Advances in Industrial Control

Andrew J. Fleming
Kam K. Leang

Design, Modeling and Control of Nanopositioning Systems

Springer

Advances in Industrial Control

Series editors

Michael J. Grimble, Glasgow, UK
Michael A. Johnson, Kidlington, UK

For further volumes:
http://www.springer.com/series/1412

Andrew J. Fleming · Kam K. Leang

Design, Modeling and Control of Nanopositioning Systems

Springer

Andrew J. Fleming
School of Engineering and Computer Science
University of Newcastle
Callaghan, NSW
Australia

Kam K. Leang
Mechanical Engineering
University of Nevada, Reno
Reno, NV
USA

ISSN 1430-9491 ISSN 2193-1577 (electronic)
ISBN 978-3-319-06616-5 ISBN 978-3-319-06617-2 (eBook)
DOI 10.1007/978-3-319-06617-2
Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014938477

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

To my family

Andrew J. Fleming

To Allyson, Norie, Phirin, the Newcomer and The Squeaker

Kam K. Leang

Foreword

The series *Advances in Industrial Control* aims to report and encourage technology transfer in control engineering. The rapid development of control technology has an impact on all areas of the control discipline, such as new theory, new controllers, actuators, sensors, new industrial processes, computer methods, new applications, new philosophies…, new challenges. Much of this development work resides in industrial reports, feasibility study papers, and the reports of advanced collaborative projects. The series offers an opportunity for researchers to present an extended exposition of such new work in all aspects of industrial control for wider and rapid dissemination.

The range of monographs that appear in the *Advances in Industrial Control* series is very wide and from time to time the Editors are able to welcome into the series a monograph that seems destined to become a definitive text for its field. This monograph, *Design, Modeling and Control of Nanopositioning Systems* by Andrew J. Fleming and Kam K. Leang is such an example. The monograph is a comprehensive treatise on designing and implementing control systems for nanopositioning systems. Such control modules are found in devices like the atomic force microscope. To give context to the monograph, a nanometer (nm) is the unit 1×10^{-9} m and an atomic force microscope has a resolution of 0.01 nm. Thus, for example, with the diameter of iron atoms at 0.28 nm, gallium atoms at 0.26 nm, and gold atoms at 0.27 nm, an atomic force microscope can explore the atomic topography of samples.

The narrative trajectory of the monograph assigns the first five chapters to the physical components used in nanopositioning systems, including piezoelectric transducers and position sensors. These five chapters are followed by four chapters on control topics. The control chapters cover: shunt control, feedback control, force-feedback control, and feedforward control. The concluding five chapters of the monograph report issues that affect the application and implementation of the control systems designed. Consequently these chapters cover command signal design, how to compensate for hysteresis effects, the use of charge drives, the nature of noise in nanopositioning systems, and finally the electrical issues raised by the use of piezoelectric transducers.

The authorial team has worked with these systems for some years now and is able to write from a wealth of experience. Dr. Andrew J. Fleming is an Australian

Research Fellow and a Senior Lecturer at the University of Newcastle, NSW, Australia. He is a noted expert on piezoelectric applications, and with S.O. Reza Moheimani co-authored the well received *Advances in Industrial Control* monograph, *Piezoelectric Transducers for Vibration Control and Damping* (ISBN 978-1-84628-331-4, 2006). Author Dr. Kam K. Leang is an Associate Professor at the University of Nevada, Reno, USA. With a background in Mechatronics, Dr. Leang has research interests in iterative learning control and piezo-based nanopositioning systems and applications.

In the introductory chapter, there is a useful Book Summary (Sect. 1.6) that gives the reader an indication of the level of prior knowledge the authors expect the reader to have to benefit fully from the monograph. The reader, new to nanopositioning systems, will find the monograph well structured and accessible for self-learning purposes. The control chapters are very readable and involve an interesting variety of PID control and the more advanced methods. A notable feature of the monograph is the way theory is supported by experimental assessments and case studies. The industrial control engineer will find plenty of useful explanation and discussion of the physical reasons for system design and control choices. The monograph also contains reports on aspects of control design that are often glossed over in many texts. One striking example is the work and chapter on the interplay between control design and the noise present in nanopositioning systems. The breadth and thoroughness of the material presented and the way chapters are so very well focussed should make this monograph a valuable resource for lecture and short courses in the nanopositioning field. Although the text has a strong control focus, it is thought that readers outside of the control community, for example, physicists and scientists, will also find the text accessible and interesting.

In conclusion, the monograph presents a thorough and engaging exposition of the state of the art in nanopositioning and is a valuable and welcome contribution to the literature and to the *Advances in Industrial Control* series.

Glasgow, Scotland, UK

M. J. Grimble
M. A. Johnson

Contents

1 Introduction 1
1.1 Introduction to Nanotechnology 1
1.2 Introduction to Nanopositioning 2
1.3 Scanning Probe Microscopy 3
1.4 Challenges with Nanopositioning Systems 7
1.4.1 Hysteresis 7
1.4.2 Creep 7
1.4.3 Thermal Drift 8
1.4.4 Mechanical Resonance 9
1.5 Control of Nanopositioning Systems 10
1.5.1 Feedback Control 10
1.5.2 Feedforward Control 12
1.6 Book Summary 13
1.6.1 Assumed Knowledge 13
1.6.2 Content Summary 13
References 14

2 Piezoelectric Transducers 17
2.1 The Piezoelectric Effect 17
2.2 Piezoelectric Compositions 20
2.3 Manufacturing Piezoelectric Ceramics 22
2.4 Piezoelectric Transducers 23
2.5 Application Considerations 26
2.5.1 Mounting 27
2.5.2 Stroke Versus Force 27
2.5.3 Preload and Flexures 29
2.5.4 Electrical Considerations 30
2.5.5 Self-Heating Considerations 30
2.6 Response of Piezoelectric Actuators 31
2.6.1 Hysteresis 31
2.6.2 Creep 32
2.6.3 Temperature Dependence 33
2.6.4 Vibrational Dynamics 34
2.6.5 Electrical Bandwidth 35

2.7 Modeling Creep and Vibration in Piezoelectric Actuators . . . 35
2.8 Chapter Summary . . . 39
References . . . 39

3 Types of Nanopositioners . . . 43
3.1 Piezoelectric Tube Nanopositioners . . . 43
3.1.1 63 mm Piezoelectric Tube . . . 45
3.1.2 40 mm Piezoelectric Tube Nanopositioner . . . 46
3.2 Piezoelectric Stack Nanopositioners . . . 47
3.2.1 Phyisk Instrumente P-734 Nanopositioner . . . 49
3.2.2 Phyisk Instrumente P-733.3DD Nanopositioner . . . 49
3.2.3 Vertical Nanopositioners . . . 50
3.2.4 Rotational Nanopositioners . . . 51
3.2.5 Low Temperature and UHV Nanopositioners . . . 53
3.2.6 Tilting Nanopositioners . . . 53
3.2.7 Optical Objective Nanopositioners . . . 53
References . . . 55

4 Mechanical Design: Flexure-Based Nanopositioners . . . 57
4.1 Introduction . . . 57
4.2 Operating Environment . . . 58
4.3 Methods for Actuation . . . 61
4.4 Flexure Hinges . . . 62
4.4.1 Introduction . . . 62
4.4.2 Types of Flexures . . . 64
4.4.3 Flexure Hinge Compliance Equations . . . 65
4.4.4 Stiff Out-of-Plane Flexure Designs . . . 73
4.4.5 Failure Considerations . . . 74
4.4.6 Finite Element Approach for Flexure Design . . . 75
4.5 Material Considerations . . . 75
4.5.1 Materials for Flexure and Platform Design . . . 75
4.5.2 Thermal Stability of Materials . . . 77
4.6 Manufacturing Techniques . . . 78
4.7 Design Example: A High-Speed Serial-Kinematic Nanopositioner . . . 79
4.7.1 State-of-the-Art Designs . . . 79
4.7.2 Tradeoffs and Limitations in Speed . . . 81
4.7.3 Serial- Versus Parallel-Kinematic Configurations . . . 83
4.7.4 Piezoactuator Considerations . . . 84
4.7.5 Preloading Piezo-Stack Actuators . . . 85
4.7.6 Flexure Design for Lateral Positioning . . . 86
4.7.7 Design of Vertical Stage . . . 94
4.7.8 Fabrication and Assembly . . . 97

4.7.9 Drive Electronics 98
4.7.10 Experimental Results 99
4.8 Chapter Summary 100
References 101

5 Position Sensors 103
5.1 Introduction 103
5.2 Sensor Characteristics 105
5.2.1 Calibration and Nonlinearity 105
5.2.2 Drift and Stability 107
5.2.3 Bandwidth 109
5.2.4 Noise 110
5.2.5 Resolution 113
5.2.6 Combining Errors 116
5.2.7 Metrological Traceability 117
5.3 Nanometer Position Sensors 118
5.3.1 Resistive Strain Sensors 118
5.3.2 Piezoresistive Strain Sensors 121
5.3.3 Piezoelectric Strain Sensors 123
5.3.4 Capacitive Sensors 127
5.3.5 MEMs Capacitive and Thermal Sensors 133
5.3.6 Eddy-Current Sensors 134
5.3.7 Linear Variable Displacement Transformers 137
5.3.8 Laser Interferometers 140
5.3.9 Linear Encoders 144
5.4 Comparison and Summary 147
5.5 Outlook and Future Requirements 148
References 150

6 Shunt Control 155
6.1 Introduction 155
6.2 Shunt Circuit Modeling 157
6.2.1 Open-Loop 157
6.2.2 Shunt Damping 159
6.3 Implementation 164
6.4 Experimental Results 165
6.4.1 Tube Dynamics 166
6.4.2 Amplifier Performance 167
6.4.3 Shunt Damping Performance 168
6.5 Chapter Summary 173
References 173

7 Feedback Control 175
7.1 Introduction 175
7.2 Experimental Setup 178
7.3 PI Control 180
7.4 PI Control with Notch Filters 181
7.5 PI Control with IRC Damping 183
7.6 Performance Comparison 187
7.7 Noise and Resolution 188
7.8 Analog Implementation 193
7.9 Application to AFM Imaging 195
7.10 Repetitive Control 196
7.10.1 Introduction 196
7.10.2 Repetitive Control Concept and Stability Considerations 198
7.10.3 Dual-Stage Repetitive Control 201
7.10.4 Handling Hysteresis 205
7.10.5 Design and Implementation 205
7.10.6 Experimental Results and Discussion 214
7.11 Summary 216
References 216

8 Force Feedback Control 221
8.1 Introduction 221
8.2 Modeling 223
8.2.1 Actuator Dynamics 223
8.2.2 Sensor Dynamics 225
8.2.3 Sensor Noise 226
8.2.4 Mechanical Dynamics 227
8.2.5 System Properties 228
8.2.6 Example System 230
8.3 Damping Control 230
8.4 Tracking Control 232
8.4.1 Relationship Between Force and Displacement 233
8.4.2 Integral Displacement Feedback 235
8.4.3 Direct Tracking Control 235
8.4.4 Dual Sensor Feedback 237
8.4.5 Low Frequency Bypass 239
8.4.6 Feedforward Inputs 240
8.4.7 Higher-Order Modes 241
8.5 Experimental Results 241
8.5.1 Experimental Nanopositioner 241
8.5.2 Actuators and Force Sensors 242

8.5.3 Control Design 244
8.5.4 Noise Performance 245
8.6 Chapter Summary 247
References 248

9 Feedforward Control 251
9.1 Why Feedforward? 251
9.2 Modeling for Feedforward Control 252
9.3 Feedforward Control of Dynamics and Hysteresis. 252
9.3.1 Simple DC-Gain Feedforward Control. 252
9.3.2 An Inversion-Based Feedforward Approach for Linear Dynamics 253
9.3.3 Frequency-Weighted Inversion: The Optimal Inverse. 256
9.3.4 Application to AFM Imaging 256
9.4 Feedforward and Feedback Control. 258
9.4.1 Application to AFM Imaging 261
9.5 Iterative Feedforward Control. 261
9.5.1 The ILC Problem 263
9.5.2 Model-Based ILC 265
9.5.3 Nonlinear ILC 267
9.5.4 Conclusions 271
References 271

10 Command Shaping 275
10.1 Introduction 275
10.1.1 Background 275
10.1.2 The Optimal Periodic Input 279
10.2 Signal Optimization 280
10.3 Frequency Domain Cost Functions 282
10.3.1 Background: Discrete Fourier Series 282
10.3.2 Minimizing Signal Power. 283
10.3.3 Minimizing Frequency Weighted Power 284
10.3.4 Minimizing Velocity and Acceleration. 285
10.3.5 Single-Sided Frequency Domain Calculations. 286
10.4 Time Domain Cost Function 286
10.4.1 Minimum Velocity 287
10.4.2 Minimum Acceleration 288
10.4.3 Frequency Weighted Objectives 288
10.5 Application to Scan Generation 288
10.5.1 Choosing β and K. 290
10.5.2 Improving Feedback and Feedforward Controllers 292
10.6 Comparison to Other Techniques 293

10.7 Experimental Application. 295
10.8 Chapter Summary 297
References 297

11 Hysteresis Modeling and Control 299
11.1 Introduction 299
11.2 Modeling Hysteresis 300
11.2.1 Simple Polynomial Model 300
11.2.2 Maxwell Slip Model 300
11.2.3 Duhem Model. 301
11.2.4 Preisach Model 302
11.2.5 Classical Prandlt-Ishlinksii Model 306
11.3 Feedforward Hysteresis Compensation 307
11.3.1 Feedforward Control Using the Presiach Model . . . 307
11.3.2 Feedforward Control Using the Prandlt-Ishlinksii Model 309
11.4 Chapter Summary 315
References 315

12 Charge Drives 317
12.1 Introduction 317
12.2 Charge Drives 318
12.3 Application to Piezoelectric Stack Nanopositioners 322
12.4 Application to Piezoelectric Tube Nanopositioners 325
12.5 Alternative Electrode Configurations 328
12.5.1 Grounded Internal Electrode 328
12.5.2 Quartered Internal Electrode 330
12.6 Charge Versus Voltage 332
12.6.1 Advantages. 332
12.6.2 Disadvantages. 333
12.7 Impact on Closed-Loop Control 334
12.8 Chapter Summary 335
References 335

13 Noise in Nanopositioning Systems 337
13.1 Introduction 337
13.2 Review of Random Processes. 338
13.2.1 Probability Distributions 339
13.2.2 Expected Value, Moments, Variance, and RMS . . . 339
13.2.3 Gaussian Random Variables 341
13.2.4 Continuous Random Processes 343
13.2.5 Joint Density Functions and Stationarity 343
13.2.6 Correlation Functions 344
13.2.7 Gaussian Random Processes 344

13.2.8 Power Spectral Density 345
13.2.9 Filtered Random Processes 347
13.2.10 White Noise 348
13.2.11 Spectral Density in $V/\sqrt{Hz}$ 349
13.2.12 Single- and Double-Sided Spectra 349
13.3 Resolution and Noise 351
13.4 Sources of Nanopositioning Noise 352
13.4.1 Sensor Noise 353
13.4.2 External Noise 354
13.4.3 Amplifier Noise 354
13.5 Closed-Loop Position Noise 359
13.5.1 Noise Sensitivity Functions 359
13.5.2 Closed-Loop Position Noise Spectral Density 360
13.5.3 Closed-Loop Noise Approximations with Integral Control 361
13.5.4 Closed-Loop Position Noise Variance 362
13.5.5 A Note on Units 364
13.6 Simulation Examples 364
13.6.1 Integral Controller Noise Simulation 364
13.6.2 Noise Simulation with Inverse Model Controller . . . 366
13.6.3 Feedback Versus Feedforward Control 369
13.7 Practical Frequency Domain Noise Measurements 370
13.7.1 Preamplification 370
13.7.2 Spectrum Estimation 372
13.7.3 Direct Measurement of Position Noise 373
13.7.4 Measurement of the External Disturbance 375
13.8 Experimental Demonstration 375
13.9 Time-Domain Noise Measurements 379
13.9.1 Total Integrated Noise 379
13.9.2 Estimating the Position Noise 381
13.9.3 Practical Considerations 383
13.9.4 Experimental Demonstration 384
13.10 A Simple Method for Measuring the Resolution of Nanopositioning Systems 386
13.11 Techniques for Improving Resolution 388
13.12 Chapter Summary 390
References 391

14 Electrical Considerations 395
14.1 Introduction 395
14.2 Bandwidth Limitations 396
14.2.1 Passive Bandwidth Limitations 396
14.2.2 Amplifier Bandwidth 398
14.2.3 Current and Power Limitations 398

14.3 Dual-Amplifier 399
14.3.1 Circuit Operation 399
14.3.2 Range Considerations 401
14.4 Electrical Design 402
14.4.1 High-Voltage Stage 402
14.4.2 Low-Voltage Stage 404
14.4.3 Cabling and Interconnects 405
14.5 Chapter Summary 407
References 407

Index 409

Chapter 1
Introduction

This chapter provides an introduction to the design, applications, and characteristics of piezoelectric nanopositioning systems. Particular attention is paid to the characteristics that limit speed and resolution. The performance limitations are then discussed followed by an overview of control techniques to improve performance.

1.1 Introduction to Nanotechnology

On December 29, 1959, physicist Richard Feynman gave a talk entitled "There's Plenty of Room at the Bottom" at an American Physical Society meeting at the California Institute of Technology (CalTech). Feyman's talk sparked interest in ideas and concepts behind nanoscience and nanotechnology. In his talk, Feynman described a process in which scientists would be able to manipulate and control individual atoms and molecules. Over a decade later, the term nanotechnology was coined by Professor Norio Taniguchi through his work on ultraprecision machining. Modern nanotechnology began in 1981 with the development of the scanning tunneling microscope (STM) (Binnig et al. 1982), a type of scanning probe microscope. The STM gave scientists the ability to "see" individual atoms.

The National Nanotechnology Initiative (NNI) defines nanotechnology as the manipulation of matter at the nanoscale, or more specifically at least one dimension sized from 1 to 100 nm (http://www.nano.gov/). One nanometer is one billionth of a meter, and on a comparative scale, if a marble were a nanometer, then one meter would be the size of Earth. Research and development in nanotechnology encompasses many fields, such as surface science, organic chemistry, molecular biology, semiconductor physics, and microfabrication. In general, nanotechnology involves imaging, measuring, modeling, and manipulating matter at this length scale.

What attracts scientists and engineers to work at the nanoscale is matter such as gases, liquids, and solids can exhibit unusual physical, chemical, and biological properties at the nanoscale. Therefore, scientists and engineers can develop

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_1,

novel nanostructured materials that are stronger or have different physical properties compared to other forms or sizes of the same material. For example, some materials can be developed that are better at conducting heat or electricity, or become more chemically reactive or reflect light better or change color as their size or structure is altered. Other applications of nanotechnology are equally diverse, ranging from extensions of conventional device physics to completely new approaches based upon molecular self-assembly, from developing new materials with dimensions on the nanoscale to direct control of matter on the atomic scale. Over the last several decades, billions of dollars have been invested in nanotechnology because of the variety of potential industrial and military applications.

Scanning probe microscopes such as the STM and the atomic force microscope (AFM) were invented in the 1980s to allow scientists to see and manipulate matter at the nanoscale (see Sect. 1.3 for detailed discussion). For example, the AFM uses a small microfabricated cantilever with a sharp tip (probe) located at its distal end to interact with and "feel" the sample surface (Binnig and Quate 1986; Leang et al. 2009). The tool can obtain high-resolution topographical images, and it also has the ability to directly measure various properties of a specimen. For example, the structural and mechanical properties of biological specimens such as cells and DNA have been investigated by the tool.

