Invited Review Article: High-speed flexure-guided nanopositioning: Mechanical design and control issues

Y. K. Yong,1,a) S. O. R. Moheimani,1,b) B. J. Kenton,2 and K. K. Leang2,c)
1 School of Electrical Engineering and Computer Science, The University of Newcastle, Callaghan, NSW 2308, Australia
2 Department of Mechanical Engineering, University of Nevada-Reno, Reno, Nevada 89557-0312, USA

(Received 6 June 2012; accepted 5 October 2012; published online 10 December 2012)

Recent interest in high-speed scanning probe microscopy for high-throughput applications including video-rate atomic force microscopy and probe-based nanofabrication has sparked attention on the development of high-bandwidth flexure-guided nanopositioning systems (nanopositioners). Such nanopositioners are designed to move samples with sub-nanometer resolution with positioning bandwidth in the kilohertz range. State-of-the-art designs incorporate uniquely designed flexure mechanisms driven by compact and stiff piezoelectric actuators. This paper surveys key advances in mechanical design and control of dynamic effects and nonlinearities, in the context of high-speed nanopositioning. Future challenges and research topics are also discussed. © 2012 American Institute of Physics. [http://dx.doi.org/10.1063/1.4765048]

I. INTRODUCTION

Nanopositioning systems have enabled significant advances in many fields of micro and nanotechnology.1 For instance, the simple piezoelectric tube actuator (nanopositioner) helped to create one of the most critical tools in nanotechnology, the atomic force microscope (AFM),2,3 a type of scanning probe microscope (SPM). Since its debut, the AFM has emerged as the workhorse tool for studying, interrogating, and manipulating objects and matter at the nanoscale.4 Not only is the AFM extremely versatile, but it can work with conductive and non-conductive specimens, including live biological cells. Its versatility has contributed to major breakthroughs in many research fields and industrial applications including bio-nanotechnology,5–7 nano-metrology,8,9 semiconductor manufacturing,10,11 chemical science and engineering,12 nano-machining and nanofabrication,1,13,14 material science,15–17 and high-density data storage systems.14,18,19 Likewise, piezo-actuated flexure-guided nanopositioning stages are commonly used for positioning optics and many other micro and nanoscale systems. But in recent years, the demand for high-throughput systems such as video-rate SPMs to study fast dynamic processes like the movement of biological cells in real time has posed new challenges for the design and control of high-speed and high-precision nanopositioning systems.6,20–23

This paper surveys and discusses the challenges faced in the design and control of these high-performance systems, particularly the focus is on high-bandwidth, flexure-guided nanopositioners. It is pointed out that the speed, precision, and accuracy of nanopositioners are limited by three major factors: (i) the dynamic behavior of the mechanical positioning system, (ii) the nonlinearities inherent to piezoelectric material, i.e., hysteresis and creep,24–26 and (iii) the performance of the control system.1,26,27 All these issues have to be addressed simultaneously to achieve high-performance, high-speed operation.

This paper is organized as follows. First, a brief overview of the state-of-the-art and the application of scanning probe microscopy is presented in Sec. II. Then, the details of flexure-guided nanopositioners, the main focus of this paper, are discussed in Sec. II B 3, followed by design considerations in Sec. III. A brief review of emerging MEMS-based scanners and nanopositioners are presented in Sec. IV. Afterwards, control issues are presented in Sec. V, followed by future challenges and research in Sec. VI. Finally, concluding remarks are found in Sec. VII.

II. NANOPositionING SYSTEMS, APPLICATIONS, AND scanning PROBE MICROscopy

A. Scanning probe microscopy

Nanopositioners are commonly used in scanning probe microscopy systems, such as the atomic force microscope, to position an ultra-sharp tip relative to a sample surface for studying, interrogating, and manipulating objects and matter at the nanometer scale.3 Imaging in 3D with atomic resolution is one of the most unique features of an AFM. Unlike traditional light and scanning electron microscopes (SEMs) which create images of matter by measuring the intensity of reflected electromagnetic radiation, the sharp tip of an AFM micro-probe reacts to the sample surface and thus the AFM creates images by mechanically “feeling” the surface with the micro-probe. While the resolution of light microscopy is limited by the refraction of visible light, and SEMs on the diffraction of electron beams, the resolution of scanning probe systems is directly related to the precision and accuracy of positioning the SPM probe relative to a sample surface. Additionally, the throughput of AFM (and SPMs in general) is limited by how fast the probe can be positioning over the specimen. For these reasons, recent efforts have been focused...
on improving the mechanical performance of nanopositioning systems.

To better understand the performance of the AFM, consider its main components shown in Fig. 1. An AFM includes an $xyz$ nanopositioner (such as a piezoelectric tube scanner), a micro-cantilever with a sharp tip at its distal (free) end, a laser source directed at the cantilever tip, and a position sensitive photodiode (PSD) to measure the laser light reflected off the cantilever tip. The nanopositioner is used to position the micro-probe within close proximity of the sample, followed by moving the probe tip relative to the sample. Tip-to-sample interactions cause the cantilever to react, such as by deflecting, and the deflection signal along with the lateral position of the probe tip can be used to create topographic images of the sample or to infer various properties of the sample surface. Not explicitly shown in the figure is a control system that is often used to actively control the position of the micro-probe relative to the sample surface to compensate for external disturbances or dynamic and nonlinear behavior inherent in the system.

1. Scanning configurations

The two main scanning configurations for AFM are (a) probe-scanning, where the probe tip is moved relative to a fixed sample, and (b) sample-scanning, where the sample is moved relative to a fixed probe (see Fig. 1). The advantage of probe-scanning is that all of the mechanisms, optics, and sensors are held within one unit, making for a more compact and portable system. The main disadvantage of the probe-scanning design is that the excessive mass of the AFM components (including scanner and optics) can limit the mechanical performance of the scanning mechanism, thus limiting the AFM’s throughput. To increase the instrument’s throughput, such as imaging rates, the sample scanning configuration is often preferred. However, the drawback is that the mechanical performance of the scanner changes with the mass of the sample. Relatively small sample masses are required for high performance, thus securing the samples to the stage platform with strong adhesives may be required to achieve the highest positioning speeds. A combination of probe- and sample-scanning configurations have been implemented to increase the scanning performance, limit the effects of cross coupling, and provide a more versatile platform.

2. Contact modes

The SPM in general can operate in a number of modes. For example, when operated in contact mode, the cantilever is slowly brought into “contact” with a sample. When the probe engages the sample, the probe-sample interaction causes the cantilever to deflect. The cantilever’s deflection is measured and compared to a set-point value, where a vertical feedback controller is used to minimize the difference between the measured and set-point values by moving the nanopositioner in the vertical direction. After the cantilever “lands” on the sample, the $x$- and $y$-axes of the nanopositioner are driven in a raster pattern (see Fig. 1(c)). For constant-force contact mode (see Fig. 2(a)), the vertical feedback controller actively regulates the vertical positioner such that the probe-sample interaction force remains constant during scanning. A topographic image of the sample surface is constructed from the electrical signal used to drive the $z$-axis of the nanopositioner. For constant-height contact mode (see Fig. 2(b)), the vertical feedback controller is deactivated after “landing,” such that the probe hovers at a constant height above the sample surface. An image of the sample surface is constructed from the PSD signal, which correlates with the deflection of the cantilever. The main disadvantage of contact mode is that the cantilever probe tends to exert relatively large lateral forces on the sample. High forces can potentially damage the probe and/or the sample, especially at high scan rates.

3. Dynamic modes

Soft samples such as biological cells are often imaged in dynamic modes to avoid damage induced by the cantilever probe. In this mode, the cantilever is oscillated near the sample surface at a frequency close to the cantilever’s resonance frequency. The repulsive interaction between the probe and sample changes the oscillation amplitude, phase, and frequency. These changes are modulated and used as the error signal in the vertical feedback loop. Again, the topographic image can be constructed from the control signal. Reviews of the dynamic modes can be found in the work by Ando et al. and Abramovitch et al.

B. Types of nanopositioners and speed limitations

There are many factors that limit the operating speed of SPM systems. These factors include: (i) the mechanical bandwidth of the nanopositioner, (ii) the resonance frequency of the cantilever, (iii) the bandwidth of the vertical feedback controller that regulates the probe-sample...
interaction\cite{30,36,39,46,47} and (iv) the data acquisition system.\cite{37,45,48} Herein, the focus is on the mechanical dynamics of the nanopositioner since this is often one of the key limiting factors.

