragile samples such as actin filaments and microtubules, which were easily destroyed during imaging (Fig. 49). This was due not only to the insufficient feedback speed but also to the large tapping force exerted by the oscillating tip on the sample. Thus, imaging studies in this early stage were mainly performed to evaluate the capacity of this new microscope and to determine which devices had to be improved.

A prototype dynamic PID controller [140] and an active damping technique for the z-scanner [92] were developed during this period. Using the dynamic PID controller, the tapping force was greatly reduced and the feedback bandwidth was enhanced. For example, the unidirectional movement of individual kinesin molecules along microtubules was

Fig. 49. Actin filaments and microtubules being disrupted during imaging. The number attached to each image indicates the time (s) elapsed after imaging began. (Top panel) actin filaments with myosin V in ATP; imaging rate, 80 ms/frame; scan size, 400 nm. (Lower panel) microtubules; imaging rate, 460 ms/frame; scan size, 200 nm.
observed (Fig. 50) without disassembling the microtubules [140]. However, in the observed kinesin movement, the kinesin molecules were attached to both the mica surface and the microtubules. The movement of kinesin that was attached only to microtubules was not observed. On the other hand, in the presence of AMP–PNP, wherein kinesin firmly attaches to microtubules, kinesin molecules could be observed on the top surface of the microtubules (Fig. 51). These observations imply that the tip removed the kinesin that had been attached to only the microtubules in the presence of ATP. While observing

![Image](image1)

Fig. 50. Movement of kinesin–gelsolin along a microtubule. The kimera protein is a homodimer composed of two heavy chains, each of which consists of the head–neck region of kinesin at the N-terminal and gelsolin at the C-terminal. The number attached to each image indicates the time (s) elapsed after imaging began. The arrow heads point at kinesin–gelsolin. Scale bar, 100 nm; imaging rate, 0.64 s/frame [140,141].

![Image](image2)

Fig. 51. Microtubule associated with kinesin–gelsolin in the presence of AMP–PNP. White arrows indicate kinesin–gelsolin bound to the microtubule top surface while red arrows indicate kinesin–gelsolin bound to the microtubule at the side surface. Scale bar, 20 nm.
the movement of actin filaments, we experienced a similar problem. When a mica surface was densely coated with myosin V, the gliding movement of actin filaments was captured on video (Fig. 52) [87, 141]. However, when the myosin V density was lowered, actin filaments did not appear in the images, suggesting that they were removed by the scanning and oscillating tip. These trial observations indicated that, using the dynamic PID controller, the tip–sample interaction force could be successfully reduced to a level at which protein inactivation was prevented and self-assembled polymer were not disassembled. However, the force was not reduced sufficiently to prevent the delicate protein–protein association from breaking.

9.1.2. Flash photolysis combined with high-speed AFM

It is important to observe how a protein conformation changes dynamically upon binding to its ligand. Protein molecules always exhibit Brownian motion due to the thermal agitation from surrounding water molecules. Therefore, changes in protein shape caused by this random agitation have to be distinguished from ligand-induced conformational changes. When the ligand association–dissociation kinetics is fast, the ligand-induced changes occur repeatedly during a reasonable imaging period. A structural change that is repeatedly observed can be judged to be one induced by the ligand reaction. On the other hand, when the kinetics is slow, this distinction is difficult. For example, because of the low ATPase activity of myosin V alone (≈0.05/s), it was difficult to judge whether structural changes observed only once or twice were in fact those induced by the binding of ATP (or its hydrolysis) (Fig. 53).

Fig. 52. Actin filament gliding over a surface that is densely coated with myosin V. The number attached to each image indicates the time (s) elapsed after imaging began. Scale bar, 200 nm; imaging rate, 1 s/frame [141].

Fig. 53. Myosin V in ATP. The number attached to each image indicates the time (s) elapsed after imaging began. The head portion of myosin V molecule, which is circled, appears to change its conformation from an extended form to a folded form. ATP concentration is 2 μM. Imaging rate, 0.5 s/frame; scan range, 300 nm [3].
To solve this problem, we combined the UV flash photolysis of caged compounds with high-speed AFM. To generate UV flashes, we first used a mercury lamp equipped with a UV filter and a mechanical shutter. This photolysis system was used together with caged calcium to observe Ca\(^{2+}\)-induced calmodulin (CaM) dissociation from the myosin V neck regions [140]. Myosin V possesses six IQ motifs at each extended neck domain, to which a number of CaM molecules (~5) and light chains (~1) are attached in the absence of Ca\(^{2+}\) [142,143]. In the presence of Ca\(^{2+}\), some of these CaM molecules dissociate from myosin V [144,145]. The sample stage unexpectedly moved in the x-direction during the application of a UV flash for 1 s and soon returned after shutting off the UV light. This movement was caused by thermal expansion of materials exposed to the UV light. All the calmodulin molecules (and also the light chains) dissociated instantaneously from the neck regions, which resulted in very thin necks. In addition, the coiled-coil part, which had connected the neck–neck junction and the PEST (the amino acid motif that is thought to target cytoplasmic proteins for rapid proteolytic degradation) site, appeared to unwind. These actions are in strong contrast to the recently established knowledge on Ca\(^{2+}\)-induced CaM dissociation (i.e., CaM dissociates only from the second IQ motif on each neck [146]). It appears that agitation caused by the UV-induced thermal expansion confused the feedback operation, resulting in an excessive tip–sample interaction force.

We improved the combined AFM system using attenuated high-frequency laser pulses (~50 kHz, 355 nm). Laser pulses were applied while the y-scan was performed towards the starting point after the completion of one frame acquisition. In addition, during this period, the sample stage was withdrawn from the cantilever tip. This improved method allowed us to observe the rotational movement of the myosin V head around the head–neck junction that occurred immediately after UV application to the caged ATP-containing solution [141] (Fig. 54). We also applied this method to observe the height changes in chaperonin GroEL upon binding to ATP and GroES [87,141] (Fig. 55). The ATP-induced movement of myosin V on an actin filament was also imaged, although this rapid movement was too fast to be resolved in several frames.

9.1.3. Dynein C

Single-headed dynein C from chlamydomonous flagella has a ring-shaped head from which a stem and a stalk protrude. The stalk interacts with microtubules at its globular distal end. Previous electron micrographs of dynein C revealed that the angle between the stalk and the stem differed between the nucleotide-free state and the ADP-Vi bound

![Fig. 54. Myosin V head bending upon UV-flash photolysis of caged ATP. The number attached to each image indicates the time (s) elapsed after imaging began. UV flash was applied between the first and second images. Scale bar, 30 nm; imaging rate, 80 ms/frame [141].](image-url)
state [147]. We confirmed this result using AFM in aqueous solutions. Since dynein C has a relatively high ATPase rate even in the absence of microtubules, it is easy to distinguish the ATPase-driven conformational change from Brownian motion. In fact, as shown in Fig. 56, the high-speed AFM imaging of dynein C alone in the presence of ATP revealed that the stem moved back and forth between two positions while the stalk and head were stationary [87,141]. The two positions approximately corresponded to the nucleotide-free and ADP-Vi bound states. Furthermore, the frequency of the stem movement was similar to the ATPase rate, indicating that the observed periodic stem movement is driven by the ATPase reaction.

9.2. Recent imaging studies

In this section, we introduce successful imaging studies performed on several dynamic biomolecular processes. As the original manuscripts for most of these studies are now in preparation, some of the imaging data are not displayed here.

9.2.1. Myosin V

From early imaging studies, we learned that reduction in the tapping force was essential for the establishment of high-speed AFM for biological studies. To increase the effectiveness of dynamic PID control, we developed a low-noise sensor amplifier. Owing to the

![Fig. 56. Periodic structural change in dynein C. The left panel is a schematic drawing of dynein C. The stem stays mainly at the position shown in the frames 51 and 64 (corresponding to the nucleotide-free state), and occasionally moves to the position shown in the frame 63 (corresponding to the ADP-Vi-bound state). Scale bar, 30 nm; imaging rate, 160 ms/frame [87]. The dynein C sample was gifted from K. Ooiwa and H. Sakakibara (Kansai Advanced Research Center).](image-url)
low-noise, we could use a set point $r \sim 0.9$, which made it possible to observe actin filaments gliding on a mica surface sparsely coated with myosin V (Fig. 57) [87,141]. The bound heads of myosin V were oriented at an angle relative to the actin filament, similar to the well-known arrowhead structure in muscles. From this structure, the sliding actin filament polarity was identified, and thus, it was shown that this filament moved toward the minus end (i.e., the natural direction). However, it was not evident whether large conformational changes in the bound heads occurred while the actin filament was moving unidirectionally.