In addition to imaging and investigating the surface of a sample at the nanoscale, scanning probe-based tools can be exploited for manufacturing at the nanoscale, a process also known as nanomanufacturing. Nanomanufacturing involves scaled-up, reliable, and cost-effective manufacturing of nanoscale materials, structures, devices, and systems. Nanomanufacturing also includes research, development, and integration of top-down processes and increasingly complex bottom-up or self-assembly processes. Some techniques to create nanosize features and devices include photolithography, nanoimprint, self-assembly, and the use of probe-based tools to physically shape or modify the surface of a sample.

One critical tool in nanotechnology is the nanopositioning system. Nanopositioning systems are used extensively in scanning probe microscopy and in applications that require subnanometer precision motion control. Nanopositioning systems are introduced next.

1.2 Introduction to Nanopositioning

Nanopositioning stages are mechanical positioning devices capable of developing displacements with nanometer scale resolution. A simple nanopositioning stage is illustrated in Fig. 1.1. The moving platform is centrally suspended by four leaf flexures. These flexures are designed to *flex* and deflect freely in the direction of travel but resist motion in other directions. Their purpose is to guide the motion of the platform and to provide a preloading force on the actuator.

Most nanopositioning systems employ piezoelectric stack actuators for developing force and displacement. The actuators elongate by around 0.1 % when the

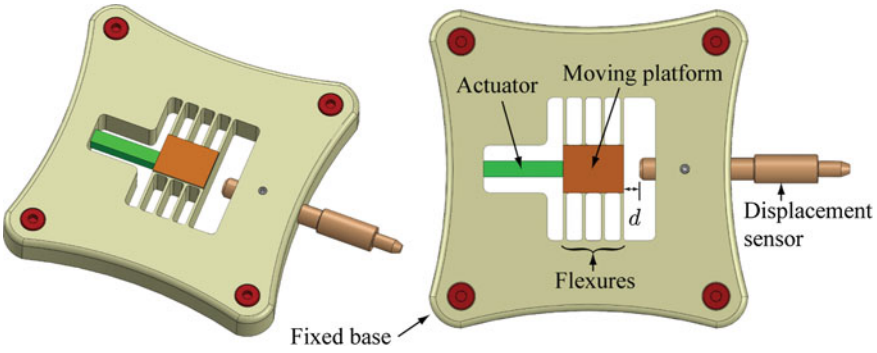


Fig. 1.1 A single degree-of-freedom positioning stage. The actuator expansion causes the platform to displace laterally. The resulting displacement d is measured by the position sensor

maximum voltage of between 60 and 200 V is applied. In Fig. 1.1, the actuator drives the moving platform through a flexure that permits only lateral deflection. This is necessary to avoid transmitting any bending or torsional forces that may be produced by the actuator. Further information on piezoelectric actuators can be found in Chap. 2.

Sources of positioning error in a nanopositioning stage include actuator nonlinearity and creep, structural vibration, and thermal drift. To eliminate these errors, a position sensor is incorporated into the stage and used within a feedback control loop to regulate the position. Figure 1.1 illustrates a position sensor that directly measures the position of the moving platform relative to the frame. The feedback controller works to equate the measured position to the command reference, thereby eliminating errors due to actuator nonlinearity, thermal drift, and other sources of disturbance.

Nanopositioning systems come in a variety of forms and are widely applied in a diverse range of scientific and industrial applications. Some examples include: fiber aligners (Wang et al. 2007), beam scanners (Potsaid et al. 2007), and lateral positioning platforms (Devasia et al. 2007). Among other applications in nanotechnology (Bhushan 2004), nanopositioning platforms are used widely in scanning probe microscopy (Salapaka and Salapaka 2008; Abramovitch et al. 2007; Meyer et al. 2004) and nanofabrication systems (Tseng et al. 2005, 2008; Tseng 2008). Examples of some commercial nanopositioning stages are pictured in Fig. 1.2. Other examples are described in Chap. 3.

1.3 Scanning Probe Microscopy

A common application of nanopositioners is in the lateral and vertical positioning stages of scanning probe microscopes (SPMs) such as the AFM. Unlike a traditional optical microscope that uses light for imaging, an AFM image is formed by scanning a microcantilever probe over the surface, as illustrated in Fig. 1.3. The AFM is one

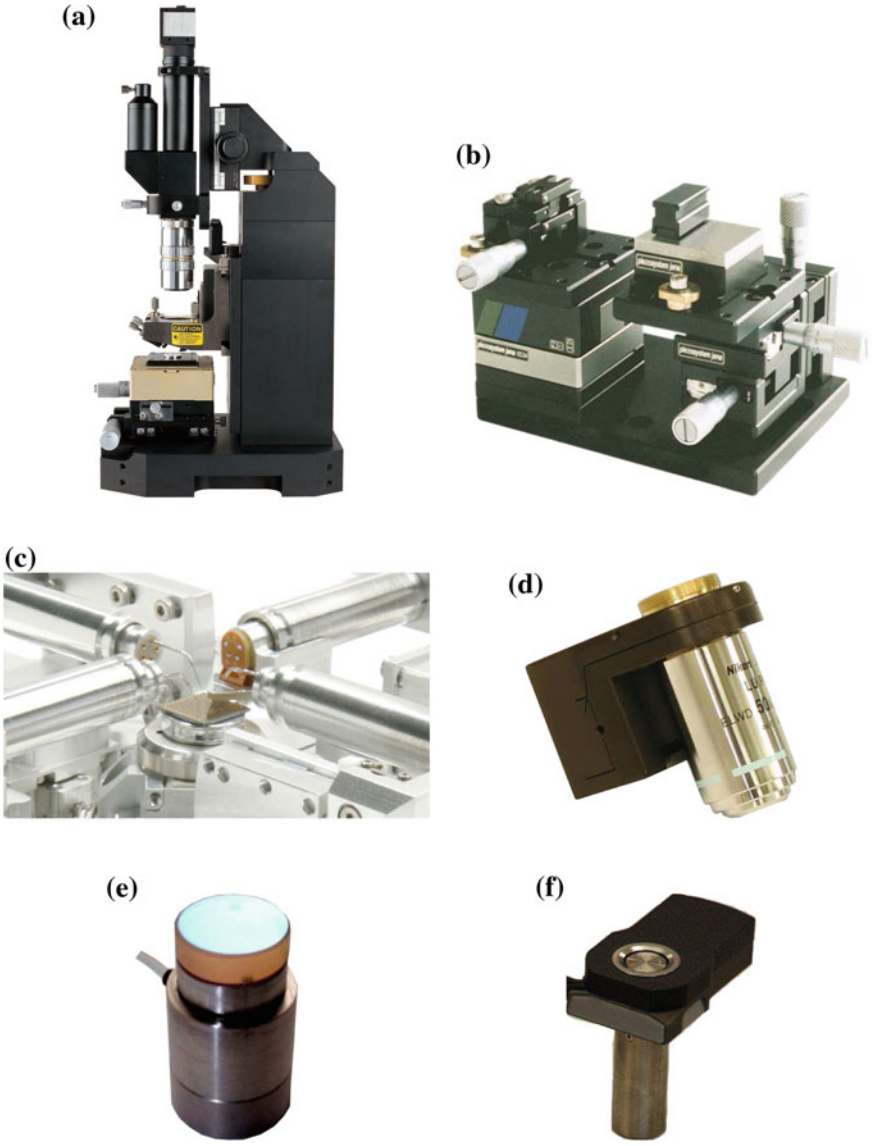


Fig. 1.2 Examples of commercial nanopositioning systems. **a** Park Systems Corp. (Korea) atomic force microscope with 2-axis sample nanopositioner. **b** Piezosystem Jena GmbH (Germany) fiber alignment system with 3-axis nanopositioner. **c** Zyvex Instruments (USA) probe station with four piezoelectric tube nanopositioners. **d** Madcity Labs Inc. (USA) microscope objective nanopositioner. **e** Queensgate Instruments Ltd. (UK) mirror tilting stage. **f** NT-MDT Co. (Russia) piezoelectric tube nanopositioner for scanning probe microscopy

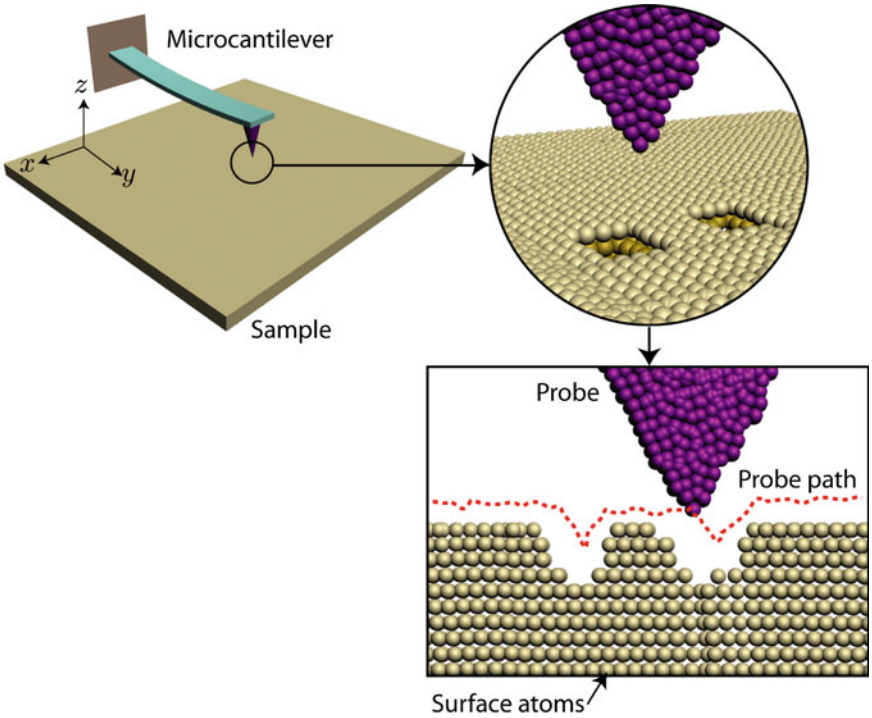


Fig. 1.3 The operation of an atomic force microscope

of the most versatile microscopes due to its ability to work with conducting and nonconducting samples in a vacuum, air, or in water (Binnig and Quate 1986). The probe is a micro-machined cantilever with a sharp tip protruding toward the sample surface. When the probe is brought into contact with the surface, the tip-to-sample interaction causes the cantilever to deflect vertically. This deflection is measured and used to construct an image of the sample. The AFM essentially "feels" the surface with a tiny, finger-like cantilever. In a vacuum, resolution of an AFM is on the order of 0.01 nm. With such high resolution, an AFM can generate topographical images of atoms, as well as to control, manipulate, and alter the properties of matter at the nanoscale (Salapaka and Salapaka 2008).

The positioning of the probe tip relative to the sample can be achieved with two basic configurations: (a) scan-by-sample or (b) scan-by-probe as shown in Fig. 1.4. In the scan-by-sample configuration, the nanopositioner, flexure-based design shown equipped with three piezo stacks, moves the sample relative to a fixed probe. The x and y axis piezos position the sample along the lateral direction (parallel to the sample surface) and a z axis stack moves the sample vertically. The deflection of the cantilever is measured optically, by reflecting a laser beam off the end of the cantilever onto a nearby photodetector. Alternatively, in the scan-by-probe arrangement shown in Fig. 1.4b, a nanopositioner moves the probe relative to a fixed sample both laterally

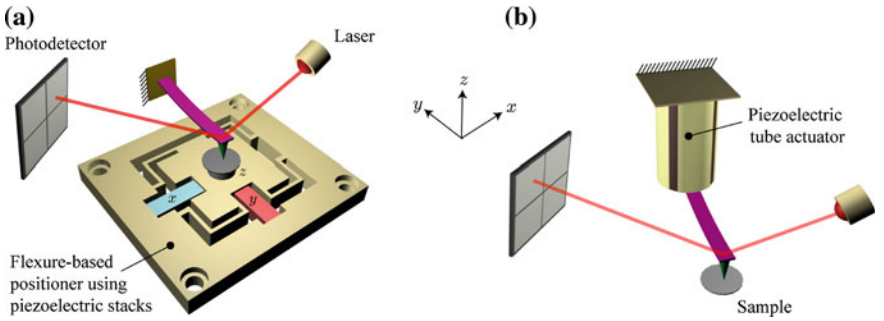


Fig. 1.4 Two positioning schemes for SPMs: **a** scan-by-sample and **b** scan-by-probe

and vertically. In scan-by-probe systems, the laser and photodector are required to move with the cantilever; however, this can be avoided by incorporating sensing elements into the cantilever itself, such as using piezoresistive, piezoelectric, or capacitive elements.

There are three basic operating modes of an AFM: contact, noncontact, and tapping mode. In contact mode, the probe interacts with the sample at very close range where the dominant force on the tip is repulsive. In this mode, the deflection of the AFM cantilever is sensed and a feedback controller is used to maintain a desired deflection. The spring constant of a contact mode AFM cantilever varies between 0.001 and 10 N/m.

Soft samples such as living cells have a contact stiffness comparable to, or less than, the cantilever stiffness, therefore, they may be deformed or damaged during contact mode operation. Noncontact mode avoids direct sample contact by exploiting attractive Van der Waals forces. In this mode, the AFM tip is hovered above the surface (at approximately 50–150 Å). By oscillating the tip, the effective stiffness of the microcantilever is effected by the force gradient of the attractive forces. The effective stiffness can be related to the sample topography by measuring or regulating the amplitude, phase, or resonance frequency of the probe. In general, noncontact mode AFM provides lower resolution than contact mode but does not pollute or damage the sample. Noncontact mode can also be used to measure long range forces such as magnetic or electric fields in samples such as hard disk media or charged insulators.

For high-resolution imaging of soft samples such as living cells, polymers, and gells, tapping mode AFM is the preferred method. In this mode, the AFM cantilever is oscillated near its resonance frequency (50 kHz–1 MHz) using a piezoelectric actuator. As the AFM tip is brought into contact with the surface, the tip lightly touches or taps the surface. When the cantilever intermittently contacts the surface, the oscillating behavior is altered by the energy loss during the tip-to-sample interaction. The change in energy is monitored and used to construct an image of the surface.

Precision positioning is needed in many AFM applications. In particular, precise position control in both the lateral and vertical directions is needed to hold the probe

at a desired location or to track a desired motion trajectory. For instance, when the AFM is used to create quantum dots (2–80 nm in size), accurate position control of the indenter tip is needed as the probe position directly affects the size, spacing, and distribution of the nanofeatures. Even 2–4 nm variation in size and spacing of the nanofeatures can drastically alter their properties (Leonard et al. 1993). Additionally, high-speed control of the probe's movement is needed for high throughput fabrication, imaging, and metrology. Without accurate motion control along a specific trajectory at high speed, oscillations can cause the tip to collide with nearby features, leading to excessive tip-to-sample forces and imaging artifacts. Large forces can damage the probe tip or soft specimens such as cells. Thus, accurate position control is critical in an AFM.

1.4 Challenges with Nanopositioning Systems

Due to their effectively infinite resolution, piezoelectric actuators are universally employed in nanopositioning applications. However, the positioning accuracy of piezoelectric actuators is limited by hysteresis over large displacements, creep, and thermal drift which is present at low-frequencies. Another major problem with nanopositioning systems is the presence of lightly damped mechanical resonances. These dynamics can result in large oscillations, particularly when step-changes or high frequency inputs are involved. The impact of these detrimental phenomena are discussed below.

1.4.1 Hysteresis

When employed in an actuating role, piezoelectric transducers display a significant hysteresis in the transfer function from the applied voltage to the resulting strain or displacement (Adriaens et al. 2000). A typical hysteresis response is plotted in Fig. 1.5. In dynamic applications, hysteresis is considered the foremost limitation to performance. It leads to poor positioning accuracy, poor repeatability, and mixing of harmonic content into the displacement response.

1.4.2 Creep

When a piezoelectric transducer is commanded by a step change in voltage, the response speed is limited only by the mechanical resonance of the host structure or transducer. Creep, illustrated in Fig. 1.6, is the phenomenon where actuator deflection slowly "creeps" upward after an increase in applied voltage. The time constant is typically a few minutes. Creep severely degrades the low-frequency and static positioning ability of piezoelectric actuators.

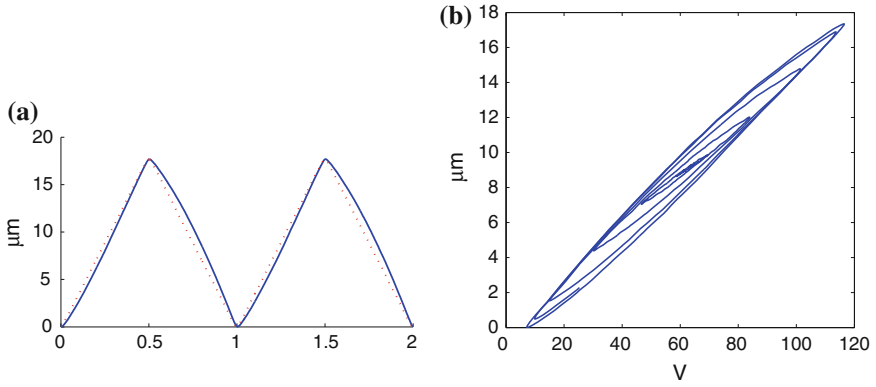


Fig. 1.5 Displacement hysteresis exhibited by the P-733 nanopositioner described in Sect. 3.2.2. **a** *Triangle input* The displacement in μm is plotted against time in seconds. **b** XY plot of displacement versus spplied voltage with an increasing amplitude sine wave input

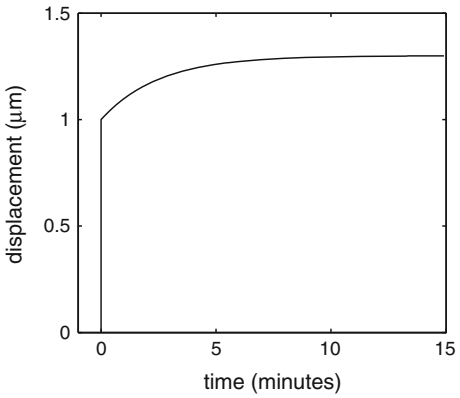


Fig. 1.6 An example of piezoelectric creep. The response to a step change in voltage is plotted over a period of 15 min

1.4.3 Thermal Drift

The properties of piezoelectric materials are highly temperature dependent. Figure 1.7 shows a 20 % increase in displacement sensitivity over a range of 50 °C. In the worst case, this would result in a drift of 0.4 % of the full range per degree of temperature drift. This is vastly more significant than the drift due to mechanical thermal drift. Such temperature dependence limits the use of piezoelectric transducers as calibrated force or displacement actuators.