1. **Piezoelectric tube scanners**

One of the simplest and most effective ways to achieve three-axis motion with nanometer resolution is to use sectored tube-shaped piezoelectric actuators.\cite{49} A schematic of a sectored piezoelectric tube scanner is shown in Fig. 3(a). Typically, such a scanner is made from radially poled piezoelectric material consisting of four external electrodes and either a matching set of sectored internal electrodes or a single internal electrode. The benefit of sectored piezoelectric tubes is that three degrees of freedom (DOF) motion with nanometer resolution can be obtained in one unit. For instance, when voltages with equal magnitude but opposite polarity are applied to a pair of opposing electrodes, one side of the tube extends while the opposite side contracts, resulting in bending of the tube. Likewise, the other set of electrodes provide motion in the orthogonal direction, resulting in motion in the $xy$-plane. To displace in the $z$ direction, a suitable signal can be applied to all four sectors. There are other configurations of electrode patterns used for scanning applications, such as an eight-segmented piezoelectric tube scanner.\cite{50} The tube is designed to reduce unwanted angular motions for scanning probe microscopy applications. Recent work has exploited the extra electrodes for both actuation and sensing\cite{51,52,53} to overcome challenges with traditional sensors such as optical displacement sensors.\cite{54}

Piezoelectric tube scanners are one of the most commonly used nanopositioners in commercial AFMs due to their simplicity and cost. However, because of the large length-to-diameter ratio of a piezoelectric tube, when it is fixed at one end, it has a relatively low mechanical resonance. This causes the actuator to be highly susceptible to scan-induced vibrations. In general, the mechanical resonance of a tube actuator is less than 1 kHz in the lateral scan directions. The maximum open-loop (without compensation) positioning bandwidth is 1/100th to 1/10th of the dominant resonance,\cite{27,29,39,49,55} thus limiting the positioning bandwidth.\cite{27,29,39,49,55} Additionally, piezoelectric tube scanners are also known for their large coupling effect from the $xy$ to $z$ axis and from $x$ to $y$ axis. The coupling effect may cause severe distortions in SPM images.\cite{55,56}
2. Shear-piezo and tuning-fork nanopositioners

High-bandwidth nanoscale positioning can be accomplished by using shear piezos [see Fig. 3(b)]. Because of their compact geometry, they have high mechanical resonances. The major drawback, however, is that shear-type piezoactuators provide limited range, typically less than 1 μm. Alternatively for fast nanopositioning, a small tuning fork can be excited to produce scanning motion. This approach has been applied to create video-rate SPMs. The tuning-fork-based sample scanners are mechanically simple, but the small dimensions of the quartz tuning fork limits the scan range and the scanning motion is typically sinusoidal.

3. Flexure-guided nanopositioners

Flexure-guided nanopositioners have emerged as the design of choice for high-speed, high-accuracy nanoscale positioning. These nanopositioners have recently appeared in several commercial AFMs. The key advantage of flexure-guided nanopositioners is their ability for high-speed scanning with low cross-coupling behavior. With an increase in application for high-speed nanopositioning, such as for monitoring fast pace biological cell events and high-speed nano-metrology, flexure-guided nanopositioners are becoming increasingly popular and critical.

Flexure-guided systems exploit the advantages of compliant mechanisms (flexible mechanisms), where a flexible monolithic joint deforms elastically to offer accurate and repeatable fine movements. These flexible structures replace the traditional joints in rigid-link mechanisms, and they offer motion guiding and mechanical amplification as illustrated in Fig. 4. Flexure-guided mechanisms are commonly machined from a single piece of material (monolithic) and no assembly of parts (i.e., between links and joints) is required. Consequently, the number of parts of a flexure-guided mechanism is substantially low, resulting in low mass. These advantages enable the design of compact, light, and fast nanopositioners. Furthermore, the absence of moving and sliding joints provide a considerable benefit, because the problems due to wear, backlash, friction, and the need for lubrication are eliminated. Thus, flexure-guided nanopositioners provide repeatable, reliable, and smooth motions to fulfill the requirement of precise and accurate nanoscale positioning for many nanoscale applications.

Flexure-guided mechanisms are categorized into two main configurations: serial or parallel (see examples in Fig. 5). Both serial and parallel-kinematic configurations have been used to design flexure-guided nanopositioners. Depending on design requirements, each configuration has its own advantages and disadvantages. A serial-kinematic system is created (i) by stacking one piezo-stack actuator in series with another actuator to obtain the corresponding displacements or (ii) by nesting one flexure-guided nanopositioner into another. This configuration allows for relatively high mechanical bandwidth for the fast scanning direction. For example, the bandwidths for the fast scan direction for two prototype systems were measured at approximately 24.2 kHz and 29 kHz for a lateral range of approximately 10 μm × 10 μm. Such scanners produced AFM images at frames rates up to 70 frames per second. However, the high mechanical bandwidth can only be achieved along one axis. This is mainly due to the relatively large mass that the base actuator would have to displace. Another disadvantage is the inability to measure and correct for parasitic motions caused by the cross-coupling effect from other axes. Nevertheless, the cross-coupling effect can be minimized with carefully designed flexures and mechanisms, such as double-hinged flexures. For raster scan purposes, the serial-kinematic design with one high-bandwidth stage is sufficient. Additionally, the serial-kinematic configuration is more cost effective as only one high-bandwidth, high-power piezo-amplifier is required.

Parallel-kinematic configurations have been used in commercial designs. Parallel structures offer high motion accuracy and high mechanical stiffness, leading to high
resonance frequencies. All actuators are fixed relative to ground (base), thus reducing the inertia of the moving sample platform. Parallel structures generally have symmetrical configurations and they are less sensitive to temperature variations. Thus, parallel-kinematic nanopositioners are attractive for accurate, fast nanopositioning applications. However, the cross coupling between the x- and y-axes can be more difficult to deal with compared to serial mechanisms. Nevertheless, with properly designed flexures, nanopositioners with cross coupling as low as −70 dB to −35 dB can be achieved.\textsuperscript{38,61,71–74} Although high-speed parallel-kinematic nanopositioners require high-power, high-bandwidth piezo-amplifiers for both the x- and y-axes, the fast axis of a parallel nanopositioner can be chosen arbitrarily for rastering in AFMs.\textsuperscript{37,38,61,62} Additionally, parallel-kinematic systems can also be used for emerging non-raster scan methods such as spiral,\textsuperscript{75,76} cycloid,\textsuperscript{77} and Lissajous-scan patterns.\textsuperscript{78,79}

4. Comparison of piezo-stack actuated nanopositioners

State-of-the-art flexure-guided piezo-driven stages\textsuperscript{80} both direct drive serial-kinematic\textsuperscript{30,33,35} and parallel-kinematic\textsuperscript{81} configurations, have been developed with mechanical resonances in the tens of kHz range. These designs employ stiff and compact stack piezoactuators [see Fig. 3(c)]. The actuation resonance frequency for a stack piezoactuator held fixed on one end can be approximated by\textsuperscript{82}

\begin{equation}
fa = \frac{1}{2\pi L} \sqrt{\frac{3E}{\rho_{piezo}}} \text{Hz},
\end{equation}

where \( L \) is the length of the actuator, \( E \) is the elastic modulus, and \( \rho_{piezo} \) is the actuator’s density. Substituting in values for \( E = 33.9 \text{ GPa} \) and \( \rho_{piezo} = 8000 \text{ kg/m}^3 \) and assuming an achievable stroke (\( R \)) of 1 \( \mu \text{m} \) per mm of piezo length, it is shown that the maximum scanning frequency is \( f_a = 567,460 R^{-1} \text{ Hz} \). When mass \( m_{end} \) is added to the free end of the actuator, the actuation resonance frequency reduces to

\begin{equation}
fa = \frac{1}{2\pi} \sqrt{\frac{3E}{L}} \left[ AL\rho_{piezo} + m_{end} \right] \text{Hz},
\end{equation}

where \( A \) is the cross-sectional area of the actuator.

For comparison, the relationship between range and resonance frequency for a variety of commercial and custom nanopositioners is shown in Fig. 6.\textsuperscript{82} The range is plotted with respect to the resonance frequency for each stage when provided. When full details are not available for multi-axis positioners, it is assumed that the resonance frequency is provided for the stage with the largest displacement, and therefore, the largest range is plotted with respect to the lowest resonance frequency. Theoretically, the maximum attainable range \( R \) for a given actuation (longitudinal) resonance is \( R = 567,460 f^{-1} \). This value is for a fixed-free piezoactuator with a modulus of elasticity of 33.9 GPa and a density of 8000 kg/m\(^3\), assuming 1 \( \mu \text{m} \) of travel per mm of piezolethn.\textsuperscript{82} This relationship is shown in Fig. 6 as a dashed line. The commercial and custom nanopositioners in Fig. 6 are well below this theoretical limit. A trend line depicted as a solid line shows that for commercial and custom nanopositioners, the relationship is approximately \( R = 30,688 f^{-0.916} \).

C. Image artifacts

There are many sources of artifacts in AFM images, such as probe geometry and sharpness, image processing techniques, nanopositioner performance, and vibrations from the environment.\textsuperscript{83} For a more complete survey on other sources of artifacts, readers are referred to Refs. 56 and 83–85.
AFM image artifacts associated with the nanopositioning system may be caused by (i) exciting the mechanical resonances of the structure,\textsuperscript{33,81} (ii) the cross-coupling effect between the \(x\), \(y\), and \(z\)-axes,\textsuperscript{56,86–88} (iii) the presence of hysteresis and creep in piezoelectric actuators,\textsuperscript{6,26,27,89,90} and (iv) the inadequate bandwidth of the vertical feedback controller due to the low \(z\)-axis resonant modes.\textsuperscript{30,33,39} Figure 7 shows the simulated image artifacts caused by the above mentioned factors.