Despite this successful observation, we were unable to observe the processive movement by single myosin V molecules along actin filaments. In low-ionic-strength solutions, myosin V tended to attach to the mica surface. In high-ionic-strength solutions, it mostly remained free from the surface but the affinity of myosin V for actin was lowered. To circumvent this problem, we reduced the cantilever free-oscillation amplitude from the usual $\sim 5$ nm to $\sim 1$ nm, thereby reducing the tip–sample interaction force but sacrificing the feedback bandwidth. With such a small amplitude and in a high-ionic-strength solution containing a low concentration of ATP, the processive movement was finally captured on video at 0.1 s/frame. The two heads of a myosin V molecule alternated between leading and trailing positions with a walking stride of $\sim 72$ nm, i.e., in a hand-over-hand manner,

![Fig. 57. Three-dimensional images of actin filament sliding movement. The number attached to each image indicates the time (s) elapsed after initiating began. Scale bar 30 nm, imaging rate, 180 ms/frame [141].](image)
as also demonstrated by single-molecule fluorescence microscopy [148–151]. High-speed AFM revealed the movement in greater detail than fluorescence microscopy. The lead lever-arm bent immediately before the rear head detached from the actin filament. This bent form was similar to that shown previously by electron microscopy [152]. The detached rear head rotated around the junction between the two lever-arms and was then attached to a forward actin. Immediately after the attachment, the new lead head was hopping and sometimes moved forward or backward by \( \sim 5 \text{ nm} \) along the actin filament. In the absence of nucleotide or in the presence of ADP, only one head of a myosin V molecule was bound to an actin filament. From the orientation of the bound head, it was clear that the bound head was in the trailing position. On the other hand, in the presence of a medium concentration of AMP–PNP, both heads were associated with an actin filament for a long time. This indicates that upon binding to AMP–PNP, the lead head conformation adapts itself so that it is able to bind to actin by rotating around the head–neck junction. Therefore, when myosin V is bound to an actin filament at both heads in an ATP-containing solution, the trailing head must contain ADP (or no nucleotide in a low ATP concentration) and the lead head must contain ATP or ADP-Pi. After the phosphate is released, the lead head rotates back to the previous orientation, which is unfavorable for actin binding. However, the lead head cannot dissociate from the actin because the ADP-bound state has a high affinity for actin. This energetically unfavorable conformation causes the lead neck to bend forwards, thereby pulling the trailing head, so that it detaches from the actin and is then conveyed forward. By these mechanical processes along the chemical reaction pathway, the two heads alternate their positions to walk along an actin filament.

9.2.2. Chaperonin GroEL

Here, we show an example of high-speed AFM imaging by which a long-lasting controversial question regarding a biomolecular reaction was solved. Chaperonin GroEL consists of 14 identical ATPase subunits that form two heptameric rings stacked back to back [153,154]. A series of biochemical studies [155,156] have shown that there is positive cooperativity in the binding and hydrolysis of ATP in the same ring, whereas there is negative cooperativity between the two rings. Owing to this negative cooperativity, it has been presumed that GroEL binds to GroES at one ring while releasing GroES from the other [157–159]. This alternate on–off switching appears to exclude the concomitant binding of GroES to both rings of GroEL. However, this issue had remained controversial [160–164], since some electron micrographs have shown a complex of GroES–GroEL–GroES with a football shape. To place GroEL on a substratum in a side-on orientation so that both rings are accessible to GroES, we prepared GroEL biotinylated at the equatorial domains [165]. GroEL was attached to streptavidin 2D crystal sheets prepared on planar lipid bilayers containing a biotin lipid. We could capture the alternate association and dissociation events at the two GroEL rings. Surprisingly, before the alternate switching took place, a football structure often appeared. Thus, the high-speed AFM observation clearly solved the controversial issue.

9.2.3. Lattice defects in 2D protein crystals

As described in the next chapter, we have been preparing substratum surfaces by forming streptavidin 2D crystals on supported planar lipid bilayers [166]. The 2D crystal surfaces are useful for the selective attachment of biotinylated molecules. When we were
observing the crystal surfaces with orthorhombic $C22_2$ symmetry, we noticed that lattice defects were sometimes formed at a few places and moved over the surfaces (Fig. 58). This movement is caused by exchanges between the defect (empty) site and one of the surrounding filled sites, as no free streptavidin exists in the bulk solution. To study the vacancy defect mobility and its participation in the crystal growth, monovacancy defects in the streptavidin 2D crystals were systematically produced by increasing the tapping force onto the sample from the oscillating tip, similarly to a method used previously [167]. Unexpectedly, this movement had a preference for one direction relative to the other [168]. The movement projected onto each lattice axis showed a time–displacement relationship characteristic of Brownian motion but the diffusion coefficients differed depending on the axes. Analysis of this dependence revealed that biotin binding affects the contact affinity between two adjacent streptavidin molecules. Streptavidin is a homo-tetramer with subunits organized in dihedral $D_2$ symmetry. Therefore, in the 2D crystals, biotin only binds to the two subunits facing the planer lipid bilayers (Fig. 59a). There are two types of subunit–subunit interactions between adjacent streptavidin molecules: interactions between biotin-bound subunits and interactions between biotin-unbound subunits (Fig. 59b).

One crystallographic axis ($a$-axis) is comprised of contiguous biotin-bound subunit pairs

---

**Fig. 58.** Movement of point defects in 2D streptavidin crystal formed on supported lipid planar bilayer containing biotinylated lipid. The point defects indicated by arrow heads are maintained as point defects during the observation. Scan range, 150 nm; imaging rate, 0.5 s/frame [168].

**Fig. 59.** (a) Schematic of a streptavidin molecule on biotinylated lipid bilayer. Two biotin binding sites occupied by biotin are indicated by the closed circles. The open circles indicate biotin binding sites facing the aqueous solution and are biotin-free. (b) Schematic of streptavidin arrays in $C22_2$ crystal. Unit lattice vectors are indicated: the $a$-axis includes rows of contiguous biotin-bound subunits, while the $b$-axis includes rows of contiguous biotin-unbound subunits [168].
whereas the other axis (b-axis) is comprised of contiguous biotin-unbound subunit pairs. A previous study on the formation of streptavidin 2D crystal on supported planar lipid bilayers suggested that the interaction between biotin-unbound subunits is stronger than that between biotin-bound subunits [169], while another study suggested the reverse relationship [170]. Our analysis of the lattice defect movement clearly revealed that the interaction between biotin-bound subunits is weaker. In addition, from the diffusion constant ratio ($D_b / D_a \sim 2.4$), the difference in the association free energy of the two types of subunit–subunit contacts were quantified to be $-0.88 \, k_B T$ ($T \sim 300$ K), which corresponds to $-0.52 \, \text{kcal/mol}$ [168].

Fusion of two point defects into a larger point defect was often observed (Fig. 60). On the other hand, fission of a multivacancy point defect into smaller point defects was rarely observed [168]. We think that the fission often occurs but cannot easily be observed because the two point defects formed immediately after the fission are quickly fused again. Immediately after the fission, a streptavidin molecule becomes facing both the resulting two smaller point defects. This streptavidin molecule contains only two intermolecular subunit–subunit contacts, and therefore, quickly detaches from the lattice and moves into either of the two adjacent point defects. The imaging rate we used (0.5 s/frame) was not sufficiently high to observe this transiently occurred fission.

During the formation of streptavidin 2D crystals in the presence of free streptavidin in the bulk solution, small point defects such as mono- and di-vacancy defects would not easily access to the free streptavidin molecules (Fig. 61a), and hence, have a tendency to remain in the crystals. However, the fusion of small point defects into a larger point defect facilitates its access to the free streptavidin molecules (Fig. 61b) and thereby promotes the removal of small point defects from the crystals. Interestingly, the defect mobility increases with increasing defect size [168]. The higher mobility of larger point defects increases their probability to encounter other point defects to form yet larger point defects with further higher mobilities. During the crystal growth in the presence of free streptavidin in the bulk solution, this acceleration effect also promotes the removal of point defects from the crystalline regions.