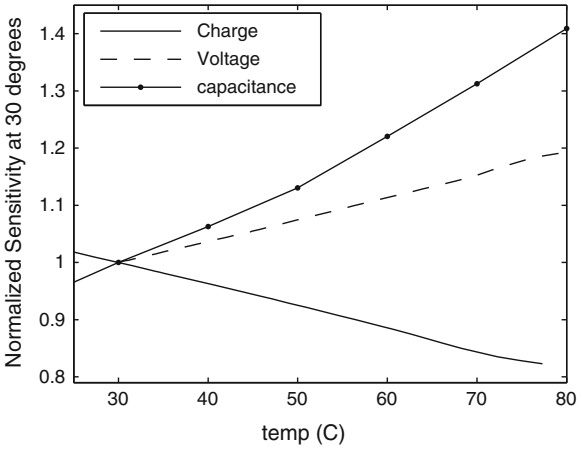


Fig. 1.7 The normalized displacement sensitivity to temperature variation of the piezoelectric tube described in Sect. 3.1.1

1.4.4 Mechanical Resonance

The greatest speed limitation of a nanopositioner is the mechanical resonances that arises from the platform mass interacting with the stiffness of the support flexures, mechanical linkages, and actuators. Since the lowest resonance frequency is typically of greatest interest, the dynamics of a nanopositioner may be approximated by a unity-gain second-order low-pass system

$$G(s) = \frac{\omega_r^2}{s + 2\omega_r \zeta s + \omega_r^2}, \tag{1.1}$$

where ω_r and ζ are the resonance frequency and damping ratio. The magnitude and phase responses of this system are plotted in Fig. 1.9. To avoid excitation of the mechanical resonance, the frequency of driving signals is limited to around 1–10 % of the resonance frequency. In applications where scan frequency is the foremost performance limitation, for example in high-speed atomic force microscopy (Ando et al. 2005; Schitter et al. 2007; Humphris 2005; Rost et al. 2005), the nanopositioner is operated in open-loop with driving signals that are shaped to reduce harmonic content. Although such techniques, reviewed in (Fleming and Wills 2008), can provide a fast response, they are not accurate as nonlinearity and disturbance remain uncontrolled.

The transient response of a nanopositioning stage can be vastly improved by actively damping the first resonance mode. This can reduce the settling time by greater than 90 % and allow a proportional increase in the scan speed. Systems with active damping also facilitate greater tracking performance as the controller gain can be significantly increased, as discussed in the following subsection.

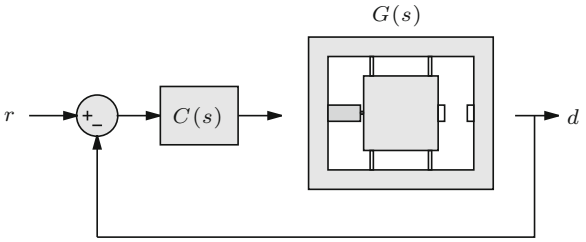


Fig. 1.8 A nanopositioner G in a displacement feedback control loop

1.5 Control of Nanopositioning Systems

1.5.1 Feedback Control

The most popular technique for control of commercial nanopositioning systems is sensor-based feedback control (Fig. 1.8) using integral or proportional-integral control (P Instruments 2009). Such controllers are simple, robust to modeling error, and effectively reduce piezoelectric nonlinearity at low-frequencies. However, the bandwidth of integral tracking controllers is severely limited by the presence of highly resonant modes. The cause of such limited closed-loop bandwidth can be explained by examining the loop gain $|CG|$ in Fig. 1.9. Here, the resonant system G is controlled by an integral controller C with gain α. The factor limiting the maximum feedback gain and closed-loop bandwidth is gain margin.

At the resonance frequency ω_r the phase lag exceeds π so the loop gain must be less than 1 or 0 dB for stability in closed-loop. The condition for closed-loop stability is

$$\frac{\alpha}{\omega_r} \times \frac{1}{2\zeta} < 1, \text{ or } \alpha < 2\omega_r\zeta. \tag{1.2}$$

As the system G is unity gain, the feedback gain α is also the closed-loop bandwidth ω_{cl} (in radians per second). Thus, the maximum closed-loop bandwidth is proportional to the product of damping ratio ζ and resonance frequency ω_r, that is,

$$\text{max. closed-loop bandwidth} < 2\omega_r\zeta. \tag{1.3}$$

This is a severe limitation as the damping ratio is typically on the order of 0.01, so the maximum closed-loop bandwidth is less than 1 % of the resonance frequency. The maximum closed-loop bandwidth can also be estimated directly from the frequency response by replacing the factor 2ζ with $1/P$, where P is the linear magnitude of the resonance peak divided by the DC gain, that is

$$\text{max. closed-loop bandwidth} < \frac{\omega_r}{P}, \tag{1.4}$$

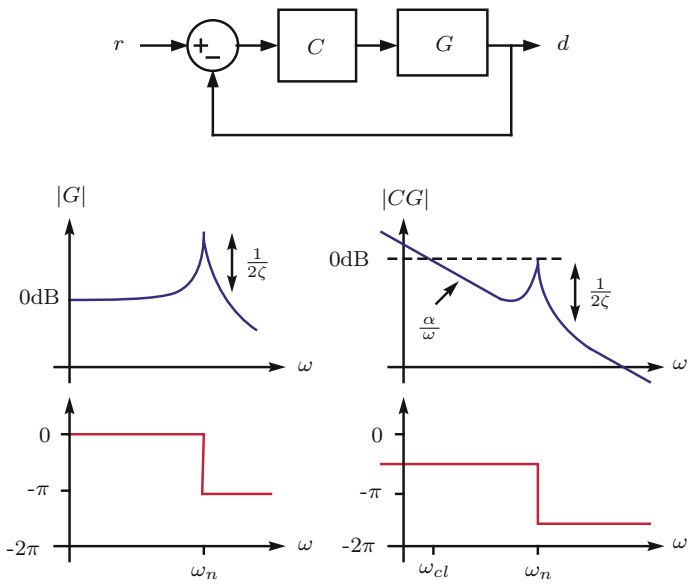


Fig. 1.9 A nanopositioning system G controlled by an integral controller $C = \alpha/s$. The frequency response of G and the system loop gain CG are plotted on the *left-hand side* and *right-hand side*, respectively

Techniques aimed at improving the closed-loop bandwidth are based on either inverting the resonance using a notch filter (Abramovitch et al. 2008) or damping the resonance with a damping controller (Aphale et al. 2008). Other feedback-based approaches include state-feedback (Okazaki 1990), gain scheduling (Merry et al. 2009), robust control (Korson and Helmicki 1995; Salapaka et al. 2002), and repetitive control (Aridogan et al. 2009; Merry et al. 2011; Shan and Leang 2012).

Inversion techniques are popular as they are simple to implement and can provide excellent closed-loop bandwidth, up to or greater than the resonance frequency (Abramovitch et al. 2008). The major disadvantage of inversion-based techniques is the requirement for an accurate system model. If the system resonance frequency shifts by only 1 %, a high-gain inversion-based feedback controller can become unstable. In most applications this is unacceptable as the load mass and resonance frequency can vary significantly during service. As a result of this sensitivity, high-performance inversion-based controllers are applied in applications where the resonance frequency is stable, or when the feedback controller can be continually recalibrated (Abramovitch et al. 2008).

Damping control is an alternative method for reducing the bandwidth limitations imposed by mechanical resonance. Damping control uses a feedback loop to artificially increase the damping ratio ζ of a system. Due to Eq. (1.2), an increase in ζ allows a proportional increase in the feedback gain and closed-loop bandwidth. Although damping controllers alone cannot increase the closed-loop bandwidth to

beyond the resonance frequency, they have the advantage of being insensitive to variations in resonance frequency. In addition, as damping controllers suppress, rather than invert, the mechanical resonance, they provide better rejection of external disturbances than inversion-based systems (Devasia et al. 2007).

A number of techniques for damping control have been demonstrated successfully in the literature, these include Positive Position Feedback (PPF) (Fanson and Caughey 1990), polynomial-based control (Aphale et al. 2008), shunt control (Fleming and Moheimani 2006), resonant control (Sebastian et al. 2008) and Integral Resonance Control (IRC) (Aphale et al. 2007, 2008). These techniques can successfully damp a system resonance with modest insensitivity to variations in resonance frequency. However, like all feedback control systems, the tracking controller gain is still limited by stability margins and the positioning resolution is still dominated by sensor-induced noise.

To demonstrate the limitations imposed by sensor noise, consider a nanopositioner with feedback control derived from a high performance capacitive sensor with a range of $\pm 100\ \mu$m and root-mean-square (RMS) noise of $20\,\text{pm}/\sqrt{\text{Hz}}$. An estimate of the RMS positioning noise can be found by multiplying noise density by the square-root of closed-loop bandwidth. i.e.,

$$\text{RMS Noise} = \sqrt{\text{Bandwidth}} \times \text{Noise Density}. \tag{1.5}$$

For example, with a closed-loop bandwidth of 100 Hz, the positioning noise is 0.2 nm RMS or approximately 1.2 nm peak-to-peak (if the noise is normally distributed). For atomic resolution, the closed-loop bandwidth must be reduced to below 1 Hz, which is a severe limitation.

1.5.2 Feedforward Control

Feedforward or inversion-based control is commonly applied to both open- and closed-loop nanopositioning systems that require improved performance (Devasia et al. 2007; Butterworth et al. 2008). Good reference tracking can be achieved if the plant model or its frequency response are known with high accuracy. In addition to improved performance, other attractive characteristics of inversion-based control are the lack of additive sensor noise and the ease of implementation, particularly in high-speed applications (Schitter and Stemmer 2004).

The foremost difficulty with inversion-based control is the lack of robustness to variations in plant dynamics, especially if the system is resonant (Devasia 2002; Butterworth et al. 2008). However, this problem only exists with static feedforward controllers. More recently, iterative techniques have been reported that eliminate both vibration and nonlinearity in systems with periodic inputs (Wu and Zou 2007). Although such techniques originally required a reference model (Wu and Zou 2007), in 2008, both Kim and Zou (2008), Li and Bechhoefer (2008) presented techniques that operate without any prior system knowledge. Both techniques

achieve essentially perfect tracking of periodic references regardless of nonlinearity or dynamics. A feedback-based repetitive controller has been designed for tracking periodic reference trajectories (Aridogan et al. 2009; Shan and Leang 2012, 2013). Unfortunately iterative feedforward and repetitive control approaches are restricted to applications with periodic references. A digital signal processor is also required.

1.6 Book Summary

This book aims to provide a practical introduction to the design and control of nanopositioning systems. It includes introductory content for the beginner and more advanced topics for achieving the maximum performance from piezoelectric nanopositioning systems.

1.6.1 Assumed Knowledge

Approximately half of the content in this book is introductory and will suit readers from diverse backgrounds in physics, electrical engineering, and mechanical engineering. The more advanced concepts such as hysteresis inversion and command shaping are targeted at control engineers aiming to achieve maximum performance from nanopositioning systems; however, an introduction to these concepts is also provided for those without a background in control theory.

It is assumed that the reader is familiar with basic linear systems and control theory, for example: transfer functions, state space systems, frequency response analysis, transient response analysis, and stability. The chapters on Hysteresis and Command Shaping will also require a working knowledge of linear algebra and optimal control theory. An understanding of electronics and circuit theory is required for the chapters on Shunt Control, Charge Drives, and Electrical Considerations. The chapter on Mechanical Design assumes basic knowledge of solid mechanics including: stress, strain, bending moments, etc.

1.6.2 Content Summary

For a newcomer to the field of piezoelectric nanopositioning, the chapters are designed by be read in order. The concepts of piezoelectricity, nanopositioning, and mechanics are introduced in Chaps. 2, 3 and 4. These chapters are followed by an introduction to position sensor technology in Chap. 5 and basic control techniques in Chap. 7.

The advanced topics begin with Shunt Control in Chap. 6 and Force Feedback control in Chap. 8. Both of these methods improve the controllability of a nanopositioner by reducing or eliminating the mechanical resonances. The servo bandwidth can also

be improved by the Feedforward and Command Shaping techniques described in Chaps. 9 and 10, respectively.

The modeling and inversion of piezoelectric hysteresis is considered in Chap. 11. This is followed by an introduction to charge amplifiers in Chap. 12 which can be an effective way of reducing hysteresis in dynamic applications. This book concludes with a detailed analysis of positioning noise in Chap. 13 and an introduction to the electrical limitations in Chap. 14.

References

Abramovitch DY, Andersson SB, Pao LY, Schitter G (2007) A tutorial on the mechanisms, dynamics, and control of atomic force microscopes. In: Proceeding of American control conference, New York, pp 3488–3502

Abramovitch DY, Hoen S, Workman R (2008) Semi-automatic tuning of PID gains for atomic force microscopes. In: American control conference, Seattle, pp 2684–2689

Adriaens HJMTA, de Koning WL, Banning R (2000) Modeling piezoelectric actuators. IEEE/ASME Trans Mechatron 5(4):331–341

Ando T, Kodera N, Uchihashi T, Miyagi A, Nakakita R, Yamashita H, Matada K (2005) High-speed atomic force microscopy for capturing dynamic behavior of protein molecules at work. e-J Surf Sci Nanotechnol 3:384–392

Aphale SS, Bhikkaji B, Moheimani SOR (2008) Minimizing scanning errors in piezoelectric stack-actuated nanopositioning platforms. IEEE Trans Nanotechnol 7(1):79–90

Aphale SS, Fleming AJ, Moheimani SOR (2007) Integral control of resonant systems with collocated sensor-actuator pairs. IOP Smart Mater Struct 16:439–446

Aphale SS, Fleming AJ, Moheimani SOR (2008) A second-order controller for resonance damping and tracking control of nanopositioning systems. In: Proceeding of 19th international conference on adaptive structures and technologies, Ascona

Aridogan U, Shan Y, Leang KK (2009) Design and analysis of discrete-time repetitive control for scanning probe microscopes. ASME J Dyn Syst Meas Control 131:061103 (12 p)

Bhushan B (ed) (2004) The handbook of nanotechnology. Springe, Berlin

Binnig G, Quate CF (1986) Atomic force microscope. Phys. Rev. Lett. 56(9):930–933

Binnig G, Rohrer H, Gerber C, Weibel E (1982) Tunnelling through a controllable vacuum gap. Appl Phys Lett 40(2):178-180

Butterworth JA, Pao LY, Abramovitch DY (2008) A comparison of control architectures for atomic force microscopes. Asian J Control (Submitted)

Devasia S (2002) Should model-based inverse inputs be used as feedforward under plant uncertainty? IEEE Trans Autom Control 47(11):1865–1871

Devasia S, Eleftheriou E, Moheimani SOR (2007) A survey of control issues in nanopositioning. IEEE Trans Control Syst Technol 15(5):802–823

Fanson JL, Caughey TK (1990) Positive position feedback control for large space structures. AIAA J 28(4):717–724

Fleming AJ, Moheimani SOR (2006) Sensorless vibration suppression and scan compensation for piezoelectric tube nanopositioners. IEEE Trans Control Syst Technol 14(1):33–44

Fleming AJ, Wills AG (2008) Optimal input signals for bandlimited scanning systems. In: Proceeding of IFAC World Congress, Seoul, pp 11 805–11 810

Humphris ADL, Miles MJ, Hobbs JK (2005) A mechanical microscope: high-speed atomic force microscopy. Appl Phys Lett 86:034 106-1–034 106-3

Kim K, Zou Q (2008) Model-less inversion-based iterative control for output tracking: piezo actuator example. In: American Control Conference, Seattle, pp 2710–2715

Korson S, Helmicki AJ (1995) An h_∞ based controller for a gas turbine clearance control system. In Proceeding IEEE conference on control applications, pp 1154–1159

Leang KK, Zou Q, Devasia S (2009) Feedforward control of piezoactuators in atomic force microscope systems. Control Syst Mag 29(1):70–82

Leonard D, Krishnamurthy M, Reaves CM, Denbaars SP, Petroff PM (1993) Direct formation of quantum-sized dots from uniform coherent islands of ingaas on gaas surfaces. Appl Phys Lett 63(23):3203–3205

Li Y, Bechhoefer J (2008) Feedforward control of a piezoelectric flexure stage for AFM. In: American Control Conference, Seattle, pp 2703–2709

Merry RJE, de Kleijn NCT, van de Molengraft MJG, Steinbuch M (2009) Using a walking piezolegs actuator to drive and control a high precision stage. IEEE/ASME Trans Mechatron 14:21–31

Merry RJE, Ronde MJC, van de Molengraft R, Koops KR, Steinbuch M (2011) Directional repetitive control of a metrological afm. IEEE Trans Control Sys Tech 19(6):1622–1629

Meyer E, Hug HJ, Bennewitz R (2004) Scanning probe microscopy. The lab on a tip. Springer, Heidelberg

Okazaki Y (1990) A micro-positioning tool post using a piezoelectric actuator for diamond turning machines. Precis Eng 12(3):151–156

P Instruments (2009) Piezo nano positionoing: inspirations

Potsaid B, Wen JT, Unrath M, Watt D, Alpay M (2007) High performance motion tracking control for electronic manufacturing. ASME J Dyn Syst Meas Control 129:767–776 (mirror Scanner)

Rost MJ, Crama L, Schakel P, van Tol E, van Velzen-Williams GBEM, Overgauw CF, ter Horst H, Dekker H, Okhuijsen B, Seynen M, Vijftigschild A, Han P, Katan AJ, Schoots K, Schumm R, van Loo W, Oosterkamp TH, Frenken JWM (2005) Scanning probe microscopes go video rate and beyond. Rev Sci Instrum 76(5):053 710-1–053 710-9

Salapaka SM, Salapaka MV (2008) Scanning probe microscopy. IEEE Control Syst Mag 28(2): 65–83

Salapaka S, Sebastin A, Cleveland JP, Salapaka MV (2002) High bandwidth nano-positioner: a robust control approach. Rev Sci Instr 73(9):3232–3241

Schitter G, Åström KJ, DeMartini BE, Thurner PJ, Turner KL, Hansma PK (2007) Design and modeling of a high-speed afm-scanner. IEEE Trans Control Syst Technol 15(5):906–915

Schitter G, Stemmer A (2004) Identification and open-loop tracking control of a piezoelectric tube scanner for high-speed scanning-probe microscopy. IEEE Trans Control Syst Technol 12(3): 449–454

Sebastian A, Pantazi A, Moheimani SOR, Pozidis H, Eleftheriou E (2008) A self servo writing scheme for a MEMS storage device with sub-nanometer precision. In: Proceeding of IFAC World Congress, Seoul, pp 9241–9247

Shan Y, Leang KK (2012) Dual-stage repetitive control with prandtl-ishlinskii hysteresis inversion for piezo-based nanopositioning. Mechatronics 22:271–281

Shan Y, Leang KK (2013) Mechanical design and control for high-speed nanopositioning: serial-kinematic nanopositioners and repetitive control for nanofabrication. IEEE Control Syst Mag (special issue on dynamics and control of micro and naoscale systems) 33(6):86–105

Tseng AA (ed) (2008) Nanofabrication: fundamentals and applications. World Scientific, Singapore

Tseng AA, Jou S, Notargiacomo A, Chen TP (2008) Recent developments in tip-based nanofabrication and its roadmap. J Nanosci Nanotechnol 8(5):2167–2186

Tseng AA, Notargiacomob A, Chen TP (2005) Nanofabrication by scanning probe microscope lithography: a review. J Vac Sci Technol 23(3):877–894

Wang Z, Chen L, Sun L (2007) An integrated parallel micromanipulator with flexure hinges for optical fiber alignment. In: Proceeding of IEEE international conference on mechatronics and automation, Harbin, pp 2530–2534

Wu Y, Zou Q (2007) Iterative control approach to compensate for both the hysteresis and the dynamics effects of piezo actuators. IEEE Trans Control Syst Technol 15(5):936–944

Chapter 2
Piezoelectric Transducers

Due to their high stiffness, compact dimensions, and extremely high positioning resolution, piezoelectric actuators are used exclusively in the vast majority of nanopositioning systems. This chapter introduces piezoelectric actuators and describes their electromechanical properties, with a focus on those that are relevant in nanopositioning applications.