1. Vibrations

During high-speed rastering, a fast triangular signal is applied to drive one of the lateral axes of the nanopositioner. The sharp corners of the input signal tend to excite the structural dynamics, causing excessive oscillations. As shown in Fig. 7(b), vibration-induced artifacts lead to ripple-like features that appear and distort the image. Examples of experimentally measured vibration-induced image artifacts are found in the literature.\textsuperscript{20,53,91,92}

2. Cross coupling

Piezoelectric tube nanopositioners are known for their cross-coupling effect from \(x\) and \(y\) to the \(z\) axis. The coupling effect is commonly known as “bowing”, where the topographic image is overlaid by a curve [see Fig. 7(g)]. This is due to the bending of the fixed-free piezoelectric tube scanner.\textsuperscript{55,56} The severity of the coupling effect is dependent on the lateral scanning range of the scanner. Some nanopositioners also exhibit cross coupling behavior between the lateral axes due to the non-orthogonality between the two axes.\textsuperscript{56} Parasitic motions between the two axes may result in a skewing effect as shown in Fig. 7(c). Dynamic cross coupling may also exist, where out-of-plane modes are excited by, for example, actuation modes.\textsuperscript{20}

3. Hysteresis

Piezoelectric materials exhibit hysteresis behavior, a non-linear behavior between the input voltage and output displacement of the actuator.\textsuperscript{24,26} For instance, when voltages are applied to a piezoelectric actuator, it experiences a smaller displacement per volt at the start of a scan line than at its end. This behavior results in gradually elongated features in an AFM image as shown in Fig. 7(d). Approaches to deal with hysteresis involved active feedback control\textsuperscript{25} or model-based feedforward control.\textsuperscript{27}

4. Creep

When a large offset voltage is applied to a piezoelectric actuator, it first responds very quickly, by moving to the intended position, which corresponds to the applied voltage. However, the actuator then slowly creeps (over an extended period of time).\textsuperscript{24,25,56,93} This phenomenon can adversely
FIG. 7. Simulated AFM image artifacts. (a1 and a2) Ideal grating without artifacts. (b) Artifacts caused by vibrations, (c) x-to-y cross-coupling motions, (d) the presence of hysteresis at the lateral axes, (e) the creep and drift effect, (f1 and f2) insufficient vertical feedback controller bandwidth, and (g1 and g2) xy-to-z cross-coupling motions.

5. Thermal drift

Material expansion and contraction due to thermal effects can cause significant drift and positioning error. Material expansion is inversely proportional to the melting point of a material. The change of length (from \( l_0 \) to \( l_f \)) of a solid material for a change in temperature (from \( T_0 \) to \( T_f \)) is given by

\[
\frac{l_f - l_0}{l_0} = \alpha(T_f - T_0),
\]

where \( \alpha \) is the coefficient of thermal expansion (CTE) and has units of \( (^\circ C)^{-1} \) or \( K^{-1} \). For nanoscale motions, thermal effects cannot be ignored. Careful material selection and good mechanical design are effective ways to minimize thermal expansion effects. For example, the CTE for aluminum is \( 23 \times 10^{-6}/^\circ C \), while for Super Invar alloy it is only \( 0.3 \times 10^{-6}/^\circ C \), over 70 times lower. Effective practices also include carefully matching the stage material with the materials of surrounding support structures. Also, using materials with high thermal conductivity ensures quick thermal equilibrium, thus minimizing slow transient behavior.

6. Insufficient vertical feedback control bandwidth

The maximum scan speed of a SPM is limited by the bandwidth of its vertical feedback control loop. The effect of inadequate vertical feedback bandwidth on the resulting image is simulated in Fig. 7(f). During a high-speed scan, the nanopositioner with a low-bandwidth vertical feedback loop is unable to track sharp changes in the sample topography. This leads to the “smudging” of the feature edges in the resulting image. In contact mode AFM, when an integral controller is used to regulate the tip-sample force, the vertical loop’s bandwidth can be estimated as \( \omega_n/P \), where \( \omega_n \) is the resonance frequency and \( P \) is the peak magnitude. Thus, by increasing the resonance frequency, e.g., by designing a stiff mechanism, and reducing the peak magnitude, e.g., by using feedback control to decrease quality factor of the nanopositioner, the maximum feedback control bandwidth can be improved significantly to facilitate high-speed operations.

III. DESIGN CONSIDERATIONS

High-speed nanopositioners employed for video-rate SPM imaging require an imaging rate of at least 30 frames per second. In this case, the frame rate establishes the required scan rate for the fast-axis direction. For example, for 30 frames/s imaging with 100×100 pixel resolution, the scan rate of the fast axis should be at least 3 kHz. The parasitic motions along the \( x \), \( y \), and \( z \) axes should be within 1% or less (≤−40 dB). A scan range of approximately 10 μm × 10 μm or more is preferred for imaging samples such as biological cells. In order to achieve the above characteristics and to avoid image artifacts, discussed above, it is necessary to consider (i) the mechanical stiffness of the nanopositioner and flexure design, (ii) material selection, (iii) manufacturing techniques, (iv) actuator properties and electrical issues, and (v) control design methods.
A. Flexure hinges

There are different profiles of flexures, namely, circular,\(^55,95–98\) corner-filleted,\(^{65,99,100}\) and conical.\(^{65,97,101,102}\) These are illustrated in Fig. 8. Each profile provides unique properties to suit different requirements. For example, circular flexures provide accurate rotational motions, i.e., the center of deflection can be estimated to be at the center of the circular flexure.\(^{97,100}\) Circular flexures provide high out-of-plane stiffness but small in-plane deflection (bending). They are suitable for applications which require accurate positioning over a relatively small range. Circular flexures have been used extensively in many flexure-guided micro/nanopositioners. Therefore, many design equations are available to predict their stiffness in various directions.\(^{65,95,96,99,103,104}\) A review of the accuracy of these design equations and a guideline on how to select these design equations can be found in Ref. 98. Corner-filleted flexures, also known as beam-type or leaf-spring flexures, have relatively low bending stiffness, making them attractive for high-displacement-orientated applications. For the same deflection, elliptical flexures have lower maximum stress than corner-filleted flexures. Therefore, elliptical flexures are superior in achieving better flexibility with longer fatigue life. There are other flexure profiles such as parabolic and hyperbolic. Their detailed descriptions and characteristics can be found in Refs. 102 and 105. Table I provides a brief comparison of different flexures in terms of their stiffness, motion accuracy, and maximum stress.

Double-hinged flexures (see Fig. 9) with a “rigid” connection link in the center section have been used in designing micro/nanopositioners.\(^{20,106,107}\) It was reported that the vertical stiffness of a beam flexure can be increased by thickening the center section of a beam flexure to create a double-hinged flexure. Particularly, a good comparison of stiffness of a beam flexure and a double-hinged flexure can be found in Ref. 20. The researchers noted in that study that the effective vertical stiffness of the flexure can be improved to increase the out-of-plane stiffness by (i) increasing the number of flexures used in parallel, (ii) decreasing the flexure length, and (iii) thickening the center section of a beam flexure to create a serial-compliant double-hinged flexure.

It is also important to minimize the \(x\)-to-\(y\), \(y\)-to-\(x\), \(xy\)-to-\(z\), and \(z\)-to-\(xy\) parasitic motions (or cross couplings). For most of the AFM nanopositioners, a \(z\)-actuator is mounted on,\(^5\),\(^3\),\(^7\),\(^6\) or recessed within a \(xy\) stage body.\(^{20,21,38}\) Inertial forces induced by the \(z\)-actuator are insufficient to deform the \(x\) and \(y\) flexures, thus the \(z\)-to-\(xy\) parasitic motion is negligible.

For constant force (topography) AFM imaging, the fastest movements are performed by the vertical \(z\)-actuator.\(^{29,39,108}\) Because the vertical \(z\)-stage is typically set up in a serial-kinematic configuration with the \(x\)- and \(y\)-stages (directly attached to the lateral stage), the inertial forces in the \(z\)-axis stage can induce vibrations on the lateral stage if not properly designed with counterbalancing.\(^{20,28,67}\) Several configurations which limit the inertial effects of a stage on its corresponding nesting stage through counterbalancing are described in Refs. 33 and 28. These configurations [shown in Figs. 10(a)–10(d)] include (a) face mounting, (b) flexure sandwiching (mounting both faces of the actuator to flexures), (c) rim mounting, and (d) inserting the piezoactuator in a hole and allowing the end faces to be free. A novel circular plate flexure designed to further constrain two counterbalanced piezo-stacks to help improve the mechanical dynamics was first presented in Ref. 30 with detailed analysis of the performance discussed in Ref. 28. Similarly, a diaphragm flexure designed to counterbalance inertial forces of piezo-stacks can also be found in Ref. 47.

To capture high-quality AFM images at high-speed, it is important to ensure that the actuation modes occur before the out-of-plane modes (e.g., torsional and twisting modes).\(^{20,28}\) Actuation modes can be measured accurately by placing a sensor in front of a moving target at its direction.