9.2.4. Intrinsically disordered regions of proteins

In addition to biological processes, high-speed AFM imaging is also useful for identifying structurally flexible regions of proteins, particularly intrinsically disordered (ID) regions of proteins.
Biomolecular binding specificity has been described for a quite while by the famous “lock and key” mechanism, posing that a protein must be folded to give a high degree of geometrical precision in molecular binding. This concept is recently being overturned by a surprising yet credible finding that many proteins in the cell appear to be unfolded most of the time. Unstructured segments mostly comprise flexible linkers that play important roles in the assembly of macromolecular complexes and in the recognition of smaller biomolecules. The functional importance of ID regions has recently been recognized, particularly in transcription, translation, and cellular signal transduction [171,172]. However, there are no useful techniques for analyzing unstructured segments of this sort at the single-molecule level. Indeed, X-ray crystallography and electron microscopy do not allow us to directly observe the ID regions. Conventional slow AFM also cannot reveal such a thin and flexible entity. Under ambient conditions, thin and flexible unstructured polypeptides tend to form lumps, or become flattened due to their strong attachment to a surface. Such flattened polypeptides are difficult to visualize by AFM. In addition, they are not easily adsorbed onto substrate surfaces in solution, even with the assistance of adhesive chemicals. Recently, we observed the facilitates chromatin transcription (FACT) protein in solution by high-speed AFM [173]. FACT is a heterodimer and displaces histone H2A/H2B dimers from nucleosomes, thereby facilitating RNA polymerase II transcription [174,175] and chromatin remodeling [176]. The image of oligomerized FACT clearly revealed undulating tail-like structures protruding from the main body of FACT (Fig. 62) [173]. The image of a FACT monomer indicated that FACT contains two tail-like structures of different lengths [173]. This finding coincided with the ID regions predicted from the amino acid sequences. In fact, the AFM images of deletion mutants lacking either of the two predicted ID regions showed either of the shorter and longer tail-like structures. Their macroscopic contour lengths of the shorter and longer ID regions showed wide distributions and were on average 17.8 nm and 26.2 nm, respectively.

We analyzed mechanical properties (stiffness) of the observed tail-like structures. The stiffness of a polymer chain is described by the persistence length $p$ of the chain, i.e., the length over which the average correlation in the tangent direction decays by $1/e$. In two-dimensions, the mean square point-to-point distance of the chain is given by [177]...
where \( l \) is the contour length between two points on the chain. Here, note that the persistence length depends on the hierarchy of the concerned polymer structures. For example, the persistence length of a double helical DNA strand is different from those of thicker strings formed by super-coiling of a DNA strand. In the AFM images obtained for the tail-like structures of FACT, polypeptide chains are not resolved. Therefore, the persistence lengths that can be obtained from the images are those for overall (macroscopic) string structures (see inset in Fig. 63). We analyzed the tail-like structures of the deletion mutants. The mean square point-to-point distance \( \langle r^2(l) \rangle_{2D} \) as a function of \( l \) was well fitted by Eq. (46) for both cases, giving persistence lengths of 10.5 nm and 12.1 nm for the shorter and longer tail-like structures, respectively (Fig. 63). These similar values indicate that the two ID regions have physically similar structures.

The Young modulus \( E \) is a material property that represents its stiffness and is related to the macroscopic persistence length as \( E = \frac{p k_B T}{I} \) [178], where \( I \) is the momentum of inertia of the cross-section of the object. Assuming that the tail-like structures of FACT are thin rods with a radius \( \rho \), \( I \) is given by \( I = \pi \rho^4 / 4 \). We estimated \( \rho \) (1–1.6 nm) from the width of the tail-like structures on several AFM images obtained using different cantilever tips, taking into account the tip radiuses. The width is less affected than the height by the substrate surface and its estimation is reliable when the sample height is similar to or smaller than the tip radius. Thus, the Young modulus of the two tail-like structures was estimated to be 9–58 MPa from their similar macroscopic persistence lengths. This value

\[
\langle r^2(l) \rangle_{2D} = 4pl \left[ 1 - \frac{2p}{l} \left( 1 - e^{-l/2p} \right) \right],
\]
estimated as a macroscopic stiffness of the ID regions is 1–2 orders of magnitude smaller than those reported for globular proteins such as an actin monomer (~2 GPa) [179] and lysozyme (0.5–2 GPa) [180,181]. The value for an actin monomer is estimated by the force–extension relationship measured using single actin filaments. The value of 0.5 GPa for lysozyme is estimated by AFM force-indentation experiments using a tip with a radius of ~20 nm [180] while the value of 2 GPa for lysozyme is estimated by measuring the speed of ultrasound propagation through a lysozyme solution [181]. Therefore, these values represent the macroscopic stiffness of these proteins. Interestingly, Young’s modulus esti-
mated as macroscopic stiffness of bovine carbonate anhydrase II is reported to be in a range of 2–6 MPa when the sample is incubated in the presence of 2–6 M guanidine HCl [182]. Although these values of Young’s modulus are somewhat smaller than the value for the ID regions of FACT, the structures of the ID regions are considered to be roughly similar to those of denatured proteins.

The unique high-speed AFM capability to detect unstructured regions of proteins is also useful for successful protein crystallization, since the predicted ID regions can be confirmed and deleted before crystallization to facilitate crystal growth.

9.3. Imaging using commercially available high-speed AFM

The full scale commercialization of high-speed AFM has not been achieved yet. To the best our knowledge, two manufacturers have produced prototype instruments. Infinitesima (Oxford, UK) has recently produced VideoAFM™ equipped with a fast x-scanner made of a quartz tuning fork, based on the studies by Miles’ group [183,184]. A fast feedback system is not implemented in this prototype. Although its marketing is restricted to only Japan at present, Olympus (Tokyo, Japan) has produced Nano Live Vision™ which is a modified version of the microscope previously developed by us [54]. One of the users of the latter product has recently filmed the dynamic processes of GroEL–GroES interaction, which was regulated by the ATPase reaction of GroEL immobilized onto a mica surface in an end-up orientation [185]. Moreover, the group observed the 1D diffusion of a restriction enzyme along a DNA strand, which was followed by the cleavage reaction [186]. A more recent study successfully captured the translocation and looping of DNA by a type-III restriction enzyme, EcoP15I [187]. Another group observed the formation and dissociation of a streptavidin–biotinylated DNA complex [188].

10. Substrata for observing dynamic biomolecular processes

To visualize dynamic biomolecular processes, the substratum surfaces on which a sample is placed should meet various requirements; for example, (a) the surface must have an appropriate binding affinity for the sample so that the sample does not move too fast on it, (b) the functional activity of the sample should be retained on the surface, (c) the surface should selectively attach a specific component in a multicomponent sample, and (d) the sample should attach to the surface in a desired orientation. Mica (natural muscovite or synthetic fluorophlogopite) has frequently been used as a source substratum owing to its surface flatness at the atomic level over a large area. It has net negative charge and is therefore quite hydrophilic. A bare mica surface adsorbs various proteins by electrostatic interactions. Except in some cases (such as GroEL attachment in an end-up orientation), the orientation of adsorbed proteins is not unique, and the selective attachment of a specific species is not expected. When the dynamic processes of a single species of protein are to be observed without other proteins, a mica surface is useful. We can control the affinity for a specific protein by varying the ionic strength or pH, or by adding divalent cations such as Mg$^{2+}$. However, interesting biomolecular processes often occur in multi-component systems.