2.1 The Piezoelectric Effect

In 1784, Charles Coulomb conjectured that electricity might be produced by pressure (Ballato 1996); however, no conclusive experiments were performed to validate the claim, until 1880, when Pierre and Jacques Curie[1] discovered that certain crystals (such as quartz, sodium chlorate, boracite, cane sugar, and Rochelle salt) when subjected to mechanical stress produce electric charge (see timeline in Fig. 2.1). One year later, the French physicist Lippmann predicted, based on thermodynamic analysis, the converse effect: strain as a result of an applied voltage. That same year the Curie brothers verified Lippmann's prediction (Mason 1946). Subsequently, the discovery was named the *piezoelectric* effect from the Greek word *piezein*, meaning to press or squeeze, and the Curie Brothers were credited with the discovery (Ballato 1996; Cady 1946).

The piezoelectric effect is illustrated by the simple two-dimensional model for quartz shown in Fig. 2.2. Lord Kelvin conceived the model in 1893 to explain the piezoelectric effect (Ballato 1996; Mason 1946). The piezoelectric effect is based on the unique characteristic of certain crystalline lattices to deform under pressure, and as a result, the centers of gravity of the positive and negative charges separate, creating a dipole moment (product of charge value and their separation). The resulting

[1] Pierre Curie was born in 1859 and died of an accident with a horse carriage in 1906. Jacques was born in 1855 and lived till 1941. The discovery of the piezoelectric effect was made in Jacques' laboratory (Mason 1981).

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_2, © Springer International Publishing Switzerland 2014

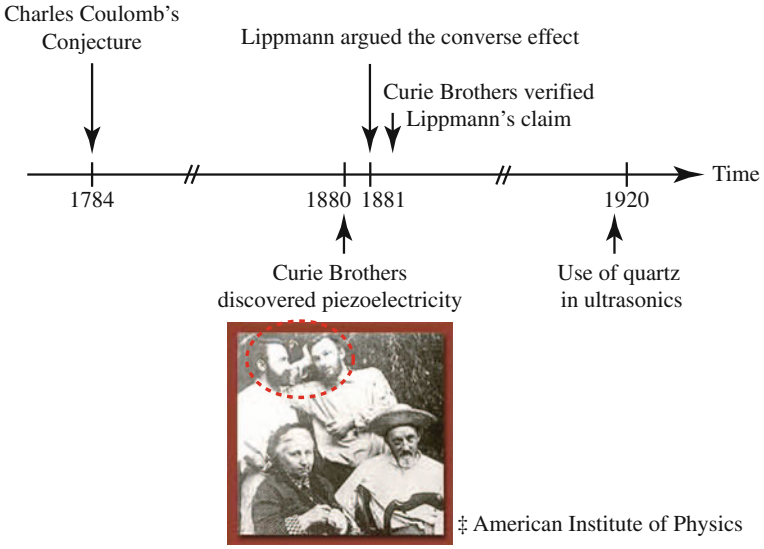


Fig. 2.1 Piezoelectricity timeline (Cady 1946)

dipole moment induces an electric charge which can be measured across the surface of the material. Conversely, an applied voltage induces a mechanical strain in the crystalline lattice (Mason 1946). The circles in Fig. 2.2 represent positive (silicon) and negative ions (oxygen pair) of the unit cell of quartz, where the small solid circle represents the center of gravity for the positively charged ions and the small open circle represents the center of gravity for the negatively charged ions. In Fig. 2.2a, the centers of gravity for both positive and negative ions coincide in the equilibrium state, therefore yielding no dipole moment. On the other hand, as the crystal is compressed by mechanical pressure, a relative displacement of the centers of gravity between the positive and negative ions induces a dipole moment as illustrated in Fig. 2.2b. Consequently, an electric potential develops along the axis of polarization; the electric potential can be measured across the surface of the crystal. Likewise, by applying a voltage across the crystal the converse effect, mechanical strain induced by an electric potential, is achieved as illustrated in Fig. 2.2c and d. For example, Fig. 2.2c shows two electrodes of opposite sign, one applied to the top and the other applied to the bottom of the unit cell. As the applied field increases, it causes the corresponding ions to move in a favorable direction, consequently inducing deformation in the crystal lattice and mechanical strain is achieved (Callister 1994). By reversing the sign of the electrodes, strain in the opposite direction is achieved as depicted in Fig. 2.2d.

Interestingly, the piezoelectric effect only occurs in crystals with no center of symmetry. Of the 32 possible classes of crystals, 20 are piezoelectric and 12 are not; therefore, this effect depends on the type of symmetry existing in the crystal. According to Ballato (1996), substances such as bone, wood, and ice exhibit the

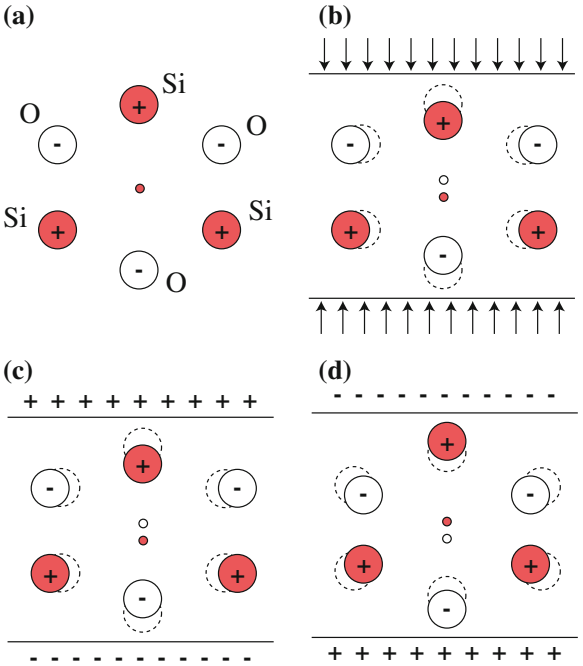


Fig. 2.2 A two-dimensional model of a unit cell for a quartz crystal illustrating the piezoelectric effect. This model was first conceived by Lord Kelvin in 1893 (Mason 1946). The *large solid circles* represent positively charged ions and the *small solid circle* represents their center of gravity. Likewise, *large open circles* represent negatively charged ions, and their center of gravity is represented by the *small open circle*. **a** The equilibrium state where there is no net dipole moment, i.e., the centers of gravity for positive and negative ions coincide; **b** mechanical stress induces an electric dipole—separation of centers of gravity for positive and negative ions; **c** and **d** an applied field produces mechanical strain. (Figure is adapted from Mason (1946)

piezoelectric effect due to the asymmetric nature of the molecules that make up the material.

Piezoelectric materials, either by mechanical stress or applied voltage, produce electric dipoles. Materials which exhibit a spontaneous polarization (i.e., electric dipoles) in the absence of an applied stress or electric field are referred to as ferroelectrics.[2] All ferroelectrics exhibit the piezoelectric effect; however, the converse is not necessarily true. For example, quartz exhibits the piezoelectric effect, but the crystal structure does not yield a spontaneous polarization, i.e., no net dipole moment in its equilibrium state because the centers of gravity for the positive and negative ions coincide as shown in Fig. 2.2a. On the other hand, the microscopic crystallites of the man-made lead-zirconate-titanate [$Pb(Ti,Zr)O_3$], otherwise known as PZT,

[2] Ferroelectricity was discovered in the late 1940s (Berlincourt 1981; King et al. 1990).

exhibit a spontaneous polarization due to the arrangements of atoms within the unit cell at room temperature.

In the 1960s, the naturally occurring monocrystalline piezoelectric materials were superseded by man-made polycrystalline ceramics such as PZT. The word "ceramics" is derived from the Greek word *keramikos*, which means "burnt stuff." It is the high-temperature heat treatment process that gives the material its unique properties. PZT ceramics are relatively easy to produce and exhibit exceptional efficiency in converting electrical energy to mechanical energy and *vice versa*. High efficiency enables the generation of large forces or displacements from relatively small applied voltages. For this reason, PZT's are the most commonly used piezoelectric material for solid-state actuation.

2.2 Piezoelectric Compositions

The piezoelectric effect in naturally occurring materials such as quart, sodium chlorate, and Rochelle salt is extremely small. Polycrystalline piezoelectric ceramics were developed with enhanced performance. Examples include potassium dihydrogen phosphate, barium sodium niobate, barium titanate, lithium niobate, lithium tantalate, and the popular PZT. Although barium titanate ($BaTiO_3$) was the first ferroelectric material used for piezo-based applications in the 1950s, PZT ceramic has since then replaced barium titanate because PZT exhibits nearly twice the piezoelectric effect (Cady 1946; Berlincourt 1981; King et al. 1990; Uchino 1991).

PZT ceramics are used extensively for solid-state actuators. Commercially available PZT ceramics come in two flavors, "hard" and "soft." Hardness in this case refers to the material's resistance to depolarization, and should not be confused with mechanical hardness. Hard PZT is doped with acceptor dopants that create oxygen (anion) vacancies. In contrast, soft PZT is doped with donor dopants, which create metal (cation) vacancies (Damjanovic and Newnham 1992). Other dopants have been used to affect aging and sensitivity, for example. The major differences between hard and soft PZT's are the operating voltage and their sensitivity to an applied field. For instance, hard PZT operates in the kilovolt range, where as soft PZT can be driven with several hundred volts. Therefore, hard PZT's are suited to high power applications, where as soft PZT's are favored in low-power generators and motor-type transducers. The average extension of a hard PZT is over half that of soft PZT; good soft PZT sensitivity is approximately 6 Å/V (King et al. 1990). Hard PZT exhibits 5–10-times less hysteresis and other nonlinearities compared to soft PZT.

Due to the many different compositions and dopants, a wide selection of PZT ceramics are available, and even some applications have their own custom formulation. The United States (U.S.) Navy established a naming scheme for the different types of PZT's, for example PZT-4 (Navy Type I), PZT-5A (Navy Type II), PZT-8 (Navy Type III), PZT-BT (Navy Type IV), PZT-5J (Navy Type V), and PZT-5H (Navy Type VI) (Etzold 2000). For additional information on other PZT types, refer

Table 2.1 Properties of common types of hard and soft PZT's at 24 °C (Morgan 1997)

Property (units)	PZT-4	PZT-8	PZT-5A	PZT-5H
ρ (kg/m^3)	7.6	>7.5	7.8	7.5
κ_{33}	0.700	0.640	0.710	0.750
κ_{31}	−0.330	−0.300	−0.340	−0.390
d_{33} (Å/V)	2.85	2.25	3.74	5.93
d_{31} (Å/V)	−1.22	−0.97	−1.71	−2.74
d_{15} (Å/V)	4.95	3.30	5.85	7.41
g_{33} ($\times 10^{-3}$ Vm/N)	24.9	25.4	24.8	19.7
g_{31} ($\times 10^{-3}$ Vm/N)	−10.6	−10.9	−11.4	−9.1
g_{15} ($\times 10^{-3}$ Vm/N)	38	28.9	38.2	26.8
Q	500	1000	75	65
T_c (°C)	325	300	365	195
$Y^{E_{11}}$ (GPa)	82	87	61	62
$Y^{E_{33}}$ (GPa)	66	74	53	48
$Y^{D_{11}}$ (GPa)	99	99	69	71
$Y^{D_{33}}$ (GPa)	126	118	106	111

κ is the coupling coefficient; d is the piezoelectric charge constant; g is the piezoelectric voltage constant; Q is the quality factor; T_c is the Curie temperature; Y^E is the short circuit elastic constant; Y^D is the open circuit elastic constant; and the Poisson's ratio for all ceramics is approximately 0.31

to Morgan Electro Ceramics (www.morganelectroceramics.com).[3] The U.S. Navy's type designation has been adopted by some vendors (e.g., Morgan Electro Ceramics), but many others have established their own convention.

Hard PZT materials include PZT-4, PZT-4D, and PZT-8. Because of their resistance to depolarization, these materials are best suited for high voltage applications. In particular, PZT-4 is suited to ultrasonic cleaning, sonar, and other high power acoustic radiation applications. Soft PZT materials include PZT-5A, PZT-5B, PZT-5B, PZT-5J, PZT-5H, and PZT-5R. PZT-5A has high sensitivity, permittivity, and time stability. For large range positioning applications, PZT-5B is often used because of its increased sensitivity and piezoelectric characteristics compared to PZT-5A. For fine positioning, PZT-5H is preferred due to its extremely high permittivity, coupling, and piezoelectric constant. However, the material has a lower time stability and the lowest Curie temperature of the soft PZT's. Table 2.1 is a representative list of the most commonly used PZT's and their properties.

Another potential material for solid-state actuation is lead magnesium niobate (PMN), an electrostrictive material (Damjanovic and Newnham 1992). The displacement response of PMN is proportional to the square of the applied field. Within a small temperature range, between +20 and +35 °C, PMN exhibit significantly lower hysteresis, between 2 and 3 %, and no creep. However, PMN has higher capacitances compared to PZT materials and thus require more power for dynamic applications.

[3] http://www.morganelectroceramics.com/tutorials/piezoguide10.html

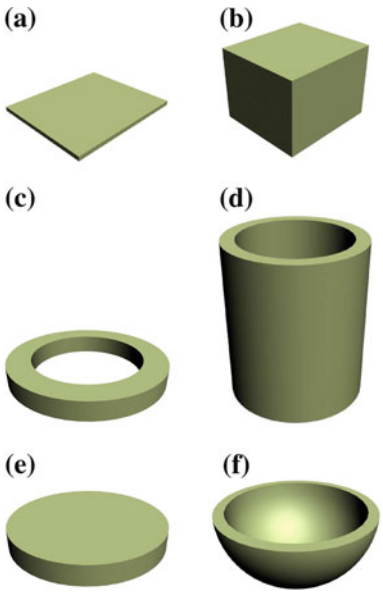


Fig. 2.3 Examples of commercially available shapes of piezoelectric ceramics

PMN is more sensitive to temperature variation compared to PZT, especially above 10 °C.

2.3 Manufacturing Piezoelectric Ceramics

The manufacturing process for piezoelectric ceramic involves a number of steps. First, the raw materials are combined and put through a ball milling process to create a powder. Then, the powder is heat treated to form a polycrystalline phase. Afterwards, the material undergoes additional ball milling and the resulting powder is mixed with a binder, then formed (by pressing) into specific shapes such as bars, plates, rods, discs, tubes, etc., as shown in Fig. 2.3. The shaped material undergoes additional heat treatment, first burning out the binder, followed by sintering. Finally, the ceramic is polished, ground, and electrodes are applied. The most common type of electrode is glass-loaded paint printed or sprayed onto the ceramic surface and then heated to create a good electrical contact. Gold is also used for good conductivity with minimum thickness as well as platinum and palladium, but they are more expensive.

After manufacture, the piezoelectric ceramic consists of randomly oriented domains; a domain is a microscopic region of material with a net polar orientation. Because of the random domain orientation, the material produces no net effect when mechanically stressed or when voltage is applied. However, through a process called *poling*, the material can be made to exhibit considerable piezoelectric effect.

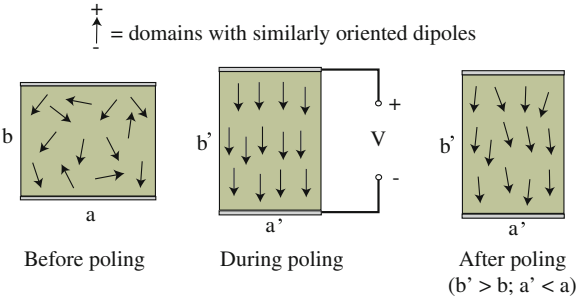


Fig. 2.4 The poling process

Basically, the poling processes forces the dipoles in the material to align in a favorable direction as shown in Fig. 2.4. The process involves heating the material near its Curie temperature, typically between 100 and 300 °C, then applying a strong electric field while cooling the material. The heating process allows movement of the individual crystallites and the application of a strong electric field causes the dipoles to align with the field in favor of a net effect (Berlincourt 1981; King et al. 1990). As the field is maintained during the cooling process, the majority of the dipoles maintain their alignment. The dimensions of the material after poling permanently changes as shown in Fig. 2.4. In the figure, the poling axis is the dimension between the poling electrodes. During poling, the material increases its dimensions parallel to the poling axis and the dimensions along the electrodes decrease. After poling, the ferroelectric material exhibits considerable piezoelectric effect.

2.4 Piezoelectric Transducers

Piezoelectric transducers are available in many shapes and forms (Physik Instrumente 2009). In addition to their traditional application in microphones, accelerometers, ultrasonic transducers, and spark generators (APC International Ltd 2002), piezoelectric transducers are now used in applications such as structural vibration control (Moheimani and Fleming 2006; Giurgiutiu 2000), precision positioning (Devasia et al. 2007), aerospace systems (Bronowicki et al. 1999), and nanotechnology (Tseng et al. 2005).

Figure 2.5 shows the basic modes of deformation for a piezoelectric element that can be exploited for nanopositioning. Based on these deformation modes, unimorph, bimorph, stack, and tube piezoelectric actuators have been developed.