![FIG. 8. Different types of flexures. (a) Circular. (b) Corner-filleted. (c) Elliptical.](image)

![FIG. 9. Beam flexure design: (a) trimetric view of constant cross section corner-filleted flexure beam, (b) trimetric view of corner-filleted serial-compliant flexure, (c) top view, (d) top view with applied load, and (e) expanded view of corner-filleted flexure hinge.](image)

| TABLE I. A comparison of different flexure geometries. |
|----------------|-------------------------------|----------------|----------------|
| Flexures       | Stiffness                     | Motion         | Max.           |
|                | In-plane                      | Out-of-plane   | accuracy       | stress         |
| Circular       | High                          | High           | High           | High           |
| Corner-filleted| Low                           | Low            | Low            | Low-med        |
| Elliptical     | Low-med                       | Low-med        | Med-high       | Low            |
| Parabolic      | Med-high                      | Med-high       | Med-high       | Med            |
| Hyperbolic     | High                          | High           | High           | Med-high       |
of motion. If the first dominant mode is designed to be the actuation (or piston) mode, the control design problem, in some cases, is greatly simplified. It should be pointed out that the out-of-plane modes are difficult to measure and possibly undetectable (depending on the sensor location), and they could be uncontrollable. These modes can, therefore, limit the operating bandwidth of a nanopositioner when they occur prior to the actuation mode. On the other hand, if required, control techniques can be implemented to suppress actuation modes effectively and to also provide accurate tracking.

An effective strategy to ensure that the actuation modes occur before the out-of-plane modes is presented in Ref. 20. It is reported that resonance frequencies for translation $u_i$ and rotational motions $\theta_i$ ($i = x, y, z$) can be estimated by $f_{u_i} = \sqrt{k_i/m_i}/2\pi$ and $f_{\theta_i} = \sqrt{k_{\theta_i}/J_i}/2\pi$, where $m_i$ and $k_i$ are the effective translational mass and stiffness, and $J_i$ and $k_{\theta_i}$ are the effective mass moment of inertia and rotational stiffness, respectively (refer to Fig. 11 for the simplified modeling of the vibration modes). The researchers’ strategy involves carefully designing the nanopositioner geometry and flexure dimensions to ensure that the out-of-plane stiffness-to-mass ratios ($k_z/m_z$, $k_{\theta_y}/J_y$, $k_{\theta_z}/J_z$) are higher than that of the actuation ratio $k_x/m_x$. The nanopositioner presented in their work is a serial-kinematic mechanism where the $x$-axis is the fast axis, with a first mechanical resonance (actuation mode) at 24.2 kHz. Therefore, the stage was designed to have out-of-plane stiffness-to-mass ratios higher than $k_z/m_z$. For parallel-kinematic structures with non-symmetric $x$ and $y$ axes, the nanopositioner would require the out-of-plane ratios to be higher than both $k_z/m_z$ and $k_{\theta_y}/J_y$ ratios.

**B. Flexure stiffness**

Flexure stiffness can be determined analytically or numerically. For instance, for a given flexure or beam geometry, the flexure stiffness can be derived using Castigliano’s second theorem (total strain energy) \cite{20, 109, 110}. As an example, consider a fixed/free beam of rectangular cross section, then the total strain energy is

$$U = U_{axial} + U_{torsion} + U_{bending} + U_{shear},$$

$$U = \int_0^L \left[ \frac{F^2}{2AE} + \frac{T^2}{2GJ} + \frac{M^2}{2EI} + \frac{\alpha V^2}{2GA} \right] dx,$$

where $L$ is the beam length, $A$ is the cross-sectional area of the beam, $h$ is the height, $t$ is the thickness, $E$ is Young’s modulus, $F$ is the force, $T$ is the torque, $M$ is the bending moment, $V$ is the volume flow rate, $\alpha$ is the density, $G$ is the shear modulus, and $I$ is the moment of inertia.

![FIG. 10. Six piezoactuator configurations for vertical positioning: (a) Dual base-mounted piezoactuators, (b) single piezoactuator slid in between two flexures, (c) single piezoactuator held at sides, (d) single piezoactuator slid in hole,\(33\) (e) flexure-guided z-stage with single base-mounted piezoactuator,\(38\) and (f) dual base-mounted flexure-guided piezoactuators.\(20,28\)](image1)

![FIG. 11. Generic flexure-guided stage simplified to single degree-of-freedom systems modeling the dominant modes.](image2)
modulus, \( G = \frac{E}{2(1+v)} \) is the shear modulus, \( v \) is Poisson’s ratio, \( J = h t^3 \left[ \frac{1}{12} - 0.21 \left( 1 - \frac{1}{12} \right) \right] \) is the approximate torsional moment of inertia,\(^{111} \) \( I = \frac{ht^3}{12} \) is the second moment of inertia about the vertical z-axis, \( V \) is the shear force, and \( \alpha \) is a shape factor for the cross section used in the shear equation (for a rectangular cross section \( \alpha = 6/5 \)).\(^{109,111,112} \)

Applying Castigliano’s second theorem, the displacement of a point in a given direction \( u_i, \theta_i \) is the partial derivative of the total strain energy with respect to the applied force, i.e.,

\[
\begin{align*}
  u_i &= \frac{\partial U}{\partial F_i}; \\
  \theta_i &= \frac{\partial U}{\partial M_i}.
\end{align*}
\]

The compliance values are found by dividing the displacement by the applied load, i.e.,

\[
C_{ui,F_j} = \frac{u_i}{F_j}; \quad C_{\theta_i,M_j} = \frac{\theta_i}{M_j};
\]

and are used to form the compliance matrix \( C \). The compliance matrix is defined as the ratio of the displacement \( U = [ux \ uy \ uz \ \theta x \ \theta y \ \theta z]^T \) for a given load \( L = [Fx \ Fy \ Fz \ Mx \ My \ Mz]^T \). Hence, the displacement vector is

\[
\begin{bmatrix}
  u_x \\
  u_y \\
  u_z \\
  \theta_x \\
  \theta_y \\
  \theta_z
\end{bmatrix} =
\begin{bmatrix}
  C_{11} & 0 & 0 & 0 & 0 & 0 & F_x \\
  0 & C_{22} & C_{23} & 0 & 0 & 0 & F_y \\
  0 & C_{23} & C_{33} & 0 & 0 & 0 & M_z \\
  0 & 0 & 0 & C_{44} & C_{45} & 0 & F_z \\
  0 & 0 & 0 & C_{45} & C_{55} & 0 & M_y \\
  0 & 0 & 0 & 0 & 0 & C_{66} & M_x
\end{bmatrix}
\]

This relationship is then used to calculate the stiffness of the flexure for different loading conditions.\(^{20,109,110} \)

Fortunately, standard pre-derived equations for stiffness are available, but limit the designer to pre-determined geometries. The analytical approach offers the advantage of a parametric expression for design and analysis.\(^{20} \)

However, for more complex geometries obtaining closed-form solutions may be challenging. Numerical techniques such as finite element analysis offer a tractable approach to determine the stiffness of various configurations and geometries. Additional advantages are that the stress values can be easily analyzed along with nonlinear material properties, complex boundary conditions, and the designer can iteratively redesign the mechanism to meet certain design objectives.

C. Material considerations

There are numerous material properties that determine the overall performance of a flexure-guided nanopositioner. Properties considered to be of great importance are the Young’s modulus of elasticity (\( E \)), density (\( \rho \)) and the CTE. Materials which have sufficient electrical conductivity to facilitate the wire electrical-discharge-machining (EDM) technique are of great interest for ultra-precision machining (the wire EDM technique will be discussed in Sec. III D). Therefore, many efforts have been made to design fast and stable flexure-guided nanopositioners by using materials with low CTE, high stiffness (high \( E \)), good electrical conductivity, and low density (\( \rho \)).

Aluminium alloys such as Al 7075 and Al 6061 are relatively inexpensive and easily machined and often the material of choice. They have been used to fabricate the vast majority of flexure-guided nanopositioners.\(^{61,70,80,113,114} \)

However, aluminium has a high CTE, as stated in Table II. Thus, it is not suitable for applications where time and temperature changes are the important factors. On the other hand, if the entire measurement can be completed in a short timeframe (i.e., in s or min) with adequate temperature control system, such as in a laboratory environment, aluminium alloys are deemed the most practical design material.

Titanium alloys have a similar \( E/\rho \) value compared to aluminium alloys, but they have a much lower CTE value. Thus, titanium alloys are suitable for applications where positioning accuracy needs to be maintained over a long period of time under varying temperature conditions. However, titanium is not always preferred over aluminium, because it is a more difficult material to machine.\(^{115} \)

Invar, a nickel-iron alloy, which exhibits even smaller values of CTE compared to titanium, can be used for applications with extreme thermal constraints. However, Invar alloys are considerably more expensive and somewhat difficult to machine.\(^{116} \) These alloys are highly magnetic due to their large iron and nickel content.\(^{117} \) Therefore, they are not suitable for applications where magnetic characteristic is a consideration.

D. Manufacturing techniques

The appropriate manufacturing technique for a given design depends on the scale of the positioning stage and the material from which the mechanism is to be machined. The positioning accuracy of a nanopositioner depends on the precision with which its flexures are machined.\(^{118} \)

In particular, it is crucial to ensure all cuts of flexures have parallel sides without tapering. For example, the flexures of the nanopositioner in Fig. 12 have to be as identical as possible to avoid parasitic motions (or out-of-plane motions), caused by uneven stiffness of flexures.