It is desired to have, in the first step, a surface that does not initially attach to the sample. Then, this surface is modified so that only a specific species of molecules can attach to it through a linker in a desired orientation. A membrane surface with zwitterionic polar
head groups such as phosphatidyl choline (PC) and phosphatidyl ethanolamine (PE) is known to resist protein adsorption \[189,190\]. Streptavidin is also useful for the specific attachment of a biotinylated protein since it also resists nonspecific protein adsorption. Mica-surface-supported planar lipid bilayers \[191\] can be easily formed from liposomes as their structure is disrupted by the strong hydrophilic interaction with the mica surface \[192\]. Various lipids with functional groups that are attached to polar groups (e.g., biotin attached to PE, Ni-NTA attached to phosphatidylserine (PS)) are commercially available. They enable the specific attachment of proteins labeled with biotin or his-tags onto planar lipid bilayers. Dioleoyl-phosphatidyl-choline (DOPC) is useful for the preparation of 2D streptavidin crystals when it is used together with biotinylated lipids \[193,194\]. It contains an unsaturated hydrocarbon in each of the two alkyl chains, which causes the bending of the chains and therefore weakens the interaction between neighboring DOPCs. This weak interaction lowers the phase-transition temperature of DOPC lipid to \(-20^\circ\text{C}\), thereby affording considerable fluidity to the planar bilayer at room temperature and thus facili-

Fig. 64. Streptavidin on mica-supported planar lipid bilayers. (a) 2D streptavidin crystals formed on bilayers of DOPC + DOPE-biotin. Scale bar, 30 nm. (b) Streptavidin on bilayers of DPPC + DPPE-biotin. Scan range, 200 nm; imaging rate, 0.18 s/frame. (c) Streptavidin on bilayers of DPPC + DOPE-biotin. Scan range, 200 nm; imaging rate, 0.18 s/frame. The number attached to each image in (b) and (c) indicates the time (s) elapsed after capturing the first frame. The apparent difference between (b) and (c) in the size of streptavidin molecules are due to the difference in the sharpness of the cantilever tips used \[3\].
tating the 2D crystal formation of streptavidin (Fig. 64a). The densely packed streptavidin does not diffuse easily. If less diffusibility is necessary, the packed streptavidin can be cross-linked using glutaraldehyde, which does not affect its ability to bind to biotin. We used this surface for the selective attachment of biotinylated GroEL in a side-on orientation. Dipalmitoyl-phosphatidyl-choline (DPPC) contains no unsaturated hydrocarbons in the alkyl chains; therefore, its phase transition temperature is high (≈41 °C) and it is suitable for preparing planar lipid bilayers with low fluidity. For example, when planar bilayers are formed using DPPC at a high temperature (≈60 °C) together with a suitable fraction of DPPE-biotin, streptavidin, which is sparsely attached to the surface, hardly diffuses at room temperature (Fig. 64b). When DOPE-biotin is used together with DPPC, the sparsely attached streptavidin diffuses at a moderate rate (Fig. 64c) [3].

Some degree of protein adsorption onto a planar lipid bilayer surface is sometimes necessary to observe dynamic interactions between an anchored protein and a floating counterpart protein. When the counterpart protein contains a flexible structure, it exhibits very rapid Brownian motion even when the protein is in contact with an anchored protein. As such a rapidly moving structure cannot be imaged even by high-speed AFM, it is necessary to find appropriate lipid polar groups to slow the motion. Although the relationships between the properties of protein adsorption onto planar bilayer surfaces and the lipid composition have been studied [195–197], we need more extensive studies on this issue.

11. High-speed phase-contrast imaging

11.1. Phase mapping

11.1.1. Compositional mapping on blended polymers

To demonstrate that high-speed phase-contrast imaging is possible using small cantilevers and the fast phase detector described in Chapter 7, we first imaged poly(styrene–butadiene–styrene) (SBS) block copolymers as a test sample, as SBS films are often used for phase imaging with tapping-mode AFM [114,198]. The SBS polymer was dissolved in toluene and then cast on a freshly cleaved mica surface, which was followed by drying under ambient conditions. Prior to the observation, the sample was heated at 70 °C for 1 h. The small cantilever used had a resonant frequency of 658 kHz and a quality factor of ≈2 in water, and a spring constant of ≈0.2 N/m.

It is known that several types of phase separation patterns are formed using SBS films depending on the evaporation speed of the solvent [199]. Fig. 65 shows typical (a) topographic and (b) phase images obtained simultaneously at a rate of 84 ms/frame in distilled water. Here, the phase detection timing was set at a regime in which the tip was approached the surface. Clear phase-contrast images were obtained even at such a high imaging rate. In the phase-contrast images shown in Fig. 65b and d, darker regions correspond to an advanced phase with an average shift of approximately +6° relative to the phase at brighter regions. Since the image contrast in the error signal is faint, the phase-contrast predominantly reflects the compositional heterogeneity of the SBS film. To identify the compositions of different region in the topographic image, the peak-to-peak set point amplitude was reduced from 11 nm to 7 nm while imaging. Fig. 65c and d shows topographic and phase-contrast images obtained with the reduced set point amplitude, respectively. The higher topographic region shown in Fig. 65a was dented by the increased loading force, as shown in Fig. 65c. The cross-section height profiles shown
in Fig. 65e reveal that it was dented by more than 5 nm. Thus, the compositions of the higher and lower regions in Fig. 65a are identified to be the less stiff poly-butadiene (PB) and the stiff poly-styrene (PS) domains, respectively. The darker (advanced phase) area in the phase-contrast image shown in Fig. 65d is wider than that in Fig. 65b. This is because the contact area between the tip and the surface was increased by the increased loading force. However, the magnitude of the phase shift is independent of the loading force, as can be seen in the cross-section phase profiles shown in Fig. 65f. Therefore, the observed phase-contrast is caused by only compositional heterogeneities and is independent of the surface topography.

Fig. 66 shows (a) topographic and (b) phase-contrast images obtained at 896 ms/frame on a different SBS film. The configuration of this sample is very different from that shown in Fig. 65. This large difference was caused by a difference in the solvent evaporation condition. In this case, the topographic and the phase-contrast images are not consistent with each other. The compositional heterogeneity revealed in the phase-contrast image is not clearly seen in the topographic image.

Thus, it was clearly demonstrated here that the combination of small cantilevers and the fast phase detector enables highly sensitive and rapid phase detection of the cantilever
oscillation. This capability allows us to simultaneously observe the surface topography and the surface material heterogeneity map at speed greater than 10 frames/s.

11.1.2. Dependence of phase-contrast on detection timing

To determine conditions for the optimum phase-contrast as well as to obtain an insight into the phase-shift mechanism, we investigated the dependence of the phase-contrast on the phase detection timing within a cantilever oscillation period [119]. Fig. 67a shows an oscillation signal of a cantilever tip that is tapping an SBS film in water. The cantilever oscillation “0 nm” corresponds to the situation that the tip is touching the SBS surface. The oscillation signal is slightly distorted from the sinusoidal wave due to the induction of nonharmonic components attributed to the cantilever’s low quality factor. Fig. 67b–d shows typical phase images obtained at various detection timings. At the phase detection trigger points in the region indicated by (i) in Fig. 67a, we observe only a faint phase-contrast as shown in Fig. 67b. Since the image contrast of the error signal is similar to this phase-contrast, this faint phase-contrast is mainly due to the fluctuations in amplitude. On the other hand, the phase-contrast image obtained at the trigger points in region (ii) shows a relative phase advance of approximately +5° in the PB regions as shown in Fig. 67c. Then, at the trigger points in region (iii), we can clearly observe the phase-contrast with a relative phase advance of approximately +10° in the PB region as shown in Fig. 67d. The phase-contrast image obtained at the trigger points in region (iv) shows a similar contrast to that shown in Fig. 67c. Thus, the magnitude of the phase shift strongly depends on the timing of the trigger point; namely, it depends on the movement direction of the oscillating tip. The largest phase shift is observed when the tip approaches the surface. However, when the tip withdraws from the surface, a phase shift barely occurs. At the positions where the tip velocity is almost zero, the phase shifts only slightly. We used

Fig. 67. Phase-contrast images of a PS–PB block co-polymer film depending on the phase detection timing. (a) Cantilever oscillation signal and various trigger regions for phase detection. The amplitude set point \( r \) is 0.63 and the peak-to-to-peak free-oscillation amplitude is 8.7 nm. The oscillation signal of “0 nm” corresponds to the distance where the tip is in contact with the surface. The phase-contrast images shown in (b), (c), and (d) are obtained at trigger points within regions (i), (ii), and (iii), respectively. The imaging rate is 1.2 s/frame, and the scan area is \( 350 \times 350 \) nm\(^2\) with \( 200 \times 200 \) pixels. (e) Phase image obtained in air. Broken circles indicate the poly-butadiene regions. The imaging rate and scan area are 8 s/frame and \( 400 \times 400 \) nm\(^2\) with \( 256 \times 256 \) pixels, respectively [119].
three different cantilevers with resonant frequencies of about 600 kHz, 800 kHz, and 1.2 MHz. These showed a similar trigger-timing dependence. In addition, the same tendency was observed regardless of the imaging rate, free oscillating amplitude, and set point amplitude. Surprisingly, the phase-contrast images observed under the ambient conditions revealed a different contrast (Fig. 67e). The phase was slightly delayed in the PB regions encircled by dotted lines; identification of these PB regions was performed using the same indentation method as that employed to obtain the image shown in Fig. 65c. Thus, the mechanism of the phase shift significantly differs between air and liquid-environments. In addition, the distortion of the sinusoidal waveform in air is smaller than that in water, and the strong dependence of the phase-contrast on the trigger-timing is not observed. This is because the quality factor is relatively high in air.