Unimorphs and bimorphs are bender style actuators with large range of motion, but low force. Sawyer in 1931 developed the first bender actuator using Rochelle salt bars (Sawyer 1931). Bimorph actuators consist of two ceramic elements bonded together, and can be configured serially or in parallel. The parallel configuration

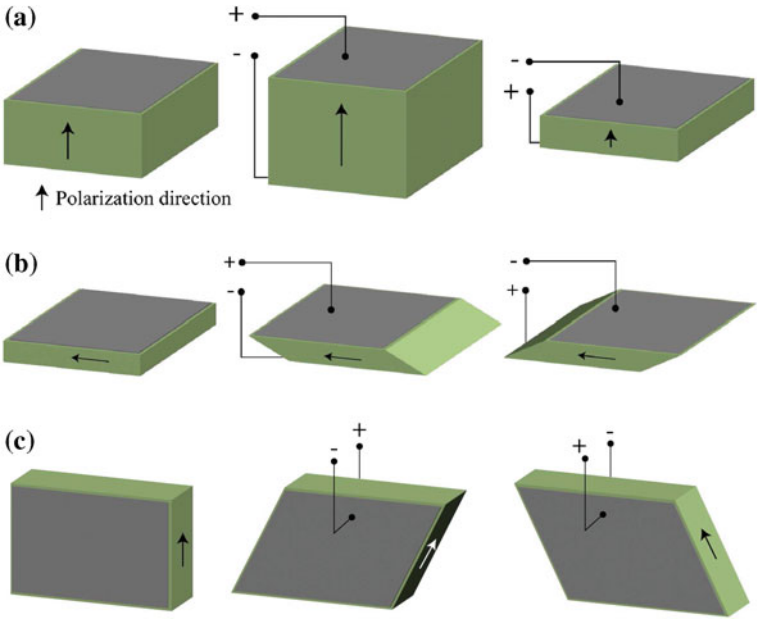


Fig. 2.5 Basic modes of piezoelectric element deformation. *Arrow* indicates direction of polarization

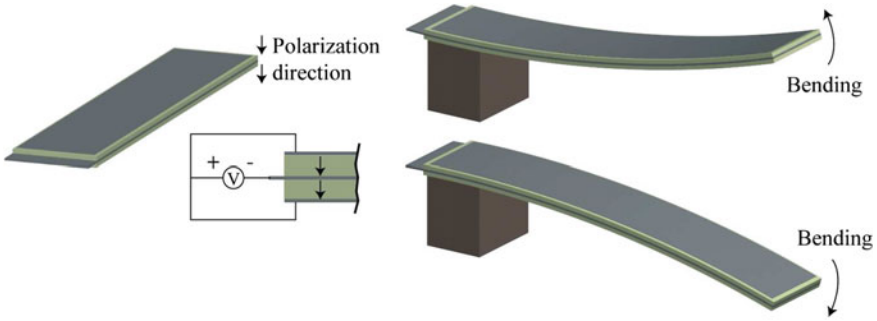


Fig. 2.6 Piezoelectric bimorph actuator

in Fig. 2.6 shows an electrode sandwiched between two piezo plates. For this configuration, the static deflection at the end can be estimated by (APC International Ltd 2002)

$$\Delta x = \frac{3d_{31}L^2V}{t^2}, \tag{2.1}$$

where L and t are the bender's length and thickness, respectively, and d_{31} is the strain coefficient (displacement normal to the polarization direction). On the contrary,

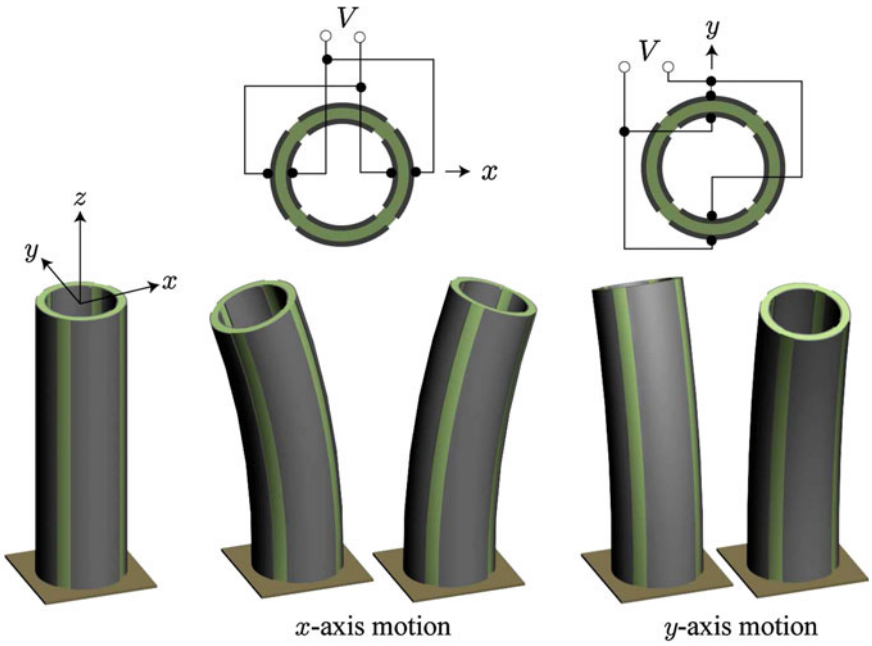


Fig. 2.7 Quarter-sectored piezoelectric tube actuator

unimorph actuators employ only one piezoceramic element that is bonded to an elastic shim, such as aluminum, brass, or steel. Bending motion for both unimorph and bimorph actuators is due to the difference in expansion and/or contraction between the opposing plates.

Quarter-sectored tube-shaped piezoelectric actuators were developed for 2- and 3-D positioning and they are used extensively in scanning probe microscopes (Croft et al. 2001). The tube-shaped PZT ceramic is poled radially and the electrodes are deposited on the inner and outer circumferential surfaces of the tube as shown in Fig. 2.7. If the inner electrode is held at ground and the two opposing electrodes are driven by ±V, then the resulting static deflection of the tube's distal end can be estimated by (Chen 1992),

$$\Delta x \approx \frac{2\sqrt{2}d_{31}L^2V}{\pi D_i t}, \tag{2.2}$$

where L, t, and D_i are the tube's length, thickness, and inside diameter, respectively. Compared to the bender style actuators discussed above, tube-shaped actuators are stiffer because of their cylindrical geometry.

Piezoelectric stack actuators emerged after the development of poled ceramic transducers of PZT (Ramsay and Mugridge 1962). A stack actuator is made by bonding thin layers of piezoelectric materials between electrodes such that the

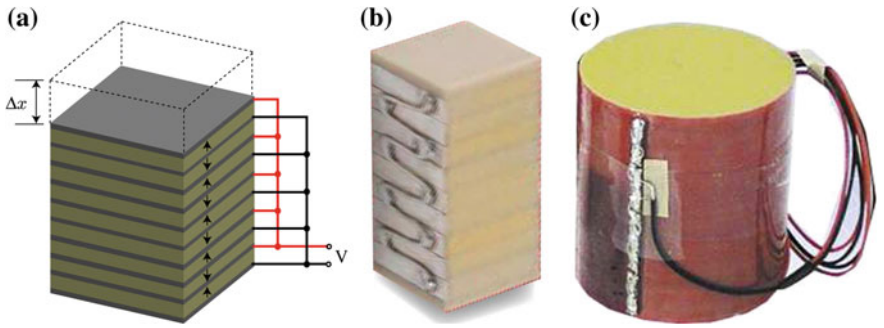


Fig. 2.8 Piezoelectric stack actuators. **a** Electrode configuration; **b** monolithic stack actuator; and **c** multilayer stack actuator (Physik Instrumente 2009)

polarization direction is aligned with the direction of stroke and blocking force. All the elements are connected in parallel as depicted in Fig. 2.8a. The thin ceramic layers (100 μm thick) wired in parallel enables the stack to be operated at 100 V or less, with an achievable stroke of 0.2 % of the stack height (APC International Ltd 2002). Due to their high stiffness and force output, stack actuators are used extensively in high-speed nanopositioning designs (Ando et al. 2002; Schitter 2007; Leang and Fleming 2008). Because the ceramic layers are connected in parallel, the overall capacitance of stacks is high compared to tubes and bender actuators, and thus power requirements must be carefully considered, in particular for dynamic applications. The static axial elongation of a stack actuator is given by

$$\Delta x = n d_{33} V, \tag{2.3}$$

where n is the number of ceramic layers and d_{33} is the strain coefficient along the axial direction of the stack.

Shear actuators make use of the shear-strain coefficient d_{15}, whereby an electric field is applied perpendicular to the polarization direction to induce shape change (see Fig. 2.5). The strain due to shear can be has much as twice the deformation of a comparable size material based on d_{33}. Some advantages include high force output and bipolar operation. When thin shear actuators are used for high-speed nanopositioning applications, the range is relatively small (Rost et al. 2005).

2.5 Application Considerations

The following discussion highlights some of the important considerations to ensure high-performance operation when designing with piezoactuators. Topics include mechanical, electrical, and thermal considerations.

2.5.1 Mounting

Piezoelectric materials are brittle and thus, proper support and the elimination of off-center loading are essential to prevent premature failure. For example, lateral or bending forces must be avoided when possible.

As illustrated in Fig. 2.9, stack actuators are more prone to damage caused by off-axis and tensile loads due to their design. Therefore, they cannot tolerate shear or bending forces and only the axial extension of the actuator is used. Bare stacks should only be mounted at their ends. Stacks with casings can be mounted at their ends or circumference. If this is unavoidable, the resultant force must be directed as much as possible axially. Only pure axial force should be allowed to be transferred between the stack and a coupled mechanical component. To minimize point-loading on the ends of a bare stack, a face plate can be combined with a ball tip or smooth coupler to distribute the load. Avoid misaligned contact planes which produce localized stress and nonaxial loading that can damage the stack.

When an actuator is glued to a substrate or other component, it is recommended that a very thin layer of glue be used. Epoxy-type adhesives are recommended and room temperature adhesives for mechanical assembly is recommended. The pressure during the curing process should fall between 2 and 5 MPa. Flexible electrically conductive glues should be used for electrical connections to minimize failure due to fatigue.

Operation in a humid environment is not recommended as this will increase the chances of arcing between the electrodes. If this is unavoidable, consider surrounding the actuator with a nonconductive coating. High temperatures, especially those near the material's Curie temperature, can depole the piezoelectric actuator and must be avoided entirely. In addition, hysteresis losses in the material during actuation may causes excessive heating of the material and must be taken into account during high-speed operation.

2.5.2 Stroke Versus Force

In nanopositioning, the main priority for a piezoelectric actuator is its displacement or stoke and positioning resolution. One must also keep in mind that a piezoelectric actuator generates a combination of stroke and force. The stroke of a piezoelectric actuator depends on the mounting arrangement (i.e., boundary conditions), the applied preload, and the interconnecting components. Figure 2.10 shows an idealized stroke versus force curve for a piezoelectric actuator. When unloaded (zero force), the actuator generates the largest stroke/displacement $\Delta x_{\max}$. In contrast, when constrained from expanding (i.e., zero stroke), the actuator generates the maximum force, the blocking force F_{block}.

When a piezoelectric actuator is coupled to an external spring which applies a load, the achievable stroke is less than the maximum displacement $\Delta x_{\max}$. The achievable

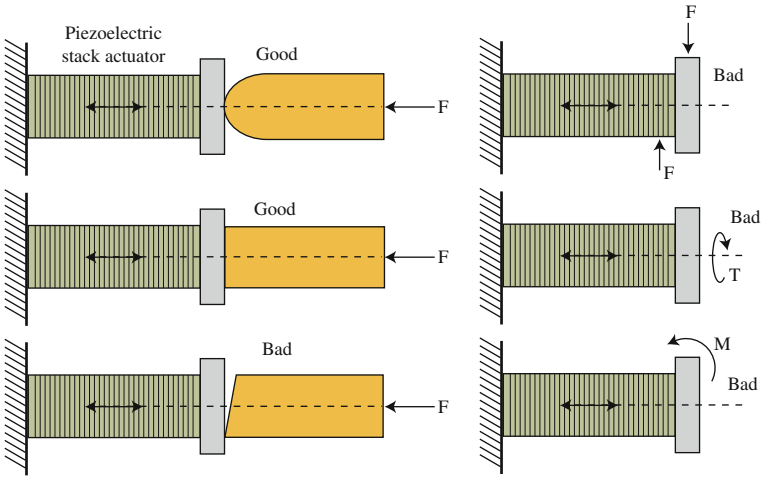


Fig. 2.9 Examples of good and bad mounting arrangements for a piezoelectric stack actuator

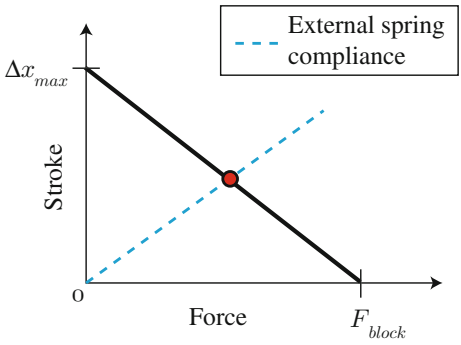


Fig. 2.10 A stroke versus force curve for a piezoelectric actuator

stroke (and force) is the intersection of the compliance of the spring with the stroke versus force curve as shown in Fig. 2.10. For example, the resulting stroke of a piezoelectric actuator with stiffness k_p pushing against a spring or flexure with stiffness k_f is

$$\Delta x = \Delta x_{\max} \left(\frac{k_p}{k_p + k_f} \right). \tag{2.4}$$

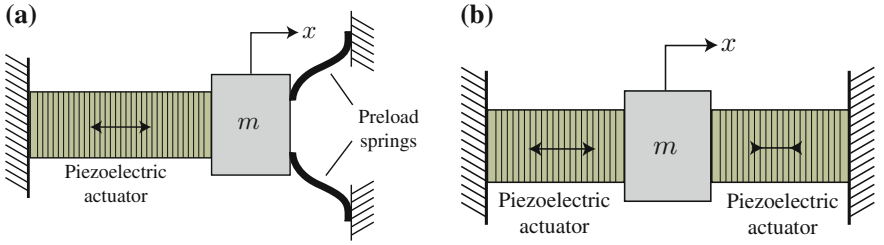


Fig. 2.11 Two possible preload configurations: **a** passive approach using springs or flexures and **b** active, antagonistic approach

2.5.3 Preload and Flexures

The fact that piezoelectric ceramic material is brittle and piezoelectric stack actuators are constructed from many layers of piezoelectric materials either glued or fused together, tensile loads must be avoided. Excessive tensile forces caused by inertial loads occur during high-speed positioning. In this case, preload is applied to the actuator to compensate for the damaging forces during dynamic operation.

Two possible preload configurations are shown in Fig. 2.11. The passive method (a) involves springs (or flexures) to create the necessary preload force to counteract the inertia forces generated by the effective mass of the piezoelectric actuator and the tip mass m. Some advantages include simplicity and low cost; however, the actuator always acts against a force and possible resonances due to the springs may emerge. The active approach (b) involves two opposing piezoelectric actuators, one pushing and the other pulling. The advantages include no additional resonances due to springs and the actuators do not work against an opposing force; however, because there are two actuators, either the cost of drive electronics increases or if they are driven by a single amplifier, the overall actuator capacitance and electrical power increase.

In some cases, preload enhances the mechanical performance of a piezoelectric actuator. For example, some ceramics show enhanced strain (Bryant et al. 2005).

The speed of response is dictated by the mass of the parts to be moved and the output of the voltage amplifiers. It is noted that frequencies in the 10 MHz range can be generated by piezo-based ultrasonic transducers.

During expansion and contraction, a piezoelectric actuators behaves like a corkscrew, twisting as it displaces due to small manufacturing imperfections. Twisting must be minimized for two reasons. First, the twisting can cause parasitic motion such as runout which may affect the precision of the positioner. The use of adequate guiding mechanisms such as flexures and ball tips will minimize runout. Second, the twisting can induce lateral and bending forces that can damage the piezoelectric actuator. The forces can be decoupled using flexures and ball tips. Additionally, contact surfaces must be ground smooth to avoid off-center loading.

2.5.4 Electrical Considerations

In general, the capacitance of a piezoelectric actuator varies with the applied voltage, load, and even temperature. The capacitance value is important for calculating the required electrical power for a given dynamic response. It is estimated that approximately 5 % of the power consumption is dissipated into heat.

One wants to minimize the capacitance of the piezoelectric actuator, since this will minimize the electrical current/power consumption. In fact, it is required to minimize the dielectric constant of the piezoelectric actuator. The low dielectric materials also has an advantage; they operate at higher temperatures and offer better stability against depoling (APC International Ltd 2003). One exception to this rule is low dielectric ultrasound PZT materials. Such materials provide very low strain and are not suited for actuation (APC International Ltd 2003).

Driving piezoelectric materials at or near their maximum rated voltage may reduce the mean time to failure. Some commercial vendors provide endurance data comparing the predicted time to failure. For example, the mean time to failure of piezoelectric actuators from Tokin Corporation of Japan operating at the severe operating conditions (150 VDC, 40 °C, 90 % relative humidity) is predicted at 5,000 h. However, when the actuator is operated at the recommended operating conditions (100 VDC, 25 °C, 60 % RH) the mean time failure increases to 24,500 h. An empirical formula that predicts the mean time to failure t_m is given by

$$t_m = 500 \times 3.2\frac{150}{V} \times 4.9\frac{90}{\text{RH}} \times 1.5\frac{40 - T}{10}, \tag{2.5}$$

where V is the drive voltage, RH is the relative humidity (for 60 % RH = 60), and T is the ambient temperature. More details on electrical considerations can be found in Chap. 14.

2.5.5 Self-Heating Considerations

Piezoelectric actuators during dynamic operation experience self-heating. The self-heating increases with actuation frequency and amplitude. As discussed below in Sect. 2.6.3, the piezomechanical and electrical properties of PZT 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. However, heat management can be challenging due to the low thermal conductivity of PZT. 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. Companies such as APC offer specialized actuator configurations, such as the "ThermoStable" technique, to improve heat management (APC International Ltd 2003).

2.6 Response of Piezoelectric Actuators

So far in this discussion it has been assumed that piezoelectric transducers expand and contract proportionally to applied voltage. Unfortunately, this assumption is not accurate and is particularly erroneous when considering moderate or high electric fields, and when the frequency of operation becomes high. There are three significant sources of error that degrade and complicate the response of piezoelectric transducers. These are discussed below under the headings: Hysteresis, Creep, and Temperature Dependence. In addition, problems also arise from the highly capacitive nature and structural dynamics of piezoelectric actuators. These restrict speed and are discussed in the final two headings: Actuator (or vibrational) Dynamics and Electrical Bandwidth.

2.6.1 Hysteresis

Hysteresis, which is a nonlinear behavior between the applied electric field and the mechanical displacement of a piezoelectric actuator, is believed to be caused by irreversible losses that occur when similarly oriented electric dipoles interact upon application of an electric field (Jiles and Atherton 1986). Hysteresis is significant over large-range displacements (Barrett and Quate 1991; Adriaens 2000). Figure 2.12a and b shows the effect of hysteresis in an experimental piezo-based system. A typical hysteresis curve which depicts the nonlinear relationship between the output displacement and applied input voltage is shown in Fig. 2.12c. This nonlinear effect leads to distortion in scanning probe microscopy (SPM)-based imaging as shown in Fig. 2.12d. Although the actual features are oriented in a parallel fashion, hysteresis causes the features to appear curved and distorted. More specifically, the distortion is caused by plotting the information collected about the sample topology with respect to the desired position of the probe. Because of hysteresis, the probe does not achieve the desired position, therefore leading to the distorted image. In addition to poor positioning accuracy, hysteresis causes poor repeatability and the mixing of harmonic content into the displacement response. Hysteresis can be avoided by operating in the linear range, i.e., over short range displacements; however, this limits the achievable positioning range. Controlling the charge delivered to the piezoelectric transducer, rather than the voltage, helps to minimize hysteresis (Fleming and Moheimani 2005). More on hysteresis, modeling, and control methods to mitigate its effects can be found in Chap. 11.