Standard milling, turning, and drilling techniques are best suited for metals, such as aluminium, titanium, and steel. These techniques are best for feature sizes above 1 mm, and at best the machining tolerance is on the order of \( \pm 0.001 \) in (\( \approx \pm 25 \mu m \)). Some of the high-precision machining methods that are often employed to construct nanopositioning platforms are discussed in the following.
1. **Wire electrical-discharge-machining (EDM)**

Wire EDM is one of the most frequently used machining techniques for fabricating monolithic structures such as flexure-guided nanopositioners, owing to its ability to machine complicated shapes with relatively fast cutting speed and accuracy. This method of machining was developed in the 1940s and is based on the erosion of a metallic material in the path of electrical discharges that form an arc between a continuously moving electrode tool (wire) and the workpiece. Dielectric fluid (deionized water) is ejected into the sparking area to flush away the eroded particles. Wire diameters of approximately 100 μm are often used. Traditional machining techniques are used to remove the bulk of the stock before performing EDM. Dimensional accuracy on the order of ±12 μm can be achieved using the EDM process without severely tapering sides and a good surface finish (1.95Ra or less). Note that Ra is the arithmetic mean of the absolute roughness profile ordinates.

2. **Diamond tool machining**

Diamond tool machining is an ultra high-precision machining method which provides high geometrical accuracy, good surface quality, capability to machine complicated profiles and can be used to cut very hard materials. However, this machining process can be labor-intensive, leading to high fabrication cost.

3. **Laser machining**

The laser machining technique operates on the simple principle of using a laser to heat and melt the workpieces, leaving a high-quality surface finish. Laser machining operations have large production rates; therefore, they are used extensively in mass-production of industrial parts. For fabricating nanopositioners, the designer should consider the following challenges with this machining process. For example, cuts may not have parallel sides, a workpiece may have recast layers, i.e., materials on both sides of the cut may re-solidify after cooling, leading to different properties for these layers, and cracks may appear on surface of a workpiece due to poor surface finishing.

### E. Piezo-stack actuators

Piezo-stack actuators are widely used to drive high bandwidth, flexure-guided nanopositioners. These solid-state actuators generally produce small strains, i.e., approximately 0.1% of their original length. However, one of main advantages of piezoelectric ceramic actuators is that they can generate mechanical stresses in the order of tens of megapascals. That is, they can provide tens of newtons of force over a millimeter squared area. Additionally, the response time can be in the micro-second time scale and an acceleration rate of $10^4 \text{ g}$ can be achieved. Therefore, piezoelectric materials are quite suitable for applications that require fast actuation.

A piezo-stack actuator is constructed by “stacking” and glueing multiple thin layers of piezoelectric ceramics (as thin as 100 μm) together such that the polarization direction is aligned with the direction of stroke. All layers are electrically connected in parallel, as depicted in Fig. 3(c). This enables the actuator to be operated at 200 V or less. The force generated by a piezo-stack actuator is proportional to its $d_{33}$ constant. By combining a piezo-stack actuator with a flexure-guided amplification mechanism it is possible to realize nanopositioning platforms that can achieve a relatively large stroke with sub-nanometer positioning resolution.

1. **Preloading piezo-stack actuators**

Because a piezo-stack actuator is made by gluing together multiple layers, the actuator is highly sensitive to tensile forces (pulling forces). Tensile forces can cause cracking of the ceramic material as well as separation of the layers. For a non-preloaded actuator, manufacturers often suggest that the tensile load should be within 5%–10% of the compressive load limit in order to avoid damaging the stack. During high-speed operations, the stack experiences significant pushing and pulling forces due to the inertial forces associated with its own mass and the additional mass that it carries. A preload must be applied to compensate for the tensile force, simply to protect the actuator. Despite these challenges, stack actuators are used extensively in high-speed nanopositioning designs because of their high stiffness, fast response, and high force generation.

For flexure-guided mechanisms, a preload is commonly applied to the piezo-stack actuator using flexures (see Fig. 13). This is an elegant approach since it eliminates the need for additional components such as mechanical springs. The preload on the piezo-stack actuator is achieved when it is first installed into the flexure-guided structure by pressing the piezo-stack against the flexures using fasteners [see Fig. 13(a)] or metal shims. Other preloading methods include applying weights to deform the flexure elastically, allowing a piezo-stack actuator to be placed in its designated position. Weights are then released to restore the original position of the flexures [see Fig. 13(b)]. The flexures will be slightly deformed and the resultant reaction force will hold the actuator in place. It is also recommended that high-strength epoxy be used to help secure the actuator in place.

The required preload can be calculated from the sample mass and its acceleration (computed from the known scan...
trajectory) using Newton’s second law. A procedure for estimation of preload is discussed in Ref. 35.

2. Travel range

The travel range of a piezo-stack actuator reduces when the stiffness connected flexures increases. One can imagine the flexures as a spring load (see Fig. 14) and the total available displacement of a piezo-stack actuator can be estimated as

$$\Delta L = \Delta L_0 \left( \frac{k_T}{k_T + k_e} \right), \quad (7)$$

where $\Delta L$ is the resultant displacement, $\Delta L_0$ is the nominal displacement without external force, $k_T$ is the piezo-stack’s stiffness, and $k_e$ is the stiffness of the external spring (or flexure mechanism). In most nanopositioner designs, this effect is insignificant since the stiffness of the flexures is designed to be 1/20–1/10th the stiffness of the piezo-stack actuator.$^{35,61}$ However, for nanopositioner designs where a large travel range is a priority, the effect of external springs or flexures should be considered carefully to minimize the impact on achievable range.

The free stroke of most piezo-stack actuators is approximately 0.1% of their length, which is insufficient for applications where a large displacement is a necessity. Multiple piezoactuators per degree-of-freedom have been used to increase travel range (and in some cases bandwidth$^{28}$), but at the cost of increased power to drive the piezoactuators at high frequencies.$^{33,81}$ Designs which involve mechanical amplification have been implemented to increase the range without having to increase the actuator’s length.$^{61,69,70,80,107,114,128,129}$

A parallelogram flexure guiding system as shown in Figs. 15(a) and 15(c) is a commonly used amplification mechanism. The amplification ratio of this mechanism can be estimated as $r = (a + b)/a$. The bridge-type amplification mechanism, illustrated in Figs. 15(b) and 15(d), is also commonly used to amplify displacement of a piezo-stack actuator.$^{114,130,131}$ For this bridge-type amplification mechanism, the actuator is extended initially, and then retracted to provide the input displacement. The amplified displacement of this mechanism is perpendicular to the direction of the actuator. The amplification ratio of this mechanism can be estimated as$^{130}$

$$r = \left( \frac{\sqrt{l^2 \sin^2 \alpha + d_1 (2l \cos \alpha - d_e)} - l \sin \alpha}{d_1} \right) / d_i.$$

Figures 15(c) and 15(d) illustrate the finite-element (ANSYS software) simulated displacements of the two flexure-guided amplification mechanisms, respectively.

![FIG. 13. Preloading a piezo-stack actuator. (a) Preloading using flexures. The deformation of the flexures are exaggerated in the diagram. (b) Preloading using fastener. Reprinted with permission from Y. K. Yong, B. Bhikkaji, and S. O. R. Moheimani, IEEE/ASME Trans. Mechatron. (to be published).](image1)

![FIG. 14. Displacement of a piezo-stack decreases as the spring load increases. Original diagram can be found in Ref. 71.](image2)

![FIG. 15. Commonly used amplification mechanisms. (a) Schematic of a level-type mechanism. The amplification ratio can be estimated as $r = (a + b)/a$. (b) Schematic of a bridge-type mechanism. The amplification ratio can be estimated by $r = \left( \sqrt{l^2 \sin^2 \alpha + d_1 (2l \cos \alpha - d_e)} - l \sin \alpha \right) / d_i$. (c) FE simulated displacement of the level-type mechanism. (d) FE simulated displacement of the bridge-type mechanism.](image3)
Adding a mechanical amplification mechanism into a nanopositioner is an effective way of increasing its achievable stroke. However, the added mass of the mechanical amplifier together with the flexible linkages lowers the mechanical resonance of the structure, which leads to a lower bandwidth. To achieve a high resonance frequency, the design of the nanopositioner has to be compact and rigid. Consequently, the length of the flexures and the amplification levers have to be short. However, short flexures and levers will reduce the overall motion of the device. This forces the designer to reach a compromise between the travel range and the bandwidth.

3. Self-heating

During dynamic operation a piezoelectric actuator experiences self-heating, which is known to increase with actuation frequency and amplitude. It is also well known that electromechanical and electrical properties of a piezoelectric actuator can vary with temperature. Therefore, good thermal management is necessary for predictable performance, as well as to prevent premature failure and depoling of the piezoelectric material used. Heat management can be challenging due to the low thermal conductivity of the piezoelectric material. Creative heat sink designs can be employed to minimize the heating, but should not hinder the motion of the actuator. When an piezoelectric actuator is packaged in a metal case, the air gap can act as an insulator. Several vendors offer specialized actuator configurations that are based on methods such as the “ThermoStable” technique, in order to improve heat management.