Fig. 68a and b shows the phase lag and oscillation amplitude, respectively, recorded in water with different trigger-timings as functions of the relative tip–surface distance. In the PS region, the oscillation amplitude decreased more significantly than in the PB region, as the PB region is much softer than the PS region. The phase shows a delay for both the PS and PB regions. The phase delay in the PS region measured at a trigger point in region (iii) (see Fig. 67a) is larger than that in the PB region, as shown in Fig. 68a. This large difference results in a clear contrast in the phase image, as exemplified by Fig. 67d. On the other hand, the phase delay measured at a trigger point in region (i) does not differ significantly between the PS and PB regions, as shown in Fig. 68b. The vertical lines in Fig. 68a and b

![Fig. 68. Phase lags and cantilever oscillation amplitudes as a function of the relative tip–surface distance measured on the trigger points in (a) region (iii) and in (b) region (i) shown in Fig. 65a. Solid and dotted curves are obtained on PS and PB rich domains, respectively. (c) The phase lags measured on the PS (●) and PB (▲) regions are plotted as a function of the trigger points. The phase values are measured at the tip–surface distance indicated by broken lines in (a) and (b) [119].](image-url)
indicate the tip–sample distances where similar amplitude reductions occur in trigger regions (i) and (iii). As shown in Fig. 68c, at these tip–sample distances (which are similar), the phase delays in the PS and PB regions change with time after the tip–sample contact. The phase delay in the PS regions increases with time, while that in the PB region remained almost constant. An interpretation of this different behavior is given later.

Phase-contrast images cannot be interpreted straightforwardly because the phase is affected by a number of instrumental and non-instrumental parameters [6,7,114–117,199–202]. These studies on phase-contrast mechanisms have been performed only under ambient or vacuum conditions. Instrumental parameters such as the free-oscillation amplitude and the set point amplitude affect the phase image contrast in a complex manner. It has been reported that a flipped phase-contrast occurs when the tapping force level is being changed [114,115,203]. Such contrast flips generally occur because of a change in the tip–sample interaction between the net attractive- and repulsive-force regimes [204,205]. However, we did not observe these phase-contrast flips. In addition, the amplitude–distance and phase–distance curves (Fig. 68a and b) did not show the features of a transition from the “low-amplitude state” to the “high-amplitude state”, the transition that takes place upon moving from the attractive regime to the repulsive regime [204]. Therefore, the tip–sample interaction is mostly repulsive (e.g., high-oscillation amplitude state) in both the PS and PB regions under the present imaging conditions. Under the repulsive-force conditions, the phase-contrast of a PS/PB blend film observed under ambient conditions mainly reflects its mechanical properties [201,205]. In this case, the phase signal is more delayed in the softer PB region than in the harder PS region. This can be explained by the fact that the larger repulsive-force on the harder material induces a more positive resonant frequency shift, but the damping constants on both materials are similar [201,205]. However, in water, the phase-contrast is reversed. Hence, the phase image contrast in water appears to be dominated by other properties of PS and PB than their mechanical properties and is very likely to be affected by the adhesion force between the tip and the sample [202,206]. Both PB and PS are nonpolar materials but PB is relatively more polar than PS because the thermal treatment of PB results in the formation of some C=O groups or other polar groups such as OH [201]. Because the AFM tip used here is made of hydrophobic amorphous carbon, the adhesion force derived from the hydrophobic interaction should be larger in the PS region than in the PB region. This difference produces a larger phase delay in the PS region. In air, large capillary forces mask such chemical interactions [202,206], but the mechanical properties are emphasized by the larger quality factor in air.

Here, we propose a possible interpretation of the trigger-timing dependence of the PB/PS phase-contrast observed in water. A larger repulsive-force generated in the harder PS region produces a greater phase advance after the tip comes in contact with the surface, because a larger positive frequency shift relative to the fixed driving frequency occurs. The water medium damps the cantilever oscillation, and hence, the phase advance decreases within one oscillating cycle due to the low Q factor of the cantilever [207]. Therefore, the phase advance produced immediately after the tip–surface contact becomes smaller with time. On the other hand, the larger phase delay in the PS region caused by the adhesive force does not decay quickly within a single oscillation cycle, because of the energy-dissipative interaction. As the adhesive interaction effect is larger than the repulsive-force effect, the net phase shift is a delay rather than an advance. Consequently, when the tip withdraws from the surface, the net phase delay increases with time, as shown in
Fig. 68c. For the softer and less adhesive PB regions, the phase advance due to the repulsive-force is smaller and the phase delay due to the adhesive interaction is also smaller. However, the net phase delay in trigger region (i) is similar to the corresponding phase delay observed in the PS regions. Since the contribution of the phase advance to the net phase delay is small in the PB regions, the gross phase is almost maintained, as shown in Fig. 68c.

The insights given above lead to the following important conclusions. In phase-contrast imaging in liquids, the optimization of the trigger point is essential for obtaining a good signal-to-noise ratio. Moreover, by selecting the trigger point or by simultaneously obtaining phase-contrast images at multiple trigger points, it is possible to separately image the mechanical and chemical properties of the sample.

11.1.3. Phase imaging of myosin filament

To demonstrate the performance of the simultaneous topographic and phase-contrast imaging of biological systems, here we present images obtained using myosin filaments in a buffer solution. Skeletal muscle myosin filaments were obtained by diluting the monomer solution with a low-salt buffer [208]. The cantilever peak-to-peak free-oscillation amplitude was about 12 nm and the amplitude set point was about 0.7. Fig. 69a and b shows typical topographic and phase-contrast images, respectively, obtained at an imaging rate of 4 s/frame. The scan area was 2 μm × 2 μm, which was relatively large for high-speed imaging and therefore the imaging was performed at relatively slow speed. The surfaces of the myosin filaments appear smooth in the topographic image. On the other hand, the phase-contrast image shows the structures of the filaments in more detail. Here, the trigger-timing was tuned so that the maximum phase-contrast was obtained. Note that the phase on the myosin filaments is delayed relative to that on the mica surface. This probably indicates differences in the mechanical properties between the biological sample and the mica; the myosin filaments have higher elasticity and/or viscoelasticity than mica. The phase-contrast on the myosin filaments indicates the non-uniformity of the mechanical properties.

11.2. Phase-modulation imaging

11.2.1. Distance dependence of phase-shift signal

Recently, it has been demonstrated that FM-mode AFM has atomic-resolution imaging capability even in liquids (the details are described in Chapter 12) [121–123,132]. In

Fig. 69. (a) Topographic and (b) phase-contrast images of myosin filaments captured at 4 s/frame in buffer solution. The scan area is 2 × 2 μm² with 256 × 256 pixels.
FM mode, a cantilever is excited at its resonant frequency in a self-oscillation mechanism, which increases its effective $Q$-value. The increased $Q$-value slows the cantilever amplitude response. In addition, the bandwidth of the conventional phase-locked loop (PLL) that is used to detect the frequency shift is usually limited to a few tens of kHz. Thus, it appears difficult to increase the imaging rate of FM–AFM. On the other hand, as described in Section 11.1, the phase of the cantilever oscillation excited at a fixed frequency at near its resonant frequency is also sensitive to the tip–sample distance as well as to the surface material properties of the sample [209]. When the material properties are not significantly inhomogeneous on the sample surface, the phase signal will vary with dependence almost solely on the tip–surface distance. As mentioned in Section 11.1, a small $Q$-value can be used for the phase detection. Thus, phase-modulation mode AFM (PM–AFM) appears promising for high-speed low-force imaging. In the simple model of harmonic oscillation, the phase shift $\Delta \phi$ induced by the elastic interaction is approximately expressed by $\Delta \phi \approx -Qk/k_c$, where $k$ is the gradient of the interaction force. Our small cantilever used in high-speed AFM has a $Q$-value of $\sim 2.5$ in water but the spring constant $k_c$ is much smaller than that used in FM–AFM. Therefore, phase shifts that are detectable using our fast phase detector are produced by the tip–sample interaction.