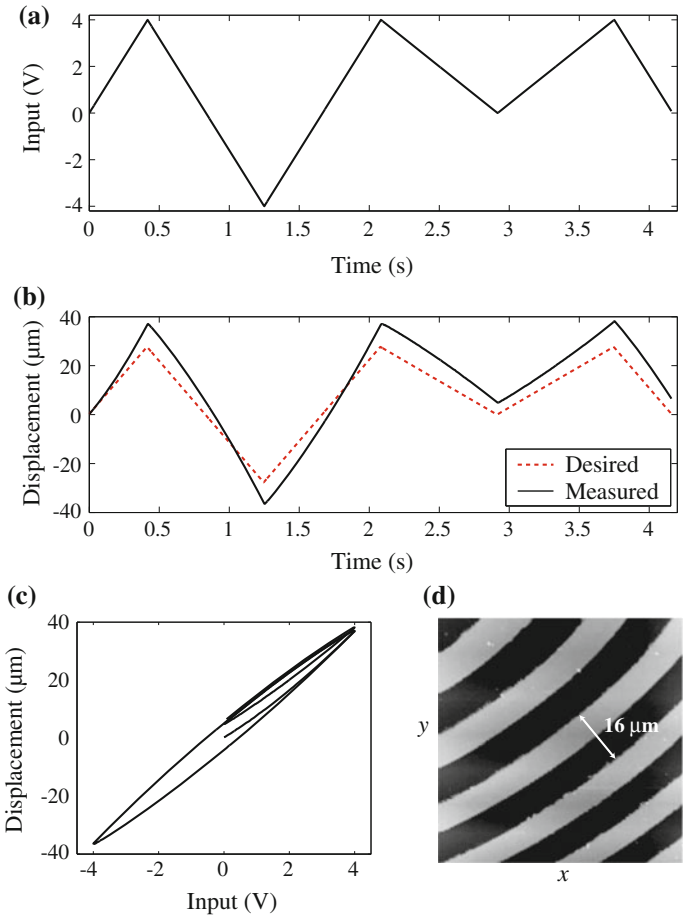


Fig. 2.12 Hysteresis and its effects on SPM: **a** applied input versus time, **b** resulting output displacement versus time, **c** displacement versus input curve (hysteresis curve) and **d** distortion in AFM imaging of 16-μm pitch encoder gratings due to hysteresis effect. The actual features are parallel

2.6.2 Creep

When a piezoelectric transducer is commanded by a step change in voltage, the response consists of high-frequency transients followed by low-frequency drift known as creep. The time constant for creep is typically a few minutes. Creep severely degrades the low-frequency and static positioning ability of piezoelectric actuators (Hues et al. 1994; Koops et al. 1999; Jung and Gweon 2000). In mechanics, creep is a rate-dependent deformation of the material when subjected to a constant load or stress (Callister 1994). Similarly, creep in piezoelectric materials is a rate-dependent

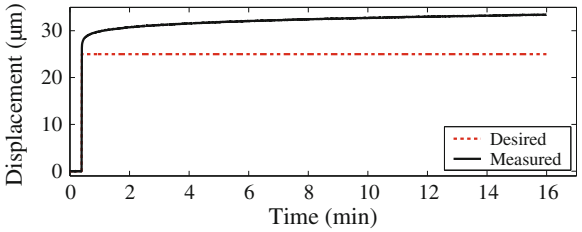


Fig. 2.13 The effects of creep in the output displacement measured over a period of 15 min

deformation due to a constant electric field. Creep manifests itself as the remnant polarization slowly increases after the onset of a constant field. Figure 2.13 shows the effect of creep in the positioning of an experimental piezoactuator. The actuator is commanded to a reference position, say 25 μm, but after a period of 15 min, the actuator's position creeps to a new position of 33.41 μm. As a result, the error due to creep is 24.44 % of the total displacement range in this case.

One method to avoid creep is to operate fast enough so that the creep effect becomes negligible (Croft et al. 2001); however, such effort prevents the use of piezo positioners in slow and static applications. For example, because of drift, it is difficult to precisely fabricate nanofeatures using AFMs when the process timescale is on the order of minutes, e.g., see (Hues et al. 1994).

Methods to compensate for creep have been well studied in the past and some examples include the use of feedback control, e.g., (Barrett and Quate 1991; Schitter et al. 2001; Schitter and Stemmer 2002; Salapaka et al. 2002), and model-based feedforward control, e.g., (Jung and Gweon 2000; Janocha and Kuhnen 2000; Jung et al. 2000; Croft et al. 2001; Krejci and Kuhnen 2001; Rifai and Youcef-Toumi 2002).

2.6.3 Temperature Dependence

Both the piezoelectric strain constant d and dielectric permittivity ε of PZT vary widely with temperature. For example, when PZT is cooled down to 77 °C or lower, the capacitance, hysteresis, and the strain constant d_{33} reduce (Lee and Saravanos 1998; APC International Ltd 2003). At low temperature, the material is less prone to depoling. Figure 1.7 shows the normalized sensitivity for a tube scanner plotted against temperature. When driven with voltage, the response increases by 10 % every 25°. Such temperature dependence effectively renders piezoelectric transducers useless as calibrated force or displacement actuators. When used within a feedback loop, the controller must have sufficient gain-margin and performance robustness to cope with the full range of system gain.

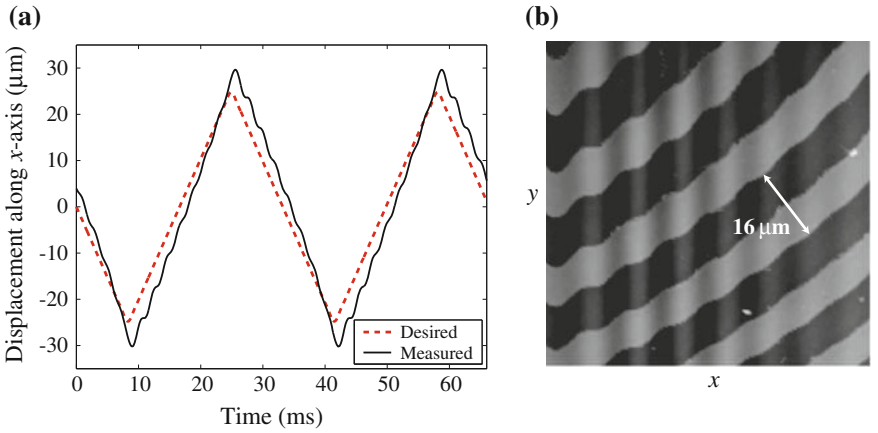


Fig. 2.14 The effects of vibration (and hysteresis) scanning at 30 Hz: **a** displacement versus time response, and **b** distorted AFM image, but actual features are parallel

2.6.4 Vibrational Dynamics

Vibration (or actuator) dynamics, such as structural resonances, limits the operating bandwidth of piezo-based positioning systems. The effect is caused by command signals exciting the flexible modes of the structure (Holman et al. 1995; Croft and Devasia 1999). For example, the frequency response of a piezo-based positioner typically reveals sharp resonance peaks. These peaks can easily be excited by certain command signals like triangle inputs applied to control the positioner. Figures 2.14a and b clearly illustrate the effect of vibration, where oscillations cause significant tracking error in the displacement versus time response (Fig. 2.14a). Such effects cause distortion in the SPM-based imaging, for example, the rippling effect in the AFM image shown in Fig. 2.14b. Typically, scan rates (i.e., scan frequencies) are restricted to less than 1/10th–1/100th of the first resonant frequency, thus limiting the bandwidth of piezo-based systems because the achievable scan rate is lower for increased resolution in positioning. However, higher operating speed can be achieved by using *stiffer* piezoactuators with higher resonant frequencies (Sulchek et al. 2000; Schitter 2007; Leang and Fleming 2008), for example, Ando et al. (2002) used a *stiff* piezo with a resonant frequency of 260 kHz in an AFM to image biological macromolecules in action.

But in general these stiff piezos have shorter effective displacement ranges. Therefore, the use of stiffer piezos to increase bandwidth also leads to reduction of positioning range.

2.6.5 Electrical Bandwidth

For actuators with high resonance frequencies, the foremost bandwidth limitations were identified in (Fleming 2008) as amplifier output impedance and cable inductance. These form a resonant low-pass filter with the load capacitance C_p. Even low output impedances of 1 Ω impose a positioning bandwidth of only 2.8 kHz with a 10 μF load (10° phase lag).

2.7 Modeling Creep and Vibration in Piezoelectric Actuators

Neglecting the effects of temperature variations and self-heating, the displacement (stoke) response of a piezoelectric actuator under an applied voltage in general consists of dynamics (vibrations) and nonlinearities, such as those associated with the actuator and/or motion mechanism. In this section, the modeling of a piezo-based nanopositioner is considered, where the focus is on the linear dynamics and the modeling of hysteresis is covered in Chap. 11. The main objective is to describe a modeling approach which enables the application of feedforward as well as feedback control for precision positioning. It is noted that the forgoing discussion also applies to other types of active material actuators, provided the material's intrinsic behavior is carefully considered.

An example of the combined dynamic and hysteresis effects measured from a tube-shaped piezoactuator is shown in Fig. 2.15. The measured output response is obtained by applying a 30 Hz triangle input signal to drive the actuator between 0 and approximately 32 μm. The oscillations shown in the figure are caused by vibrational dynamics; the slow upward-drift of the output over time is due to the creep effect; and finally, a noticeable curved distortion in the output trajectory is due to hysteresis.

To effectively model these behaviors, one must consider the operating conditions and when certain effects dominate. For instance, the amount of the dynamics on the output response depends on the operating frequency. The operating frequency is determined by the application in mind. As a result, the system's *dynamics*, be it low-frequency or high-frequency, are excited by the frequency of the input signal.

Take for example when the input frequency is close to a nanopositioning system's resonance frequencies. In this case, vibration becomes noticeably large. This is not surprising when one examines the frequency response of a typical piezoelectric actuator, where sharp resonances are common as shown in Fig. 2.16. Piezoactuators tend to be highly resonant structures due to their high stiffness and low structural damping. Therefore, input signals such as sawtooth signals can excite the piezoactuator's resonances, causing the output to oscillate or vibrate as previously shown in Fig. 2.15.

At slow operating speeds, creep is a major source of positioning error. Creep in piezoactuators is a low-frequency behavior, where the output drifts, especially when the operation is offset from the center of the piezoactuator's positioning range (see Fig. 2.15).

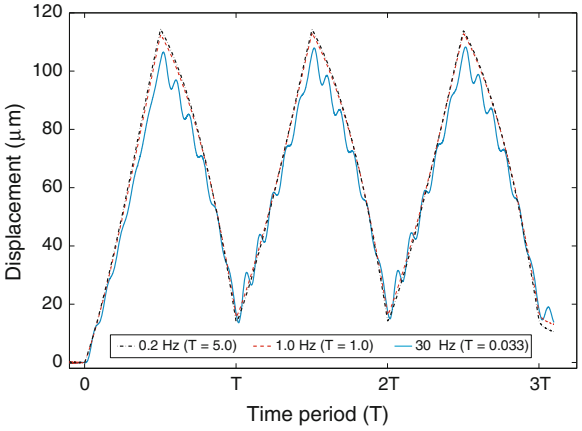


Fig. 2.15 Measured piezoactuator response (scanning at 30 Hz) showing the combined effects of vibration, creep, and hysteresis

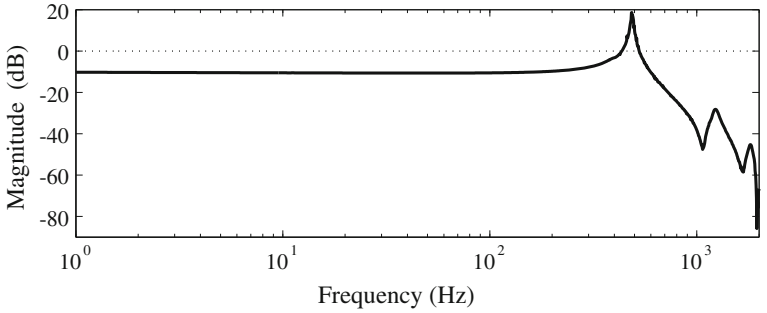


Fig. 2.16 Frequency response of a tube-shaped piezoactuator

Finally, hysteresis is significant over large-range displacements (Barrett and Quate 1991). The operation of piezoactuators in their linear range helps avoid hysteresis. In general, the linear range is within 5 % of the maximal range of motion.

The hysteresis and dynamic effects are coupled (Croft et al. 2001). For instance, when the movement of the piezoactuator is large and slow, the piezoactuator exhibits hysteresis and creep effects. As the input frequency increases, the piezoactuator's output response shows the addition of the vibrational dynamics. To model these behaviors, the cascade model depicted in Fig. 2.17a is used. The range-dependent hysteresis effect is treated as a rate-independent, input nonlinearity represented by $\mathcal{H}[u(\cdot)]$. Chapter 11 discusses various models used to represent the hysteresis behavior. The vibrational dynamics and creep effects are typically captured by the linear dynamics model $G(s)$. The cascade model structure is used extensively to model piezoactuators and similar systems (Croft et al. 2001; Tan and Baras 2005).

At very low speed, the creep effect is significant. This effect can be captured by the Kelvin–Voigt model, which consists of spring (k_i) and damper (c_i) elements

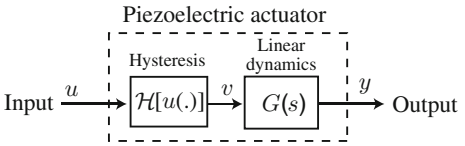


Fig. 2.17 A cascade model structure for hysteresis, vibrational dynamics, and creep effects in piezoactuators. Hysteresis, denoted by $\mathcal{H}[u(\cdot)]$, is modeled as an rate-independent, input nonlinearity that is output-range dependent. The input frequency-dependent vibrational dynamics and creep effects follow the hysteresis block, and they are captured by the linear dynamics model $G(s)$

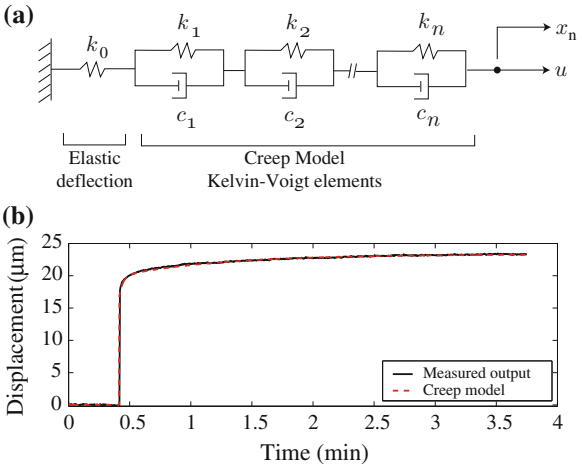


Fig. 2.18 **a** The spring-damper model for the creep effect. The parameters of the model are found by curve fitting the measured step response of the piezoactuator. **b** The time response shows the measured step response (*solid line*) and the linear creep model (*dashed line*)

(Malvern 1969; Janocha and Kuhnen 2000). The lumped-parameter model shown in Fig. 2.18a is linear, and its transfer function is

$$G_c(s) = \frac{x(s)}{u(s)} = \frac{1}{k_0} + \sum_{i=1}^{n} \frac{1}{sc_i + k_i}, \tag{2.6}$$

where $x(s)$ is the displacement of the piezoactuator and $u(s)$ is the applied input voltage. In (2.6), k_0 models the elastic behavior at dc, and the creep behavior is captured by selecting an appropriate model order corresponding to the number of spring-damper elements n. The parameters k_0, k_i, and c_i of (2.6) are determined by curve fitting the step response of the piezoactuator over, for example, a 3-min period as shown in Fig. 2.18b. The second-order model ($n = 2$) in Fig. 2.18b is

$$G_c(s) = \frac{0.4s^2 + 9.9s + 7.5}{s^2 + 20.9s + 14.7}. \tag{2.7}$$

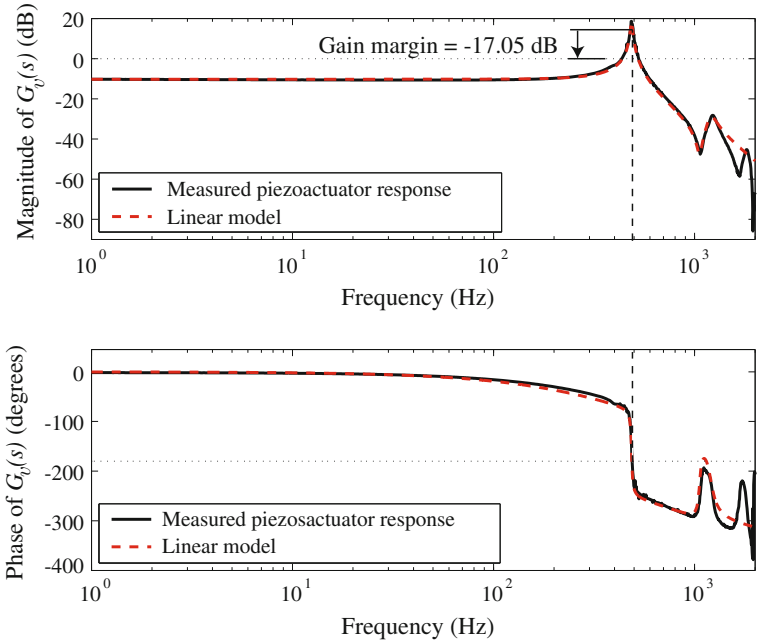


Fig. 2.19 The frequency response of a piezoactuator for modeling the vibrational dynamics. These *plots* show the measured frequency response (*solid line*), the magnitude and phase versus frequency, of the piezoactuator over ±2-μm displacement range. The *dash line* is the linear vibrational dynamics model

The creep effect can also be modeled by a logarithmic model of the form

$$x(t) = x_0 \left[1 + \gamma \log_{10} \left(\frac{t}{t_0} \right) \right], \tag{2.8}$$

where $x(t)$ is the actuator's displacement, x_0 is the nominal fast displacement to the applied voltage, γ is the creep rate, and t_0 is the settling time of the high-frequency dynamics (Rifai and Youcef-Toumi 2002).

Piezoelectric actuators can be modeled as displacement or force actuators. Details of the dynamics model where the actuator is treated as a force actuator is described in Chap. 8. On the other hand, empirical transfer function models for the vibrational dynamics can also be obtained by curve fitting the measured frequency and time responses over appropriate frequency ranges. For example, to find the vibrational dynamics model, which relates the input u to the displacement in the x, y, or z axis, a system identification algorithm is applied to the measured frequency response. Consider a tube-shaped piezoactuator with transverse range of approximately 100 μm. First, the frequency response along the x-axis is measured over a displacement range of less than 5 % of the maximal range to avoid the hysteresis effect. To avoid creep, the response is measured over a wide frequency range, in this case, 1 Hz to 2 kHz.

Frequency responses can be obtained using commercially available dynamic signal analyzers (such as Stanford Research Systems SRT785). The solid line in Fig. 2.19 shows the measured frequency response curve for the piezoactuator in the x axis.

Using a system identification algorithm, such as the function "invfreqs" in Matlab, a transfer function model is fitted to the measured response. The dash line shown in Fig. 2.19 is the model given by

$$G_v(s) = \frac{a_2 s^2 + a_1 s + a_0}{s^6 + b_5 s^5 + b_4 s^4 + b_3 s^3 + b_2 s^2 + b_1 s + b_0}, \tag{2.9}$$

where $a_2 = 7.2 \cdot 10^{13}$, $a_1 = 2.3 \cdot 10^{16}$, $a_0 = 3.2 \cdot 10^{21}$, $b_5 = 1.1 \cdot 10^4$, $b_4 = 9.5 \cdot 10^7$, $b_3 = 7.0 \cdot 10^{11}$, $b_2 = 2.0 \cdot 10^{15}$, $b_1 = 5.6 \cdot 10^{18}$, and $b_0 = 1.0 \cdot 10^{22}$.

2.8 Chapter Summary

This chapter has dealt mainly with the electromechanical properties of piezoelectric actuators, key design considerations, and the modeling of the piezoactuator dynamics, such as vibration and creep.