4. Electrical Considerations

Due to the highly capacitive nature of piezo-stack actuators, amplifiers with large current and power dissipation capabilities are needed to drive the actuators at high-speed. The electrical and drive issues related to high-speed operation of piezo-stack actuators are discussed in the literature. The current $I_p$ and the associated power dissipation $P_d$ in a linear amplifier can be estimated by

$$I_p = s CV_p$$  \hspace{1cm} (8)$$

$$P_d = I_p (V_s - V_p)$$  \hspace{1cm} (9)$$

where $s$ is the Laplace variable, $C$ is the piezo-stack capacitance, $V_p$ is the load voltage and $V_s$ is the supply voltage. For a nanopositioner that is to be operated at high speeds, selecting an actuator with a small capacitance is prudent. This will ensure that the actuator can be driven by an amplifier with sufficient current and power dissipation capabilities. Typically, a smaller piezo-stack actuator possesses a smaller capacitance, but it has a lower stiffness and provides a smaller stroke. These issues should be factored into the selection of a piezo-stack actuator for high-speed nanopositioning. A comparison of piezo-stack actuators with various cross-sectional areas versus their maximum travel ranges and stiffness is found in Ref. 20.

IV. MEMS-BASED NANOPOSITIONERS

Research on microelectromechanical systems (MEMS) that started about 30 years ago as a scientific curiosity, has evolved into a multibillion dollar industry. Despite its immense commercial success, MEMS development is still in a highly exploratory phase with numerous ideas considered to be worth investigating. An area that has attracted interest in recent years is MEMS nanopositioning. The reason can be traced back to the research on probe-based data storage, over the last ten years, which is aimed at developing a new form of non-volatile memory for mobile and archival storage.

In a probe-storage memory, write and read functions are based on the mechanical scanning of a storage medium with respect to an array of thousands of probes, each of which is equipped with a nanometer-sharp tip at its extremity. These probes are used to store digital information, as tiny indentations, on a polymer storage medium. The scanning is performed using a 2D nanopositioner, often referred to as a scanner, that moves the storage medium relative to the probe array. A very attractive feature of probe-storage is its ability to achieve extremely high data storage densities, on the order of several Tb/in. Within this framework, feature sizes as small as 25 nm have already been demonstrated, making control and nanopositioning capabilities of critical importance to this technology.

A. Actuation methods in MEMS nanopositioners

1. Electromagnetic MEMS nanopositioners

A number of MEMS 2D nanopositioners for probe-based data storage has appeared in the literature. The preferred type of actuation in probe-storage literature appears to be electromagnetic. A MEMS electromagnetic actuator consists of a tiny permanent magnet and a coil, and is a linear device, i.e., the displacement is a linear function of the current applied to the coil. This makes the control design problem more straightforward to address. Due to the current-driven nature of these actuators, they can be operated at relatively low voltages, which makes them attractive for mobile applications. MEMS electromagnetic actuators must be assembled separately and then integrated with the MEMS device since they are not fully compatible with standard micro-machining processes. This is a disadvantage of this form of actuation. Energy consumption of electromagnetic actuators is also relatively high.

The MEMS nanopositioner designed by IBM, and reported in Ref. 137 consists of a square-shaped scanner, 6.8 mm on each side and a pair of electromagnetic actuators. The actuators can move the scan table in x and y directions about 120 μm. Each actuator consists of a pair of permanent magnets glued into a silicon frame with a miniature coil. The motion of each actuator is coupled to the scan table by means of a mass-balancing scheme that makes the scanner robust against external shocks and vibrations. Other electromagnetic nanopositioners are reported in the literature. However, none appears to be as refined as described in Ref. 137.
2. Electrostatic MEMS nanopositioners

Electrostatic actuation is the most efficient method of actuating MEMS devices. Electrostatic MEMS actuators are well studied and their design and fabrication are straightforward.\textsuperscript{141} Amongst all possible forms of electrostatic actuation, comb drives are the most widely used. A comb drive consists of two interdigitated structures, shaped like combs, that attract each other when charged oppositely. The force that is generated by a pair of comb fingers, however, is rather small. Thus, many fingers are often needed to generate sufficient force. The areal efficiency of electrostatic actuation is, therefore, less than electromagnetic actuation.

Compared with an electromagnetic actuator, a comb drive is much easier to fabricate. Its consumption of energy is also much lower. However, a comb drive is typically operated at higher voltages, which may require additional dc to dc conversion circuitry in mobile applications. Carley et al.,\textsuperscript{142} reports the design of an electrostatically actuated nanopositioner that achieved a travel range of 50 $\mu$m at a 120 V drive. This design was further refined by Alfaro and Fedder in 2002,\textsuperscript{143} where a stroke of 100 $\mu$m was obtained. Other examples of electrostatically actuated MEMS nanopositioners reported in the literature include Refs. 144–146.

3. Electrothermal MEMS nanopositioners

Chevron-type electrothermal actuators are the most widely used form of electrothermal actuation in MEMS.\textsuperscript{147} These actuators use an array of silicon beams facing each other in pairs to generate one-directional stroke. When the beams are heated, they expand and ultimately buckle. The beams are designed with a pre-bend angle. Hence, the structure has a tendency to move in-plane (parallel to the substrate). These actuators can operate under low voltages, can generate large forces, and enjoy a high degree of vibration resistance due to their stiff structures.

Electrothermal actuators depend on the dissipation of electrical energy through Joule heating in resistive layers at any position except rest. Thus, there is a penalty for holding a position. This amounts to a significant disadvantage when compared with electrostatic devices, which dissipate power only during the times that they are charging and discharging. In this respect, when power consumption is a critical performance parameter and voltage and scan range constraints are loose, electrothermal actuation cannot compete with electrostatic actuation. To reduce power consumption, greater thermal isolation is needed. However, the consequent reduction in thermal dissipation negatively impacts the speed of the device.

A one-degree-of-freedom nanopositioner with electrothermal actuation and sensing is reported in Ref. 148 and its control design problem is discussed in Ref. 149. A 2D nanopositioner with electrothermal actuation and capacitive sensing is reported in Ref. 150 but only open-loop results are presented. A new method of electrothermal bi-directional actuation based on z-shaped beams is reported in Ref. 151.

B. Sensing for control in MEMS nanopositioners

To measure the motion of the scan table, position sensors are incorporated into MEMS nanopositioners. The most widely used displacement measurement technology in MEMS is capacitive sensing. The reasons are the simplicity of the sensor element itself, low power consumption, and good stability over temperature.\textsuperscript{141} Many capacitive transducers display a nonlinear capacitance vs. displacement characteristic. Thus, feedback is commonly used to convert the signal to a linear output. Alternatively the actuator’s static nonlinearity can be inverted and then cascaded at the input to render the overall system linear. This sensing method has been used in two nanopositioners reported in Refs. 152 and 140 where nanometer-level positioning accuracies have been achieved. They have also been used in Ref. 153 where these electrostatic actuators are exploited for simultaneous sensing. An alternative approach to simultaneous sensing and actuation in electrostatic MEMS drives is introduced in Ref. 154.

An innovative electrothermal position-sensing method for MEMS-based probe-storage devices was reported in Ref. 155. In this approach displacements of the scan table, along x and y directions, are measured using two pairs of electrothermal position sensors, placed directly above the scan table. The sensors consist of thermally isolated, resistive strip heaters made from moderately doped silicon. Displacement of the scan table results in a change in the temperature of these heaters and subsequently a change in their electrical resistance. By measuring resistance variations, global position of the scan table can be determined. The first lateral resonance frequencies of this nanopositioner are about 120 Hz. A resonant controller\textsuperscript{156, 157} that was designed for this stage achieved a positioning accuracy of approximately 0.25 nm, which is comparable with the diameter of an atom.\textsuperscript{19} A comparison of electrothermal and electrostatic sensing methods in terms of sensor noise can be found in Ref. 158. An analysis of electrothermal sensing (and actuation) is presented in Ref. 159.

C. On-chip scanning probe microscopy

There have been several attempts to use MEMS technology for scanning probe microscopy, dating back to 1992.\textsuperscript{160} This reference reports the design of parallel plate MEMS actuators that move a micro-machined probe instead of a sample for scanning tunnelling microscopy. The travel range of this nanopositioner was 200 nm, which limited its application. More recently there have been efforts to build a fast MEMS scanner, for high-speed STM, with lateral resonance frequency of above 10 kHz and a scan range of at least 3 $\mu$m, as reported in Ref. 161. Apart from scanning tunnelling microscopy, there have been earlier efforts to build MEMS scanners for atomic force microscopy, e.g., the scanner reported in Ref. 162, which includes an AFM-like probe. However, it is not clear if the device was used to study surfaces.

Feasibility of on-chip atomic force microscopy was demonstrated in a recent study, reported in Refs. 163 and

164. The device, shown in Fig. 16, is a nanopositioner fabricated using a multi-user silicon-on-insulator (SOI) micromachining process. It consists of a central positioning stage (scan table), actuating combs, connecting springs, and the substrate layer. Two comb-drive actuators are positioned at adjacent edges of the device in a unidirectional pull configuration to actuate the stage in the x and y directions. The central positioning stage is a solid 3 mm × 3 mm structure, and is connected to the electrostatic actuators using beam springs which are oriented parallel to the direction of actuation of the combs. The nanopositioner is designed such that the cross coupling between the two lateral axes are reduced to a minimum. The main structural features of the device including the positioning stage, beam springs and comb drive actuators are fabricated from 25 μm-thick doped silicon.