Fig. 70 shows the amplitude–distance and phase-shift–distance curves observed on a mica surface in a buffer solution. The phase is advanced (i.e., the resonant frequency is increased) due to the repulsive interaction and its sensitivity is 28 mV/$\text{n}m$. The amplitude signal starts to decrease slowly at the distance marked with ‘A’ in Fig. 70. From distance ‘B’, it decreases more quickly. For imaging in the AM mode, the set point amplitude should be set less than that at the distance B to obtain sufficient image quality. The slope of the amplitude–distance curve at distance B is approximately 60 mV/$\text{n}m$. Meanwhile, the phase-shift signal decreases almost linearly from the distance A with an average slope of 280 mV/$\text{n}m$. Because of this higher sensitivity of the phase signal, the PM mode (the phase signal is used for the feedback control of the tip–sample distance) enables low-invasive imaging on the sample as well as at higher spatial resolution than the AM mode.

Fig. 70. Amplitude-distance (dotted line) and phase shift-distance (solid line) curves measured on a mica surface in a buffer solution.
11.2.2. PM imaging of GroEL

Chaperonin GroEL was used as a test sample for the PM imaging, as its stacked ring structure is fragile and is often bisected by strong tip forces [210]. Fig. 71a–d shows successive topographic images taken in the AM mode at 800 ms/frame. The set point amplitude was approximately 90% of the peak-to-peak free amplitude of 7.5 nm. In these images, a large fraction of GroEL molecules are bisected and are sometimes dislodged by the scanning tip. On the other hand, in the PM mode, the bisection of GroEL rings barely occurred as shown in Fig. 71e–h. In addition, the hole centered on the ring structure is more clearly observed in the PM mode than in the AM mode, and the subunits are resolved in the PM mode (Fig. 72).

Fig. 71. Successive AFM images of GroEL taken by (a)–(d) the AM mode and (e)–(h) by the PM mode. Scan area is 400 x 400 nm². Imaging rate is 800 ms/frame.

Fig. 72. High-resolution image of GroEL taken in the PM mode. Scale bar, 10 nm; imaging rate, 480 ms/frame.
12. Atomic-resolution imaging in liquids

12.1. Atomic-resolution imaging by FM–AFM in liquid

Since the first demonstration of true atomic-resolution by FM–AFM in 1995 [10], a number of atomic-resolution FM–AFM images have been obtained on various surfaces in UHV. However, the high-resolution imaging capability of FM–AFM had been limited to UHV until recently. In 2005, Fukuma et al. presented a method of overcoming this limitation using an ultralow noise cantilever deflection sensor, stiff cantilevers, and a small oscillation amplitude [76]. In this section, we present the high-resolution FM–AFM images that demonstrate the significant progress in improving the spatial resolution in dynamic-mode AFM for liquid-environment.

The first true subnanometer-resolution imaging by FM–AFM in liquid was demonstrated by imaging a single crystal of polydiacetylene [poly-PTS; 2,4-hexadine-1,6-diol bis(p-toluene sulfonate)] in water, as shown in Fig. 73a and b [121]. Fig. 73a shows stripes with a spacing of approximately 0.75 nm, which correspond to the spacing of the polyacetylene main chain. The dots along each stripe represent individual PTS side chains. The image also shows molecular-scale defects. Fig. 73b reveals two protrusions corresponding to the two adjacent carbon atoms in a PTS side chain. These images demonstrate the true subnanometer-resolution (0.25 nm) of FM–AFM in liquid.
Soon after the imaging of poly-PTS, the first true atomic-resolution imaging by FM–AFM in liquid was reported by imaging a muscovite mica surface in water, as shown in Fig. 73c and d [132]. Fig. 73c shows a honeycomb-like pattern, which is characteristic of a cleaved mica surface. In the more magnified image shown in Fig. 73d, bright spots are resolved along the honeycomb pattern. These bright spots have been tentatively attributed to Al$^{3+}$ ions in the aluminosilicate honeycomb mesh. The direct imaging of individual ions irregularly aligned on the surface demonstrates that the true atomic-resolution can be obtained by FM–AFM in liquid.

12.2. Biological applications of FM–AFM

One of the most attractive applications of liquid-environment FM–AFM is the high-resolution imaging of biological systems under physiological environments. Since the spatial resolution of FM–AFM in liquid is as small as the size of a single water molecule or a solvated ion, FM–AFM should make it possible to visualize the atomic-scale interaction potential at the interface between biological systems and the physiological environment (i.e., water molecules and ions). In this section, we describe examples of such applications reported in 2007 by directly probing the atomic-scale interactions between a model biological membrane and physiological solution [64,122].

12.2.1. Direct imaging of intrinsic hydration layers

Water molecules adjacent to biological membranes could potentially play a key role in important biological processes, particularly, when they form hydration layers. The hydration layers alter the local interaction potential at the water/lipid interface, which significantly affects the function of membrane proteins and molecular transport across the membranes. FM–AFM has a unique capability of measuring local interaction potential with a nanometer-scale cross-section using a sharp tip as a probe. This capability is ideal
for directly probing the interaction that nanoscale objects (e.g., proteins and solvated ions) experience when they approach a membrane surface.

Fig. 74a shows an example of such applications: a $\Delta f$ vs. distance curve measured on a mica supported DPPC lipid bilayer in phosphate buffer saline (PBS) solution [64]. The raw data (gray line) was smoothed (black line) and was converted to a force vs. distance curve (Fig. 74b) using the equation reported by Sader and Jarvis [211]. The force curves obtained on the DPPC bilayer typically show an oscillatory force profile with one or two peaks. The average separation of the two peaks is $0.28 \pm 0.05$ nm. This distance agrees with the size of a water molecule, suggesting that the oscillatory force profile corresponds to the sequential removal of ordered water molecules.

The tip-sample distance is regulated on the basis of the assumption that $\Delta f$ increases with decreasing tip-sample separation such as in the force branches indicated by arrows (i)–(iii) in Fig. 74a. Thus, the oscillatory profile of the $\Delta f$ vs. distance curve implies existence of multiple feedback positions for a single set point value as indicated by the circles in Fig. 74a. Namely, the tip can spontaneously jump between these positions even if the set point is unchanged.

Fig. 75 shows examples of such spontaneous jumps during FM–AFM imaging of the DPPC bilayer in PBS solution [64]. In the image shown in Fig. 75a, the tip is scanned from the lowest terrace (Terrace 1) and spontaneously jumps twice. The height of the spontaneous jumps is 0.2–0.3 nm, which agrees with the size of a single water molecule. The image shows the primary and secondary hydration layers with molecular-scale corrugations corresponding to the headgroup positions as well as the headgroups themselves. Such stable imaging of the hydration layers demonstrates the excellent force sensitivity and spatial resolution of FM–AFM in liquid. It also proves that the hydration layers next to the DPPC membrane are sufficiently stable to allow such imaging.

12.2.2. Direct imaging of lipid-ion networks

Under physiological conditions, biological membranes are surrounded by an electrolytic solution containing various metal cations. So far, spectroscopy experiments have revealed that metal cations specifically interact with negatively charged moieties of the lipid headgroups [212–215]. These experiments, together with theoretical simulations [216], have led to an idea that individual ions may interact with multiple headgroups to form complex “lipid-ion networks”. This idea has been used to account for the observed effects of the ions such as the enhanced mechanical strength of membranes [217] and the reduced mobility of the lipid molecules therein [218]. However, it has been difficult to experimentally access such lipid-ion networks due to the lack of a method able to investigate local lipid-ion interactions with Ångstrom resolution. Since FM–AFM allows us to directly observe the water/lipid interface with Ångstrom resolution, it is an ideal tool for such biological applications.