References

Adriaens HJMTA, de Koning WL, Banning R (2000) Modeling piezoelectric actuators. IEEE/ASME Trans Mechatron 5(4):331–341

Ando T, Kodera N, Maruyama D, Takai E, Saito K, Toda A (2002) A high-speed atomic force microscope for studying biological macromolecules in action. Jpn J Appl Phys, Part 1 41(7B): 4851–4856

APC International Ltd. (2002) Piezoelectric ceramics: principles and applications. APC International Ltd, Mackeyville

APC International Ltd. (2003) Piezo-mechanics: an introduction. Pleasant Gap, APC International Ltd., Pennsylvania

Ballato A (1996) Piezoelectricity: history and new thrusts. In: IEEE ultrasonics symposium, pp 575–583

Barrett RC, Quate CF (1991) Optical scan-correction system applied to atomic force microscopy. Rev Sci Instrum 62(6):1393–1399

Berlincourt D (1981) Piezoelectric ceramics: characteristics and applications. J Acoust Soc Am 70(6):1586–1595

Bronowicki AJ, Abhyankar NS, Griffin SF (1999) Active vibration control of large optical space structures. Smart Mater Struct 8(6):740–752

Bryant RB, Mossi KM, Robbins JA, Bathel BF (2005) The correlation of electrical properties of prestressed unimorphs as a function of mechanical strain and displacement. Integr Ferroelectr 71:267–287

Cady WG (1946) Piezoelectricity. McGraw-Hill, New York

Callister WD (1994) Materials science and engineering: an introduction. Wiley, New York

Chen CJ (1992) Electromechanical deflections of piezoelectric tubes with quartered electrodes. Appl Phys Lett 60(1):132–134
Croft D, Devasia S (1999) Vibration compensation for high speed scanning tunneling microscopy. Rev Sci Instrum 70(12):4600–4605
Croft D, Shed G, Devasia S (2001) Creep, hysteresis, and vibration compensation for piezoactuators: atomic force microscopy application. Trans ASME J Dyn Syst Measure Control 123:35–43
Damjanovic D, Newnham RE (1992) "Electrostrictive and piezoelectric materials for actuator applications. J Intell Mater Syst Struct 3(2):190–208
Devasia S, Eleftheriou E, Moheimani SOR (2007) A survey of control issues in nanopositioning. IEEE Trans Control Syst Technol 15(5):802–823
Etzold KF (2000) Ferroelectric and piezoelectric materials. CRC Press LLC, Boca Raton
Fleming AJ (March 2008) Techniques and considerations for driving piezoelectric actuators at high-speed. In: Proceedings of SPIE smart materials and structures, San Diego, CA
Fleming AJ, Moheimani SOR (2005) A grounded-load charge amplifier for reducing hysteresis in piezoelectric tube scanners. Rev Sci Instrum 76:073707
Giurgiutiu V (2000) Review of smart-materials actuation solutions for aeroelastic and vibration control. J Intell Mater Syst Struct 11:525–544
Holman AE, Scholte PMLO, Heerens WC, Tuinstra F (1995) Analysis of piezo actuators in translation construction. Rev Sci Instrum 66(5):3208–3215
Hues SM, Draper CF, Lee KP, Colton RJ (1994) Effect of PZT and PMN actuator hysteresis and creep on nanoindentation measurements using force microscopy. Rev Sci Instrum 65(5):1561–1565
Janocha H, Kuhnen K (2000) Real-time compensation of hysteresis and creep in piezoelectric actuators. Sens Actuators, A 79:83–89
Jiles DC, Atherton DL (1986) Theory of ferromagnetic hysteresis. J Magn Magn Mater 61:48–60
Jung H, Gweon D-G (2000) Creep characteristics of piezoelectric actuators. Rev Sci Instrum 71(4):1896–1900
Jung H, Shim JY, Gweon D (2000) New open-loop actuating method of piezoelectric actuators for removing hysteresis and creep. Rev Sci Instrum 71(9):3436–3440
King TG, Preston ME, Murphy BJM, Cannell DS (1990) Piezoelectric ceramic actuators: a review of machinery applications. Precis Eng 12(3):131–136
Koops KR, Scholte PMLO, Koning WLd (1999) Observation of zero creep in piezoelectric actuators. Appl Phys A 68:691–697
Krejci P, Kuhnen K (2001) Inverse control of systems with hysteresi and creep. IEE Proc Control Theory Appl 148(3):185–192
Leang KK, Fleming AJ (2008) High-speed serial-kinematic AFM scanner: design and drive considerations. In American control conference, invited session on modeling and control of SPM, 2008, pp. 3188–3193
Lee H-J, Saravanos DA (1998) The effect of temperature dependent material properties on the response of piezoelectric composite materials. J Intell Mater Syst Struct 9(7):503–508
Malvern LE (1969) Introduction to the mechanics of a continuous medium. Prentice-Hall, Englewood Cliffs
Mason WP (1946) Quartz crystal applications. D. Van Nostrand Co., Inc, New York, pp. 11–56
Mason WP (1981) Piezoelectricity, its history and applications. J Acoust Soc Am 70(6):1561–1566
Moheimani SOR, Fleming AJ (2006) Piezoelectric transducers for vibration control and damping. Springer, Berlin
Morgan Matroc Inc (1997) Guide to modern piezoelectric ceramics, review. Morgan Matroc Inc, Bedford, pp 7–91
Physik Instrumente (2009) Piezo Nano Positioning: Inspirations 2009
Ramsay JV, Mugridge EGV (1962) Barium titanate ceramics for fine-movement control. J Sci Instrum 39:636–637
Rifai OME, Youcef-Toumi K (2002) Creep in piezoelectric scanners of atomic force microscopes. In: Proceedings of American control conference, 2002, pp 3777–3782

Rost MJ, Crama L, Schakel P, van Tol E, van Velzen-Williams GBEM, Overgauw CF, ter Horst H, Dekker H, Okhuijsen B, Seynen M, Vijftigschild A, Han P, Katan AJ, Schoots K, Schumm R, van Loo W, Oosterkamp TH, Frenken JWM (2005) Scanning probe microscopes go video rate and beyond. Rev Sci. Instrum 76(5):053710-1–053710-9

Salapaka S, Sebastin A, Cleveland JP, Salapaka MV (2002) High bandwidth nano-positioner: a robust control approach. Rev Sci Instrum 73(9):3232–3241

Sawyer CB (1931) The use of Rochelle salt crystals for electrical reproducers and microphones. Proc Inst Radio Eng 19(11):2020–2029

Schitter G, Menold P, Knapp HF, Allgöwer F, Stemmer A (2001) High performance feedback for fast scanning atomic force microscopes. Rev Sci Instrum 72(8):3320–3327

Schitter G, Stemmer A (2002) Fast closed loop control of piezoelectric transducers. J Vac Sci Technol B 20(1):350–352

Schitter G, Åström KJ, DeMartini BE, Thurner PJ, Turner KL, Hansma PK (2007) Design and modeling of a high-speed AFM-scanner. IEEE Trans Control Syst Technol 15(5):906–915

Sulchek T, Hsieh R, Adams JD, Minne SC, Quate CF (2000) High-speed atomic force microscopy in liquid. Rev Sci Instrum 71(5):2097–2099

Tan X, Baras JS (2005) Adaptive identification and control of hysteresis in smart materials. IEEE Trans Autom Control 50(6):827–839

Tseng AA, Notargiacomob A, Chen TP (2005) Nanofabrication by scanning probe microscope lithography: a review. J Vac Sci Technol 23(3):877–894

Uchino K (1991) Engineering materials handbook: ceramics and glass, vol 4. ASM International, Ohio

Chapter 3
Types of Nanopositioners

The term nanopositioner is used generally to describe a wide variety of mechanical positioning devices with resolution in the nanometer range. Typically, a nanopositioner comprises a moving platform suspended by a number of compliant mechanisms or flexures (see example in Fig. 1.1). The flexures may provide a preloading force on the actuator and guide the motion of the stage. A key feature of compliant mechanisms is that they are free from the major nonlinearities such as backlash and friction that preclude traditional mechanisms, such as roller bearings, from achieving nanometer resolution. The final defining feature of nanopositioning systems is that they utilize linear translational actuators such as piezoelectric or electrostrictive actuators. Electromagnetic and other smart material actuators are also occasionally used.

This chapter describes the operation and physical characteristics of some typical nanopositioning devices. Particular attention is paid to the ubiquitous piezoelectric tube nanopositioner and the lateral flexure-based nanopositioner. Four experimental systems are discussed that will be used in the following chapters for demonstration.

3.1 Piezoelectric Tube Nanopositioners

A piezoelectric tube scanner is a thin cylinder of radially poled piezoelectric material with four external electrodes and a grounded internal electrode. When a voltage is applied to one of the external electrodes, the actuator wall expands, which causes a vertical contraction and a large lateral deflection of the tube tip. A basic piezoelectric tube scanner with four quadrant electrodes is illustrated in Fig. 3.1a.

Piezoelectric tube scanners were first reported in Binnig and Smith (1986) for use in scanning tunneling microscopes (Meyer et al. 2004). They were found to provide a higher positioning resolution and greater bandwidth than traditional tripod positioners while being simple to manufacture and easier to integrate into a microscope. Piezoelectric tube scanners are now used extensively in scanning probe microscopes

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_3,

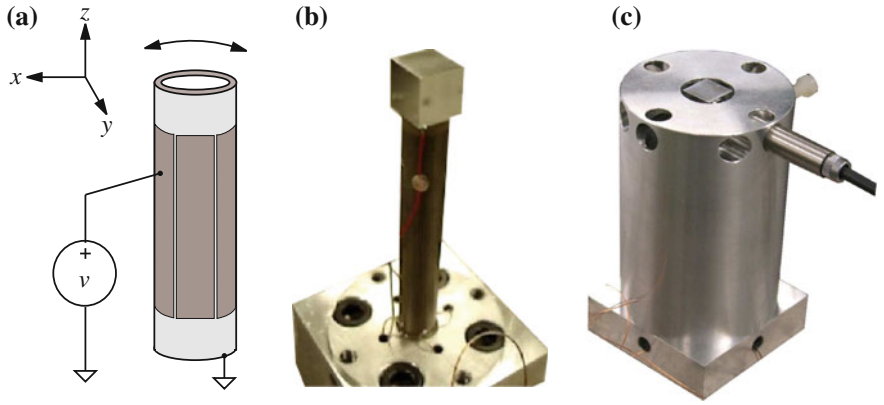


Fig. 3.1 **a** A piezoelectric tube scanner with one x-axis electrode driven by a voltage source. **b** The 63 mm tube described in Sect. 3.1.1, mounted inside an aluminum shield (**c**). A capacitive sensor is mounted parallel to a cube mounted on the tube tip

and many other applications requiring precision positioning, e.g., nanomachining (Croft et al. 1998; Gao et al. 2000), etc.

The length, diameter, and wall thickness of a piezoelectric tube scanner defines the available scan range and mechanical bandwidth. Longer, narrower tubes of around 50–80 mm are used for achieving large deflections of around 100 μm, while shorter tubes of around 15 mm are used for small deflections of 1 μm or less. Variations include: a circumferential electrode for independent vertical extension or diameter contraction, and/or sectored internal electrodes.

Small deflection expressions for the lateral tip translation can be found in Chen (1992). Measured in the same axis (x or y) as the applied voltage, the tip translation d is approximately

$$d_i = \frac{\sqrt{2} d_{31} L^2}{\pi D h} v_i \quad i = x, y \tag{3.1}$$

where d_i is the (x or y axis) deflection, d_{31} is the piezoelectric strain constant, L is the length of the tube, D is the outside diameter, h is the tube thickness, and v_i is the (x or y axis) electrode voltage. Tip deflection can be doubled by applying an equal and opposite voltage to electrodes in the same axis.

Vertical elongation due to a voltage applied on all four quadrants (or the internal electrode) is approximately

$$\Delta L = \frac{d_{31} L}{h} v. \tag{3.2}$$

where ΔL is the change in length (Chen 1992).

Although the statics and dynamics of piezoelectric tubes are inherently nonlinear and three-dimensional, when the tube has a large length/diameter ratio, the motion can be simplified. In particular, with small deflections, the vertical excursion and

tilting due to lateral deflection can often be neglected. Although there has been some recent effort to consider the coupling from lateral to vertical directions, tubes are generally designed to minimize such effects. Other design considerations are the deflection sensitivity and maximum deflection; both of which are also maximized by a large length to diameter ratio.

A consequence of designing tubes with large length/diameter ratios is low mechanical resonance frequency. This has been a fundamental problem since the inception of piezoelectric tube scanners. A lightly damped low-frequency mechanical resonance severely limits the maximum achievable scan frequency. A triangular scan rate of around 1 % of the first mechanical resonance frequency is usually assumed to be the upper limit in precision scanning applications.

The construction and characteristics of two piezoelectric tube nanopositioners are described in Sects. 3.1.1 and 3.1.2. These devices are used in the following chapters as demonstration apparatus.

3.1.1 63 mm Piezoelectric Tube

The 63 mm tube is pictured in Fig. 3.1b, c. To protect the tube and provide some immunity from environmental noise, the tube is housed in a removable aluminium shield. During assembly, the shield also serves as a jig to ensure the tube is both vertical and properly aligned while gluing. To allow displacement measurements, a polished, hollow aluminum cube 8 mm square (1.5 g in mass) is glued to the tube tip. The capacitive displacement sensor is an ADE Tech 4810 Gaging Module and 2804 capacitive sensor with sensitivity 100 mV/μm over a range of ±100 μm and bandwidth of 10 kHz.

The tube was manufactured by Boston PiezoOptics from high density PZT-5H piezoelectric ceramic. Relevant physical dimensions can be found in Fig. 3.2a. Four equally spaced quadrant electrodes are deposited around the tube circumference.

The displacement frequency response, measured using an HP 35670A spectrum analyzer, is plotted in Fig. 3.3a. The free response has a first resonance at 850 Hz and a static sensitivity of 171 nm per volt. To evaluate performance robustness in the following sections, a worst-case mass of 1.5 g is affixed to the top cube surface. The additional mass reduces the resonance frequency by 110 Hz or 13 %.

In most situations, only the first resonance frequency of a nanopositioner is considered. However, if an aggressive control strategy is employed, the higher order dynamics can also be of importance. For example, if the first resonance mode is effectively damped, the stability margins of a controller will most likely be limited by the second or third resonance mode. An understanding of higher order dynamics also quantifies the amount of modeling error in truncated or low-order models.

In Maess et al. (2008), a detailed study of the resonance modes was conducted using Finite Element Analysis (FEA) and experimental model analysis; parts of these results are repeated here. In Fig. 3.4 the first five bending modes with and without the sample holder are plotted. These results were confirmed experimentally by obtaining

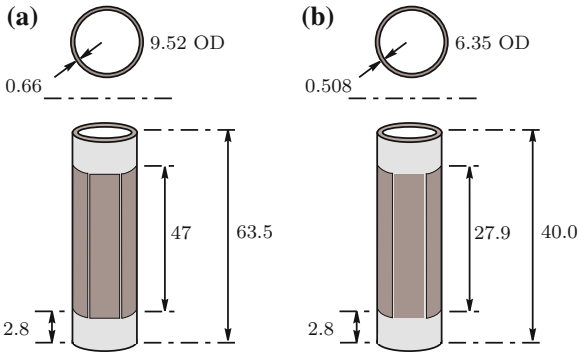


Fig. 3.2 Dimensions of the 63 and 40 mm piezoelectric tubes (in mm) described in Sects. 3.1.1 and 3.1.2

the mode shapes and frequency responses with a Polytec PI PSV300 Laser Scanning Vibrometer. The predicted resonance frequencies closely match the experimental results, which are compared in Table 3.1. The only mode predicted by FEA that was not experimentally confirmed, was the torsional mode. This mode cannot be measured using laser vibrometry as there is no velocity component normal to the surface.

The main effect of the sample holder on the mode shapes is to add mass and restrict the circular bending modes. In Fig. 3.1b, three longitudinal bending modes occur before the circular bending mode. Thus, with a sample holder attached, the behavior of the piezoelectric tube is similar to a cantilever beam. The only major difference is the longitudinal extension mode at 9.43 kHz.

Another observation from the modal analysis is that only bending motion occurs below 5 kHz. Thus, the scanner can actually be operated above the first and second resonance frequency. However, at frequencies above 5 kHz, the torsional mode is excited, so attempts to achieve lateral motion at this frequency will result in rotation of the sample holder, which is undesirable.

3.1.2 40 mm Piezoelectric Tube Nanopositioner

A second, smaller tube, is also used for experimental demonstration. This tube was also manufactured by Boston Piezo-Optics. Physical dimensions are listed in Fig. 3.2. The first resonance frequency and lateral displacement sensitivity is 1,088 Hz and 5.7 μm/μC. The frequency response is plotted in Fig. 3.3b.

A hollow 10 × 10 × 10 mm aluminum cube is glued to the tube tip to allow displacement measurement using the same capacitive sensor described in the previous section.

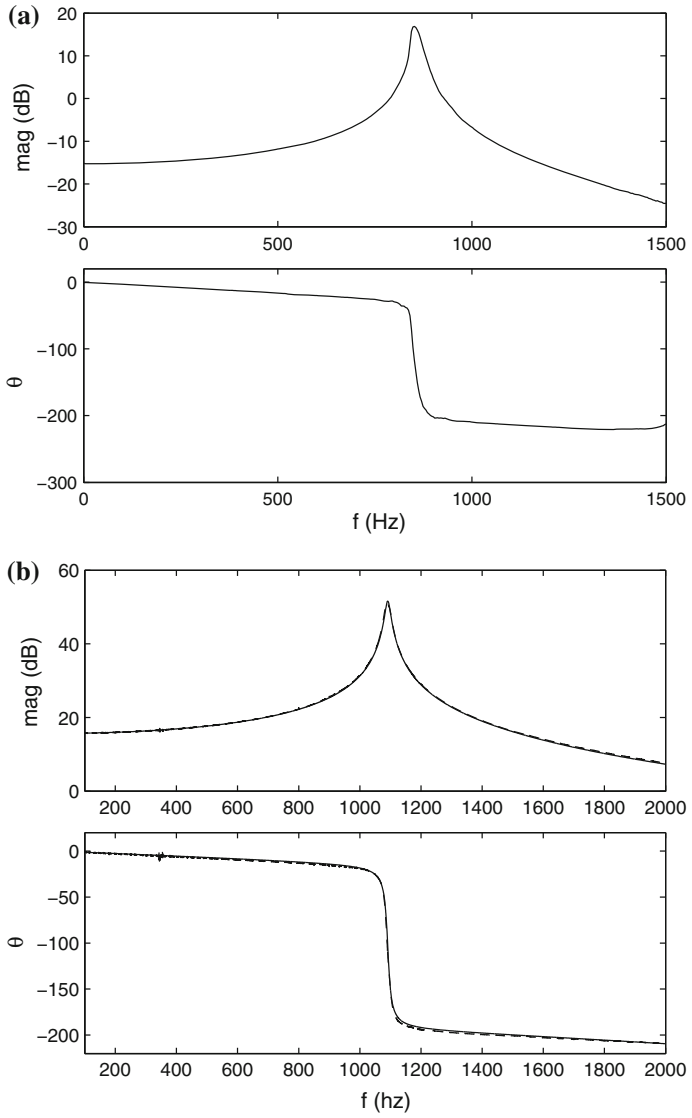


Fig. 3.3 The frequency response of the 63 mm tube is measured from the applied voltage to the tip displacement (in μm/V). The frequency response of the 40 mm tube is measured from applied charge to the tip displacement (in μm/μC)

3.2 Piezoelectric Stack Nanopositioners

Nanopositioning stages constructed from stack actuators typically comprise of piezoelectric actuators, mechanical displacement amplifiers, and a flexure guided sample platform. Although this configuration can achieve high precision with millimeter