The nanopositioner can move slightly over 15 μm in either direction, which is remarkable when compared with macro-sized nanopositioners. Its first lateral resonance frequency is about 820 Hz, which makes it comparable with piezoelectric tube scanners used in commercial AFMs. However, it can be redesigned to have a much higher resonance frequency. The scan table was designed with a repeated array of cylindrical gold features, that were imaged, in tapping mode, by operating the device in a commercial AFM, where the rastering was performed by the MEMS stage. The resulting image is shown in Fig. 17.

FIG. 17. Atomic force microscopic image of the features on the nanopositioner stage in Fig. 16. The image was obtained in tapping mode, and the rastering was performed by the MEMS stage. Reprinted with permission from A. G. Fowler, A. N. Laskovski, A. C. Hammond, and S. O. R. Moheimani, IEEE/ASME J. Microelectromech. Syst. 21, 771 (2012). Copyright 2012, IEEE/ASME.

V. CONTROL FOR NANOPOSITIONING

The performance of nanopositioning systems is greatly affected by the mechanical dynamics of the motion mechanism,20 and for piezo-actuated designs, the nonlinearities such as hysteresis can drastically limit precision.27 Therefore, control plays an important role in achieving high-quality SPM operation at high scan speeds. In this section, a review of popular control techniques for nanopositioning is presented. Readers are also referred to more detailed reviewer papers.1, 3, 26, 27

A. Performance degrading effects

The performance of a nanopositioner is affected by induced structural vibration, hysteresis and creep in the piezocantilever, cross-coupling behaviors, external disturbances, and drift due to temperature variations.165 Particularly, the vibration effect limits the operating bandwidth and is often caused by command signals exciting the flexible modes of the mechanical structure.24 Typically, scan rates (i.e., scan frequencies) are restricted to less than 1/10th–1/100th of the dominant mechanical resonance frequency. However, higher operating speed can be achieved by using stiffer piezoactuators with higher resonance frequencies, for example, Ando and co-workers166 used a stiff piezo with a resonance frequency of 260 kHz to develop an AFM to image biological macromolecules in action. Hysteresis, a nonlinear behavior between the input voltage and output displacement of the piezoactuator, causes SPM image distortion,24, 167 instability of the closed-loop,168 and loss in calibration.169, 170 By operating the nanopositioner over a relatively small range, say less than 10% of full range, hysteresis can be avoided. However, operating over such a narrow range limits the ability of the actuator for long-range motion with sub-nanometer resolution. During
slow and static positioning, creep and drift effects can result in over 30% error in positioning. To avoid these issues, the positioner can be operated at a relatively fast rate (above 1 Hz scanning frequency). Mechanical designs which lack perfect symmetry or involve high inertial loads can cause imbalances leading to cross-coupling effects which further limit precision and operating speed. Even commercially available piezoelectric tube actuators have a tendency to exhibit dynamic cross coupling as a result of imperfections in the manufacturing process. Finally, temperature changes cause the characteristics of the piezoactuator to vary with time. Even thermal expansion of the motion mechanism can result in positioning error. Although the aforementioned behaviors can be avoided, control is often required to overcome these issues to ensure high-performance, nano-precision positioning.

The control approaches for nanopositioning fall under two main categories: feedback control and model-based feedforward control, each with its advantages and disadvantages. For example, feedback control schemes such as traditional proportional-integral-derivative (PID), state-feedback, gain scheduling, and \( \mathcal{H}_\infty \) control handle modeling errors and are robust with respect to parameter variation. Although feedback control has been applied extensively in nanopositioning, the low gain margin in piezopositioners tend to limit the performance, i.e., high feedback gain tends to destabilize such systems. In practice, a compromise is sought between performance and stability; feedback gains (such as proportional, derivative, and integral terms) are adjusted to improve performance without instability. Additionally, the precision of feedback controllers can be sensitive to sensor noise and can be bandwidth limited. Feedforward control, which typically involves inverting the dynamics of the positioning system, provides good performance with accurate models and the absence of disturbances.

B. Techniques for high-speed nanopositioning

Nanopositioning systems are generally lightly damped mechanical structures. As shown in Fig. 18, which is the measured frequency response of the nanopositioning system depicted in Fig. 4(b), a sharp peak at approximately 520 Hz suggests a lightly damped resonance peak. Because of this, high-frequency command signals and/or exogenous noise can excite the dominant resonance mode(s) causing excessive vibration and oscillation in the output response. With careful mechanical design, the resonance modes can be designed to excite the dominant resonance mode(s) causing excessive vibration and oscillation in the output response. With careful mechanical design, the resonance modes can be designed to occur primarily in the actuation direction as previously discussed, thus resulting in more favorable dynamic characteristics for control. Nevertheless, both feedback and feedforward based approaches have been proposed to address the effects of the lightly-damped vibration modes.

Feedback controllers such as standard PI controllers are commonly used due to their robustness and high low-frequency gain, but PI controllers offer limited closed-loop bandwidth. It was shown that for a nanopositioner with second-order dynamics, the maximum closed-loop bandwidth \( \tilde{\omega}_{BW} \) is

\[
\tilde{\omega}_{BW} \leq 2\omega_n\zeta, \tag{10}
\]

where \( \omega_n \) is the natural frequency and \( \zeta \) is the damping ratio of the positioning system. With this, low damping ratio severely limits the closed-loop bandwidth of a PI-controlled system.

The simplest and most popular approach to suppress the sharp resonance behavior is to use notch filters or inversion-based filters in the closed-loop. By doing this, the gain-margin and closed-loop bandwidth can be improved, even up to the resonance frequency. However, one of the major disadvantages is the need for an accurate model of the resonance frequency. Even a small shift in the resonance behavior as low as 10%, say caused by changes in the payload, can lead to instability of the closed-loop system. Robust \( \mathcal{H}_\infty \)-based and loop-shaping control techniques have been considered for high-bandwidth operation. Other damping control techniques include positive position feedback (PPF) control and polynomial-based controller designs. In the PPF approach, an inner feedback loop is used to damp the highly resonant mode of the positioner. Application of the PPF controller on a commercial AFM system demonstrated damping as well as cross-coupling compensation for a piezoelectric tube scanner. Integral resonant control (IRC), a method introduced as a means for augmenting the structural damping of resonant systems with collocated sensors and actuators, damps the vibration and makes the system robust to the unmodeled dynamics and resonance frequency variations due to changes in the payload. The block diagram of an advanced version of the IRC approach which includes integral action and a feedforward input \( u_f \) is illustrated in Fig. 19, where \( G(s) \) is the plant and the regulator \( C(s) \) is

\[
C(s) = \frac{-k}{s - kD_f}, \tag{11}
\]

\( k \) is the constant, and \( D_f \) is the feedthrough term. Other active damping techniques include receding horizon control as well as active and passive shunt damping techniques for piezoelectric tube scanners.
Feedback controllers react to the measured tracking error; however, feedback controllers compensate or anticipate for deficit performance. As the operating speed of a piezobased nanopositioner increases, performance is limited by essentially dynamic and hysteresis effects.\cite{27,90} Both of these behaviors can be compensated for using feedforward control by inverting the plant dynamics and nonlinearity (see Fig. 20). Both the dynamic and hysteresis effect are often intertwined,\cite{169} but can be modeled for feedforward control using the cascade structure shown in Fig. 20. Specifically, the range-dependent hysteresis effect is treated as a rate-independent, input nonlinearity represented by $\mathcal{H}$. The vibrational dynamics (and creep effect) are captured by the linear dynamics model $G(s)$. The cascade model structure is used extensively to model piezoactuators and similar systems.\cite{24,190}

Feedforward control for piezo-scanners was first demonstrated by Croft et al.,\cite{24,191} where for a minimum phase system the feedforward input $u_f$ to track a desired trajectory $y_d$ was found by inverting the piezo-scanner’s dynamics $G(s)$ as

$$U_{ff}(s) = G(s)^{-1}[Y_d(s)]. \quad (12)$$

It is pointed out that the model-based feedforward approach cannot correct for positioning errors due to plant uncertainties and external perturbations,\cite{192} but feedback control can be used to minimize the uncertainties and then combined with feedforward control.\cite{25,72,193-195} For nonminimum phase systems, a stable inverse feedforward input can be found — although noncasual — by using the Fourier transform approach\cite{196}

$$U_{ff}(j\omega) = G(j\omega)^{-1}[Y_d(j\omega)]. \quad (13)$$

The time-domain solution $u_f(t)$ is then found by the inverse Fourier transform. This approach was applied to AFM imaging in Ref. 24 and variations of the approach which took into account input saturation and the tracking precision have been considered.\cite{25,191,197,198} Other feedforward methods include input shaping techniques\cite{199} and shaped-triangle technique for high-speed AFM.\cite{81}