The DPPC membrane surface imaged by FM–AFM exhibits various structures depending on both time and location, reflecting the mobile nature of the ions interacting with the lipids [122]. FM–AFM images reveal the existence of at least two different configurations as shown in Fig. 76a and b, which are referred to as Structures 1 and 2, respectively. Structure 1 has pairs of two protrusions separated by 0.30 nm, corresponding to the phosphate and choline groups in the PC headgroups. Structure 2 has hexagonally arranged surface groups consisting of two oval-shaped subunits with their longer axes parallel to each other. Detailed analyses of the FM–AFM images and theoretical simulations [122] have
suggested that the Ångstrom-scale contrasts found in Structure 2 represent the averaged position of mobile ions interacting with the lipid headgroups, namely, lipid-ion networks. The two subunits in Structure 2 show enhanced ion occupancies around the two negatively

Fig. 75. Example of an FM–AFM image of the DPPC bilayer in phosphate buffer solution showing spontaneous jumps during imaging. Image size: 8 nm × 8 nm. Tip velocity, 120 nm/s. Imaging speed: 85 s/frame. (b) Line-by-line flattened image of (a). Scale bar, 1 nm. Height range: 0.1 nm (black to white). Fast and slow scan directions: left to right and top to bottom. The regions indicated by the numbers 1, 2, and 3 correspond to the Terraces 1, 2, and 3 in (a). (c) Line-averaged height profile of (b) plotted along the slow scan direction [64].

suggested that the Ångstrom-scale contrasts found in Structure 2 represent the averaged position of mobile ions interacting with the lipid headgroups, namely, lipid-ion networks. The two subunits in Structure 2 show enhanced ion occupancies around the two negatively

Fig. 76. FM–AFM images of the DPPC bilayer in PBS solution. Height scale: 0.12 nm. (a) Structure 1. Tip velocity: 146 nm/s. (b) Structure 2. Tip velocity: 120 nm/s [122].
charged oxygen atoms in the phosphate group, which have been reported to be the primary binding sites for metal cations [213, 218, 219].

Sequential FM–AFM images obtained from the same area (Fig. 77) reveal that some of the surface groups change their configuration upon the formation or disappearance of lipid-ion networks [122]. For example, Subunit S1 does not pair with another subunit in Fig. 77a, whereas a pair of Subunits, S3 and S4, appears in the next image (Fig. 77b). The observed structural changes due to the formation and disappearance of the lipid-ion networks indicate that the negatively charged phosphate groups share the positive charge of cations, by which an attractive electrostatic force is exerted on all the headgroups involved in the network. The attractive interaction force mediated through such complex lipid-ion networks should bind the headgroups together and increase the global mechanical strength of the membrane. In fact, it is known that the addition of metal cations increases the mechanical strength of the DPPC bilayer [217]. The FM–AFM images shown here reveal the submolecular-scale origin of this effect of the ions on the mechanical properties of the biological membrane.

13. Future prospects

The imaging speed of high-speed AFM has reached 10–30 frames/s. Importantly, the scanning and oscillating cantilever tip hardly disturbs the physiological functions of delicate biological samples, provided the cantilever free-oscillation amplitude is suppressed to some extent. As described in Section 9.2, the current high-speed AFM with this capacity can be used to film the dynamic biological processes of purified protein systems. The video images can allow us to gain insights into their functional mechanisms in a much more straightforward manner than other techniques. However, even the current performance of high-speed AFM does not fully meet the requirements described in Chapter 4. In this chapter, we discuss whether and how the requirements can be fulfilled and what technologies must be developed further to achieve breakthroughs in this new microscopy field.

13.1. Imaging speed, interaction force, and noncontact imaging

To expand the scope of biological samples that can be studied by high-speed AFM, the imaging rate must be enhanced further and the tip–sample interaction force must be
reduced. Presently, the cantilever is the factor that prevents the achievement of both of these improvements. Generally, to improve the resonant frequency, one must compromise the stiffness and vice versa. The most advanced small cantilevers developed by Olympus deem to have almost achieved the ultimate goal of balancing these two mechanical quantities. Considering their practical use, doubling the resonant frequency to $\sim 2.4$ MHz in water seems to be the upper resonant frequency limit, provided the cantilever compliance is not sacrificed. The question, therefore, is how can this seemingly unsurpassable level be exceeded? If we can significantly improve the detection sensitivity of the tip–sample interaction, we can use stiffer cantilevers and hence attain resonant frequencies that are a few times higher. As described in Section 6.4.2, measuring the peak force instead of the amplitude will improve the sensitivity. Even when we realize this peak force detection as well as a piezoelectric actuator-based $z$-scanner with a realistic upper limit of the bandwidth of $\sim 1$ MHz, the achievable imaging rate is at most 50–60 frames/s for a scan range of $\sim 240 \times 240$ nm$^2$, $\sim 100$ scan lines and $\lambda \sim 10$ nm (see Eqs. (1) and (6)). Thus, even if we continue the present line of development, we will not be amply remunerated. Nevertheless, at present we have to pursue this line while searching for alternatives.

The ultimate minimization of the tip–sample interaction force is attained by noncontact imaging. Nc-AFM has only been realized in a vacuum environment by utilizing a cantilever with a significantly large quality factor in vacuum. If high-speed nc-AFM is realized in liquid conditions, we can use stiffer cantilevers with much higher resonant frequencies, which will promise markedly higher imaging rates. In addition, high-speed nc-AFM will make it possible to image dynamic molecular processes that occur on living cell membranes, a process that is, at present, impossible to study by AFM, since living cell membranes are extremely soft. As many important proteins exist on membranes and are involved in various physiological functions, the noncontact imaging capability will greatly increase the usefulness of high-speed AFM in life sciences. For example, extracellular stimuli (such as drugs, endogenous ligands, light or electricity) are first received by membrane proteins (receptors or ionic channels) and are then transmitted to the cell interior to produce various cellular actions. Membrane proteins also determine the efficiency of drug delivery to the interior of cells. Various diseases are linked with membrane proteins, sugar chains in membrane glycoproteins, and several intracellular molecules involved in signal transductions. Therefore, high-speed nc-AFM will also make a great contribution to the detection and diagnoses, and the discovery and evaluation of new drugs.

We recently realized that an ultrasonic technique combined with AFM (Fig. 78) has the potential of realizing noncontact imaging. In this experimental setup, the sample and cantilever tip are ultrasonically excited at different frequencies of $f_1$ and $f_2$, respectively ($f_1, f_2 >> f_c$) with $|f_1 - f_2| \sim f_c$. When a sample is attached to a uniform substratum, the wavefront of the ultrasound produced by the sample stage oscillation traces the sample topography. This wavefront can probably be detected by a cantilever tip that is close to but not in contact with the sample surface. We recently examined this possibility and found that the tip can sense a surface that it is not in contact with, as shown in Fig. 79. The cantilever started to oscillate at its first resonant frequency $f_c$ when the tip was 2 nm apart from the surface, and its amplitude increased as it approached the surface. As the tip made contact with and was pushed onto the surface, the amplitude decreased. This biphasic behavior of the amplitude signal will not cause problems with the tip–sample distance control if the DC deflection signal is used together with the amplitude signal. Thus, an ultrasonic technique combined with AFM can have the capability of noncontact imaging. However, we
also became aware of the difficulty of maintaining this noncontact detection condition even for a few minutes due to the instability of the ultrasonic wave propagation in liquids. Removing this instability is a key to the realization of high-speed nc-AFM in liquids.
The noncontact imaging capability in liquids has already been achieved by ion-conductance scanning probe microscopy (ICSPM) [220]. Owing to the progress in fabrication techniques for producing very sharp glass capillaries with a small pore at the apex, the spatial resolution of ICSPM has reached a few nm [221]. Immobile protein molecules with a size of \( \sim 14 \) nm on living cell membranes have been successfully imaged [222]. However, it seems difficult to increase the imaging rate of ICSPM; the bandwidth of ion-conductance detection cannot easily be increased, because the ionic current through the small pore of the capillary electrode is very low.

13.2. High-speed AFM for intracellular imaging

A setup similar to that shown in Fig. 78 has been used for subsurface imaging [68]. This method of microscopy has been called scanning near-field ultrasound holography (SNFUH) and has successfully been used for intracellular imaging under ambient conditions [68]. In its application, a high-frequency acoustic wave is launched from under the sample stage, and this wave then propagates through the sample. Materials embedded in the sample with different elastic moduli modulate the phase and amplitude of the propagating acoustic wave. These modulations affect the nonlinear acoustic interference that occurs at the cantilever tip and are consequently reflected in the cantilever oscillation at its resonant frequency. SNFUH has no resolution in the \( z \)-direction. However, using multiple images obtained with different launching angles of the ultrasonic wave, it is probably possible to reconstitute a 3D image. Combining SNFUH with high-speed scanning techniques will enable the high-resolution 3D imaging of various intracellular processes in living cells and organisms which take place spontaneously or as a result of their responses to extracellular stimuli.

As described in Section 13.1, the ultrasonic technique combined with AFM is capable of noncontact surface imaging. When this method is applied to living cells, the cantilever oscillation will be affected by the cell surface structures as well as by the intracellular structures. How can we distinguish these effects? The acoustic wave that appears near the sample surface consists of two components. One component is produced by the oscillation of the whole sample stage, which oscillates the sample surface together. This component is in phase with the ultrasound wave launched from under the sample stage and is used for noncontact surface imaging. We can control its phase relative to the ultrasonic wave propagated through the cantilever to minimize or maximize the acoustic-interference-produced cantilever oscillation at \( f_c \). The other component is that propagated through the sample stage and then the sample, and is used for the imaging of intracellular structures. Its phase relative to the ultrasonic wave propagated through the cantilever can also be controlled to minimize or maximizing the interference-induced oscillation at \( f_c \). Thus, we can select either cell surface imaging or intracellular imaging by tuning the phase difference altered the amplitude of the acoustic-interference-induced cantilever oscillation when the two frequencies (\( f_1 \) and \( f_2 \)) were appropriately chosen.

13.3. High-speed-recognition AFM

A recognition imaging technique (called “Trec”) has recently been developed by Hinterdorfer and colleagues [223,224]. A cantilever tip that is attached to a small number of
probe molecules through a flexible short-linker is oscillated at a frequency below the resonant frequency. Each oscillation is split into approaching and withdrawing parts. The approaching parts of oscillation signals reflect the repulsive tip–sample interaction, providing topographic images. The withdrawing parts reflect the association reaction between the probe molecule and its counterpart on a substratum, providing recognition images. The effective concentration of the probe molecule attached to the cantilever tip depends on the tether length. Assuming that only one probe molecule can interact with the counterpart molecule and a tether length is 2 nm, the concentration is approximately ~50 mM, which is sufficiently high for the association reaction to take place in 20 μs in a system with a typical association rate constant of $1 \times 10^6 \text{M}^{-1} \text{s}^{-1}$. This suggests that recognition imaging at a moderate rate (~0.4 s/frame) is possible provided the cantilever oscillation amplitude is small. As described in Chapters 7 and 11, another method of recognition imaging can be derived from phase-contrast imaging [119] using a cantilever tip functionalized with a specific probe molecule. Its phase-contrast image reveals the arrangement of specific molecules in the sample. As far as the motion freedom required to the association reaction is provided to the probe molecule, this method can use a shorter tether length (<2 nm) for the probe attachment, and therefore, the effective concentration of one probe molecule can be increased to much higher than 50 mM. Consequently, recognition imaging by phase-contrast detection may be performed much faster than by Trec.

13.4. Ultra-high-resolution high-speed AFM

As described in Chapter 12, recent studies using FM–AFM have achieved atomic-resolution imaging in a liquid-environment by minimizing the sensor noise and using small-amplitude cantilever oscillation [76,132,209]. In these studies, conventional cantilevers with a relatively small quality factor (~6) were used. Since a much larger shift in the resonant frequency occurs with small cantilevers, it appears possible to achieve subnanometer-resolution even using high-speed tapping-mode AFM (AM–AFM) equipped with a fast phase-shift detector (Fig. 43) [119] and a sharp cantilever tip. In our recent preliminary experiments using AM–AFM, we filmed atomic-resolution images of a mica surface in water at a frame rate of 6 frames/s (Fig. 80). Even at an imaging rate of 37 frames/s, atomic-resolution imaging was achieved, although further noise reduction remains to be accomplished. AFM with both ultra-high-resolution and high-speed capabilities has enormous potential for the imaging of secondary structures on protein surfaces and their dynamic changes, the mapping of rapid changes in the water layer structures on protein surfaces, the imaging of dynamic lipid–protein interactions on membranes, and the detection of small-ligand interactions with the active site on a protein surface. It is impossible to study these phenomena using “conventional” high-speed AFM.

13.5. High-speed AFM combined with optical microscope

Total internal reflection fluorescence microscopy (TIRFM) [225–227] has been used to track the dynamic biomolecular processes of individual biological macromolecules. TIRFM allows us to observe the translational or rotational motion of individual fluorescent spots that are emitted from fluorophores attached to biomolecules. Since optical
microscopy and high-speed AFM have advantages and disadvantages over each other, combining these techniques into a single instrument would therefore be useful. For example, when high-speed AFM is applied to living cells, the use of this sort of combined system is essential. From the optical image covering a wide area of the sample, we can quickly find a much narrower area to be scanned by AFM without trial imaging at various places. We can visualize the dynamic structural changes of proteins by high-speed AFM while simultaneously recording the optical signals of chemical reactions such as ATP hydrolysis at individual protein molecules. The chemical specification of an object imaged by AFM can be determined from its Raman or fluorescence signals. There have been a number of reports on conventional slow AFM combined with confocal fluorescence [228–231], conventional Raman scattering [232,233], coherent anti-Stokes Raman scattering [234,235], or TIRF [236–238] microscopes. When the simultaneous recording of optical and AFM images is required, some precautions are necessary regarding the luminescence of the cantilever tip [239–241] and the photothermal bending of the cantilever [111]. In tapping-mode AFM, the DC deflection of a cantilever does not affect AFM imaging; thus, the photothermal bending of the cantilever should not be a problem provided its bending is not too large and the light power is stable. If necessary, we can switch off the illumination light during the trace scan and switch it on during the retrace scan [228].

However, building a combined system may not be as easy as the case of conventional AFM, because high-speed AFM imposes various restrictions on the OBD detector and the scanner. Micro-electro-mechanical fabrication techniques, which have been employed to produce self-sensing and/or self-actuation cantilevers [46–48] and sensor-combined scanners [242,243], may be the key to the realization of a combined system as well as to the significant enhancement of high-speed AFM performance.
14. Conclusion

We have described various studies carried out in the last decade with the aim of increasing the imaging rate and spatial resolution of bio-AFM, as well as the basic concepts and principles behind the techniques. Moreover, we have summarized recently performed visualization studies on dynamic biomolecular processes, which were currently infeasible with other techniques. The direct and real-time observation of dynamic biomolecular processes is straightforward and can give deep insights to the functional mechanisms. Therefore, this new microscopy will markedly change our style of considering biological questions. Nevertheless, there are presently only a few setups of high-speed bio-AFM in the world which can capture dynamic biomolecular processes at 10–30 frames/s, and consequently, the user population is limited. Besides, to our knowledge, only two manufacturers are producing small cantilevers for high-speed bio-AFM and the most advanced small cantilevers are not commercially available yet. We hope that this current situation will be quickly improved by manufacturers.

In the near future, high-speed AFM will be actively used to observe a wide range of dynamic processes that occur in isolated proteins, protein assemblies, and protein–DNA complexes, which will certainly induce enormous acceleration of our understanding of how protein functions are produced. In the next stage, more complex systems including living cells and organisms will become targets of high-speed AFM after it is successfully combined with other techniques such as ultrasonic techniques and optical microscopy. The in vivo and in vitro visualization of various processes at the molecular level will become possible including the responses of membrane receptors to stimuli, nuclear envelope formation and disassembly, chromosome replication and segregation processes, phagocytosis, protein synthesis in the endoplasmic reticulum, and the targeting processes of synthesized proteins through the Golgi apparatus. Thus, high-speed AFM-based visualization techniques have great potential to bring about breakthroughs not only in biochemistry and biophysics but also in cell biology, physiology, and pharmaceutical and medical sciences. To open up such unprecedented fields, steady efforts have to be carried out towards expanding the capability of high-speed AFM and the related techniques.

Acknowledgements

We thank N. Kodera, D. Yamamoto, M. Taniguchi, A. Miyagi, H. Yamashita and all previous students for their dedicated studies for developing high-AFM. This work was supported by the Japan Science and Technology Agency (JST; the CREST program and a Grant-in-Aid for Development of Systems and Technology for Advanced Measurement and Analysis) and the Japan Society for the Promotion of Science (JSPS; a Grant-in-Aid for Basic Research (S)).

Appendix A. Supplementary data


References


C. Altenbach, J. Seelig, Ca$^{2+}$ binding to phosphatidylcholine bilayers as studied by deuteron magnetic resonance. Evidence for the formation of a Ca$^{2+}$ complex with two phospholipid molecules, Biochemistry 23 (1984) 3913–3920.