Many nanopositioning applications, for example, SPM imaging, require tracking a repeating (periodic) trajectory. The scanning motion in this case repeats from one cycle to the next. Rather than model and invert the dynamics and nonlinearity of a positioning system for feedforward control, for repeating scanning motion, the feedforward input can be found using iterative techniques for both dynamics and hysteresis compensation.\cite{171,200,201} This approach is commonly referred to as iterative learning control (ILC), where recently an approximate inverse of the plant dynamics $G$ was used in the input-update law:\cite{200}

$$u_{k+1}(\omega) = u_k(\omega) + \rho(\omega)\hat{G}^{-1}(\omega)[y_d(\omega) - y_k(\omega)], \quad (14)$$

to achieve faster convergence ($k$ is the iteration number). A block diagram of the frequency-domain implementation of the ILC control law (14) is shown in Fig. 21. Particularly, the input $u_k(t)$ and tracking error $e_k(t)$ are Fourier transformed, then the ILC algorithm is applied, producing the updated input $u_{k+1}(\omega)$. The time-domain input $u_{k+1}(t)$ is obtained by inverse Fourier transform. Then the input is applied to the system and the process is repeated.\cite{88,202} A model-less ILC approach where the inverse $G^{-1}$ is obtained from the measured input-output data from the previously iteration step was considered for AFM applications.\cite{203-205}

Recently, a feedback-based approach known as repetitive control (RC) was exploited for tracking periodic trajectories in nanopositioning applications including high-speed and metrological AFMs.\cite{20,206,207} The RC provides high gain at the harmonics of the reference trajectory by incorporating a signal generator within the feedback loop.\cite{208,209} For piezo-based nanopositioning systems, the RC can be designed for low tracking error in the presence of dynamic and hysteresis effects.\cite{210,211} Some advantages of this approach include minimal system modeling, robustness due to the feedback structure, and straightforward digital implementation; however, a priori knowledge of the period of the reference trajectory is required.

C. Minimizing cross coupling

Careful mechanical design can significantly minimize cross-coupling effects.\cite{20,86} However, during high-speed operation such as fast raster scanning in video-rate AFMs, the cross-coupling effect cannot be ignored as inertial effects and poor boundary conditions can cause out-of-plane modes to be excited.\cite{28} Control methods to deal with cross-coupling issues include inverse-based feedforward control,\cite{88} $\mathcal{H}_\infty$-based control,\cite{86,212,213} iterative feedforward approach,\cite{214} and adaptive control.\cite{215}

D. Compensating for nonlinearity

Nanopositioning systems commonly employ piezoelectric actuators for positioning, and the inherent nonlinearities such as hysteresis and creep in piezoelectric materials cause significant positioning error.\cite{24} A detailed discussion of hysteresis can be found in the three-volume collection, “The Science of Hysteresis,” edited by Bertotti and Mayergoyz.\cite{216-218}

Both feedback and feedforward based controllers have been used to minimize these effects, but in general the nonlinearity is handled with integral feedback control,\cite{219} high-gain feedback control,\cite{25} force feedback method,\cite{178} and often through an internal feedback loop, such as PID\cite{220} and

![FIG. 20. The feedforward control approach for piezoactuator system.](image-url)

![FIG. 21. Block diagram of frequency-domain ILC scheme.](image-url)
Ishlinskii model and the Prasich model\textsuperscript{221–227} to compensate controllers exploit phenomenological models such as the Prandtl-Ishlinskii model and the Maxwell slip model.\textsuperscript{232} The Prandtl-Ishlinskii model, for example, models symmetric hysteresis loops with a smaller set of parameters, and thus the model inversion is more efficient and less computationally demanding for real-time feedforward control.\textsuperscript{210, 222, 223}

By controlling the charge (rather than voltage) applied to the piezoeactuator, hysteresis can be reduced. Early demonstration of the effectiveness of charge drives have appeared in the literature.\textsuperscript{233–236} More recently, advancements in the design of charge drives for piezoelectric actuators have resulted in effective designs for operating at low and high frequencies.\textsuperscript{237–239} These charge drives have also been integrated with inversion-based feedforward control for piezoelectric actuators\textsuperscript{240} and SPM.\textsuperscript{239}

Piezoelectric creep leads to significant error when positioning over extended periods of time (for example, during slow-speed scanning operations).\textsuperscript{241–243} The effect is also present for high-frequency signals with a relatively low DC component. Control of creep include feedback-based methods\textsuperscript{244} and feedforward control.\textsuperscript{24, 93, 171, 244}

### E. Dealing with thermal drift

Thermal expansion can severely limit the precision, repeatability, and overall performance of a nanopositioning system. Specifically, the result includes temperature dependent drift in the motion and high thermal stresses which ultimately lead to cracking, warping, or loosening of components. Unfortunately, for positioning at the nanometer scale, thermal effects cannot be ignored. In a typical AFM operated in ambient temperature, even a small 1° change in temperature can cause as much as 50 nm drift.\textsuperscript{245} Feedback controllers have been used successfully to minimize thermal drift.\textsuperscript{246} For SPMs, an effective method to deal with drift is to measure and compensating for the effect through consecutive SPM images.\textsuperscript{247–250}

### VI. FUTURE CHALLENGES AND RESEARCH

High-speed nanopositioning devices have attracted significant interests in numerous ultra-precision applications. A substantial research effort, ranging from mechanical design, actuators and sensors to high-bandwidth drive and control of flexure-based nanopositioners, have been seen over the past three decades. However, one of the challenges which is still faced by researchers is the trade-off between speed and range. Therefore, more innovative solutions to achieve high-speed, large-range scanning are needed.

#### A. Dual-stage lateral scanning

Dual-stage vertical positioning techniques have been proposed for SPM applications where a long-range positioner is used to approach a cantilever to a sample and a fast short-range positioner is used for high-bandwidth scanning.\textsuperscript{31, 39, 251} Similar concept can be applied to the lateral axes when multiple locations of a sample surface are required to be interrogated rapidly. The long-range stage is used to move the cantilever to a target area and the short-range stage is used for high-speed rastering. One of the challenges of dual-stage positioners is to prevent vibrations being transmitted from the short-range stage to its surrounding structures. This can be achieved by implementing the inertial counterbalance technique on the lateral axes to minimize impulsive forces of a piezo-stack actuator.\textsuperscript{33} Another possible solution is to implement innovative control techniques for vibration cancellation and accurate tracking, such as spatial-temporal control for high-speed and high-accuracy positioning.\textsuperscript{252} This can be achieved by using high-bandwidth position sensors,\textsuperscript{20} self-sensing piezo-stack actuators,\textsuperscript{178} field-programmable gate array based control\textsuperscript{230} and field-programmable analog array based control.\textsuperscript{21, 253, 254}

#### B. Parallel probe scanning

High-speed and large-range SPM scanning can be achieved simultaneously by using multiple parallel probes.\textsuperscript{255} The deflection of each individual probe can be detected by using self-sensing cantilever with piezoelectric/piezoresistive sensors\textsuperscript{256, 257} or optical interference technique.\textsuperscript{258} However, the cantilever array system will increase the control effort due to the high-bandwidth parallel control scheme. The implementation of such scheme remains challenging.

#### C. New materials and advanced manufacturing methods

Advanced metal matrix composite (MMC) materials combine a metal and at least another part, such as a ceramic, to form a material with enhanced mechanical properties compared to traditional materials. For example, an aluminum silicon carbide metal matrix material offers up to 60% greater stiffness than traditional aluminum alloy with little to no increase in density.\textsuperscript{259} Recently, the dynamic response of flexure-guided stages can be more easily tailored using the MMC,\textsuperscript{260} for example, by using the MMC the actuation mode of the stage can be designed to occur before the out-of-plane modes. The improvement in performance of using the MMC outweighs the disadvantages that include increase cost of the material and lower manufacturability compared to traditional aluminum alloys.

### VII. CONCLUSIONS

This paper discussed the mechanical design and control issues for flexure-guided high-speed nanopositioners. These nanopositioners are designed to move samples with sub-nanometer resolution with bandwidth in the kilohertz range. Emerging applications for high-speed nanopositioners include video-rate SPMs and high-throughput probe-based nanomanufacturing and nanometrology. Novel flexure
designs and the key techniques for improving the mechanical performance were described. The performance of nanopositioning systems is greatly affected by the mechanical dynamics of the motion mechanism, and for piezo-actuated designs, the nonlinearities such as hysteresis can drastically limit high-precision high-speed positioning. To address these issues, control approaches have been recently developed. A survey of control techniques, including model-based feedforward and advanced feedback controllers, commonly used to achieve high-precision high-speed positioning was presented. Finally, emerging trends and challenges such as dual-stage lateral scanners and advanced materials that offer enhanced mechanical performance were highlighted.

ACKNOWLEDGMENTS

The two authors wish to thank the Australian Research Council for supporting their research that underlies this review. They also wish to thank all of their colleagues at the Laboratory for Dynamics and Control of Nanosystems. K. K. Leang and B. J. Kenton gratefully acknowledge support from the National Science Foundation (NSF) Grant Nos. CMMI 0910570 and DUE 0852756 and the Nevada NASA Space Grant Consortium. Additionally, they are grateful to Dr. A. J. Fleming for the insightful discussions and providing the PiezoDrive amplifiers for their experimental work.