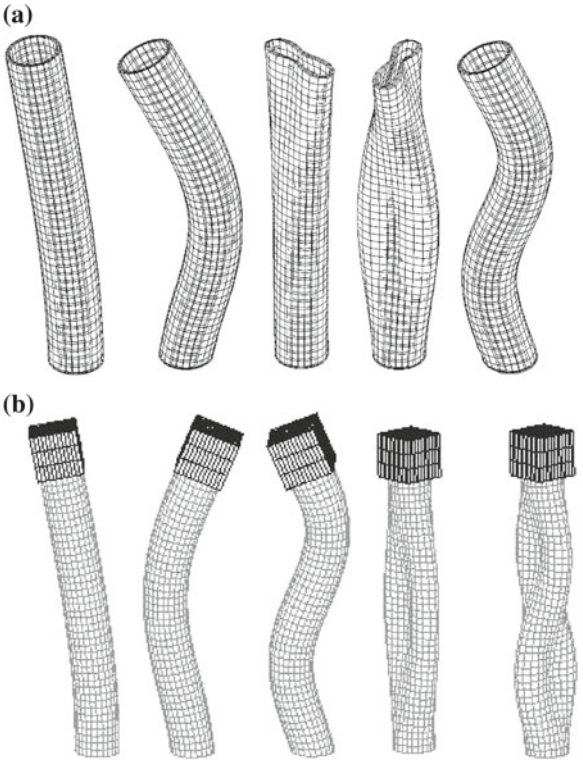


Fig. 3.4 The first five bending modes of the 63 mm tube determined by finite element analysis. **a** The first five bending modes of the 63 mm tube. The frequencies are 1.21, 6.49, 13.89, 14.50, and 15.38 kHz. **b** The first five bending modes of the 63 mm tube with a sample holder attached. The frequencies are 0.83, 4.84, 12.14, 15.19, and 16.68 kHz

Table 3.1 A comparison of the modal resonance frequencies for the 63 mm tube determined by finite element analysis and experimental modal analysis

Mode	Type	No sample holder		With sample holder	
		FEA	Exp.	FEA	Exp.
1, 2	1st long. bending	1.21	1.22	0.83	0.84
3, 4	2nd long. bending	6.49	6.55	4.89	4.84
5	1st torsional	6.97	–	5.72	–
6	1st long. extens.	11.30	11.30	9.27	9.43
7, 8	1st circ. bending	13.98	14.24	12.28	12.14
9, 10	2nd circ. bending	14.50	15.21	14.73	15.19
11, 12	3rd circ. bending	15.38	15.79	16.30	16.68

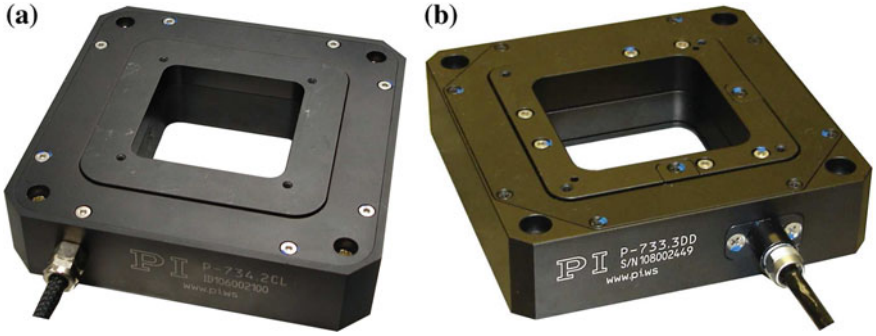


Fig. 3.5 The Physik Instrumente P-734 (**a**) and P-733.3DD (**b**) nanopositioners, with positioning ranges of 100 × 100 μm and 25 × 25 × 10 μm respectively. Both these devices are fitted with capacitive position sensors

range, the internal displacement amplifiers, large piezoelectric stacks, and platform mass contribute to a low mechanical resonance frequency. An example of such a stage is the Physik Intrumente P-734, which is shown in Fig. 3.5a and described in the following section.

3.2.1 Phyisk Instrumente P-734 Nanopositioner

A typical example of a two-axis lateral nanopositioner is the P-734 available from Phyisk Instrumente, which is shown in Fig. 3.5a. This stage has a range of 100 microns, but a resonance frequency of only 420 Hz. The frequency response of one axis is plotted in Fig. 3.6. The position is measured with a capacitive sensor, which is fitted to both axes; the accompanying electronics provides a full scale output of 6.7 V at 100 μm displacement.

In open loop or with integral control, the mechanical resonance of the P-734 limits the maximum scan frequency to 5 Hz or less with an integral controller. In the following chapters, feedback and feedforward techniques are discussed that alleviate this limitation. The only remaining limitations should be the physical limitations imposed by the mechanics of the positioner and amplifier electronics. These limitations include the maximum tensile load of the actuators and the maximum slew-rate and current limit of the amplifier, which are discussed in Chaps. 4 and 14.

3.2.2 Phyisk Instrumente P-733.3DD Nanopositioner

The P-733.3DD nanopositioner is a three-axis positioner. It has two lateral axes with a range of 30 μm and a vertical axis with a range of 10 μm. The P-733.3DD

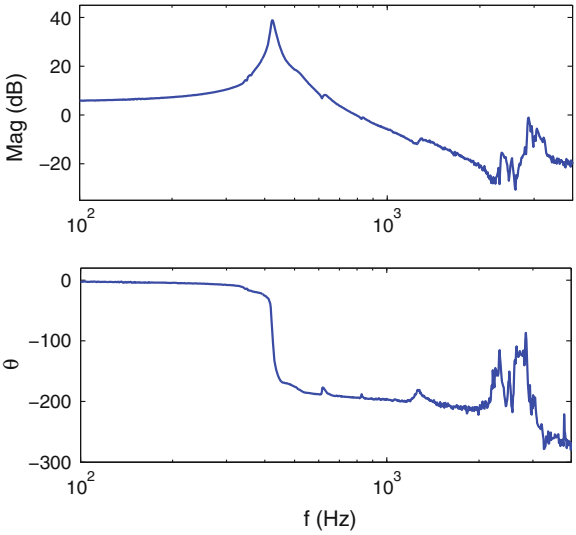


Fig. 3.6 Frequency response of the Physik Instrumente P-734 nanopositioner, measured from the applied voltage to the resulting displacement (in μm/V)

nanopositioner is a direct drive device, that is, the platform is directly connected to a stack actuator; there are no mechanical displacement amplifiers. This arrangement results in a smaller travel range, but higher stiffness and thus higher resonance frequencies, as plotted in Figs. 3.6 and 3.7. The physical properties of the P-733.3DD stage are compared to the P-734 stage in Table 3.2.

3.2.3 Vertical Nanopositioners

Vertical nanopositioners are designed to translate a load vertically over ranges of between 10 and 500 μm. They are typically used in optical microscopy for autofocusing, eliminating focus drift, and confocal microscopy where the sample or objective is scanned through a range of focal planes. Due to the applications in microscopy, vertical nanopositioners are commonly apertured to allow illumination or imaging from above and below.

Two examples of vertical nanopositioners are shown in Fig. 3.8. The Queensgate NPS-Z-500A is constructed from aluminum and titanium and has a range of 500 μm, an unloaded resonance frequency of 200 Hz, and a maximum force of 20 N. It is designed for applications including interferometry and adaptive optics. The Mad City Labs Nano-Z100 is constructed from aluminum and has a range of 100 μm, an unloaded resonance frequency of 600 Hz, and a maximum payload of 0.5 kg. The apertured design is suited to applications in microscopy.

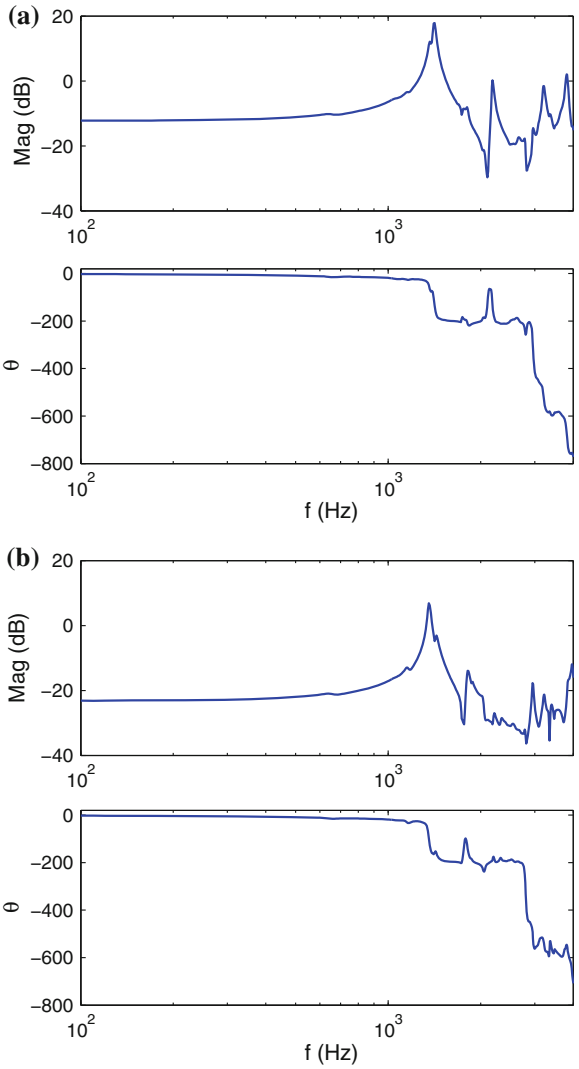


Fig. 3.7 Frequency response of the Physik Instrumente P-733.3DD nanopositioner, measured from the applied voltage to the resulting displacement (in μm/V). **a** Lateral (Y) axis. **b** Vertical (Z) axis

3.2.4 Rotational Nanopositioners

Rotational nanopositioners do not translate the sample platform, but create a rotation around the vertical axis. They are used in applications such as fiber alignment, beam steering, beam alignment, and crystallography. An example of a rotational stage is

Table 3.2 Comparison of the P-734 and P-733.3DD physical properties

	P-734	P-733.3DD
Travel range (XYZ)	100 × 100 μm	33 × 33 × 14 μm
Sensors	Capacitive	Capacitive
Stiffness (XYZ)	3, 3 N/μm	4, 4, 10 N/μm
Max pushing force	300 N	300 N
Max pulling force	100 N	100 N
Max normal load	2 kg	2 kg
Capacitance (XYZ)	6.0, 6.0 μF	6.0, 6.0, 4.4 μF
Unloaded resonance freq.	500 Hz	1.2 kHz
Resonance freq. 200 g load	350 Hz	530 Hz
Temperature range	−20–80 °C	−20–80 °C
Construction material	Aluminum	aluminum

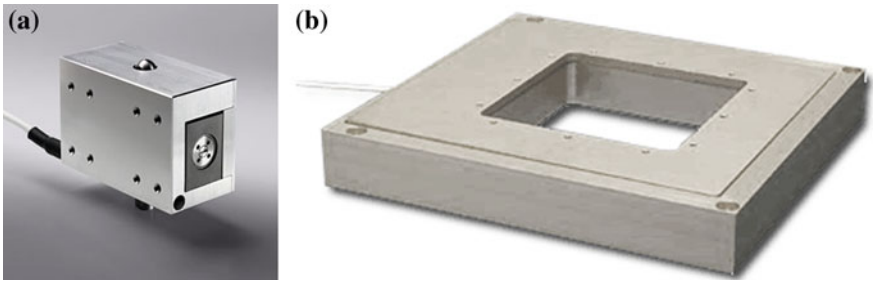


Fig. 3.8 Vertical nanopositioners. The NPS-Z-500A has a range of 500 μm and a resonance frequency of 200 Hz. The Nano-Z100 has a 100 μm range and 600 Hz resonance frequency. **a** Queensgate NPS-Z-500A. **b** Mad city Labs Nano-Z100

Fig. 3.9 Rotational nanopositioner. The Piezosystem Jena Rotor 10 has a range of 11 mrad (0.63°) and a resonance frequency of 500 Hz

the Piezosystem Jena Rotor 10 pictured in Fig. 3.9. This stage has a range of 11 mrad (0.63°) and a resonance frequency of 500 Hz.

Fig. 3.10 The Attocube ANSxy50 is designed for ultra low temperature (10 mK), high magnetic field (31 T) and ultra high vacuum (5×10^{-11} mbar). The scanning range is 30 $\times$ 30 μm at room temperature with a bandwidth of 100 Hz

3.2.5 Low Temperature and UHV Nanopositioners

Nanopositioning systems may be required to operate in extreme environments including low temperature, ultra high vacuum, and high magnetic field. It is typical that such requirements may be encountered simultaneously, for example in surface science, nanofabrication, scanning probe microscopy, and scanning beam microscopy. A further complication in such applications is the small available volume, typically a 1-inch or 2-inch footprint.

The Attocube ANSxy50 shown in Fig. 3.10 is an example of a two-axis nanopositioner that is compatible with ultra low temperature (10 mK), high magnetic field (31 T), and ultra high vacuum (5×10^{-11} mbar). The scanning range is 30 μm $\times$ 30 μm at room temperature with a bandwidth of 100 Hz.

3.2.6 Tilting Nanopositioners

Tilting nanopositioners are primarily used in beam steering and beam alignment applications. In these applications, a mirror is mounted on the moving surface, which rotates about the lateral x and y axis of the positioner.

Two examples of tilting nanopositioners are shown in Fig. 3.11. The nPoint RXY14-254 has a maximum tilting angle of 14 mrad (0.8 degrees) and an unloaded resonance frequency of 1 kHz. The Queensgate NPS-Theta-Gamma-2B has a maximum tilting angle of 2 mrad (0.11 degrees) with a resonance frequency of 1 kHz.

3.2.7 Optical Objective Nanopositioners

Optical objective nanopositioners are designed to translate a microscope objective along the optical axis. This allows the objective to be scanned through a range of focal planes in confocal microscopy. Objective nanopositioners are also used in standard microscopy for high-speed and fine resolution auto-focusing. Two examples of objective nanopositioners are shown in Fig. 3.12. The Queensgate OSM-Z-100B has a

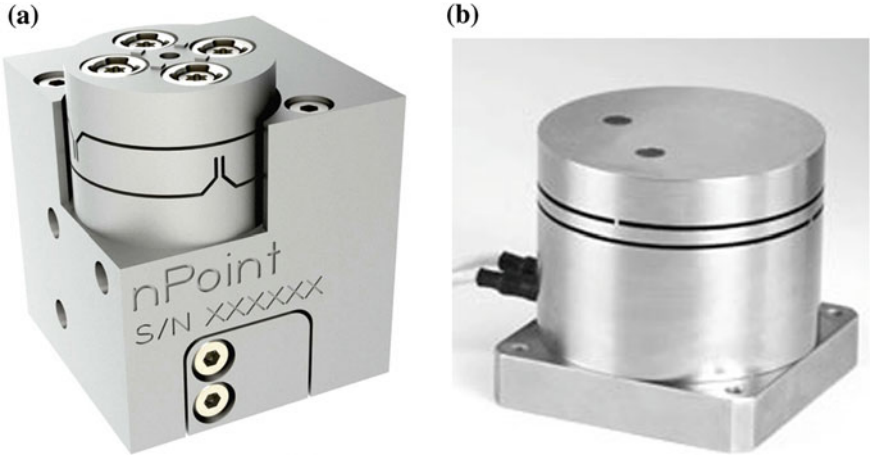


Fig. 3.11 Tilting nanopositioners. The nPoint RXY14-254 has a maximum tilting angle of 14 mrad (0.8 degrees) and an unloaded resonance frequency of 1 kHz. The Queensgate NPS-Theta-Gamma-2B has a maximum tilting angle of 2 mrad (0.11 degrees) with a resonance frequency of 1 kHz. **a** nPoint RXY14-254 **b** Queensgate NPS-$\theta\gamma$-2B

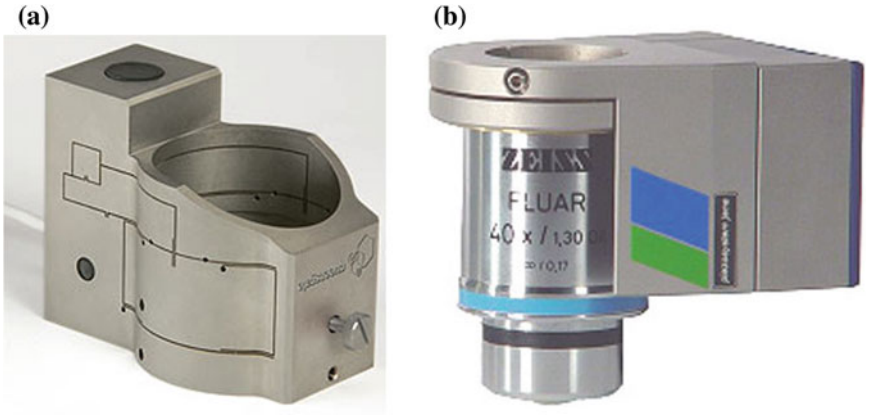


Fig. 3.12 Objective lens nanopositioners. The Queensgate OSM-Z-100B has a range of 100 μm and an unloaded resonance frequency of 600 Hz. The Piezosystems Jena MIPOS 500 has a 500 μm range and an unloaded resonance frequency of 230 Hz. **a** Queensgate OSM-Z-100B **b** Piezo Systems Jena MIPOS500

range of 100 μm, an unloaded resonance frequency of 600 Hz and maximum payload of 0.6 kg. The Piezosystems Jena MIPOS 500 has a 500 μm range, an unloaded resonance frequency of 230 Hz, and a maximum payload of 0.5 kg.

References

Binnig G, Smith DPE (1986) Single-tube three-dimensional scanner for scanning tunneling microscopy. Rev Sci Instrum 57(8):1688–1689

Chen CJ (1992) Electromechanical deflections of piezoelectric tubes with quartered electrodes. Appl Phy Lett 60(1):132–134

Croft D, McAllister D, Devasia S (1998) High-speed scanning of piezo-probes for nano-fabrication. Trans ASME, J Manuf Sci Technol 120:617–622

Gao W, Hocken RJ, Patten JA, Lovingood J, Lucca DA (2000) Construction and testing of a nanomachining instrument. Precis Eng 24(4):320–328

Maess J, Fleming AJ, Allgwer F (2008) Simulation of dynamics-coupling in piezoelectric tube scanners by reduced order finite element models. Rev Sci Instrum 79:015 105-1-015 105–9

Meyer E, Hug HJ, Bennewitz R (2004) Scanning probe microscopy. The lab on a tip. Springer, Heidelberg

Chapter 4
Mechanical Design: Flexure-Based Nanopositioners

The dynamic performance of a nanopositioning system depends on its mechanical resonance frequency, damping, the type of controller used, sensor bandwidth, and associated data acquisition hardware. Recently, speed has become a critical issue in many nanopositioning applications, such as video-rate AFM and high-throughput nanomanufacturing. One of the key limitations in speed is the system's mechanical resonance. As a result, recent efforts have focused on designing the mechanical system to have the highest possible mechanical resonance while maintaining acceptable range of motion. In this chapter, an overview of mechanical design is presented, where the emphasis is on flexure-guided nanopositioning stages for high-speed nanopositioning. The discussions will focus on systems driven by piezoelectric actuators such as plate-stacks, which are readily available from a number of commercial suppliers.

4.1 Introduction

The performance of a nanopositioning system is often dictated by the quality of the mechanical design (Yong et al. 2012). In fact, good mechanical design will minimize most position errors and improve overall accuracy. Poor mechanical design, on the other hand, can lead to more errors than the issues associated with the electronics and other components. Additionally, with good mechanical control systems can be designed to take advantages of the physical characteristics of the positioning stage. The important factors to consider for good mechanical design include: stability of shape and dimension of the positioning stage as a function of temperature; mechanical stiffness; and strength, although strength may not matter in most cases and thus will not be discussed. Cost and manufacturability are also two important factors, especially when it comes to commercialization.

In terms of speed, high mechanical resonance, that is a *stiff* mechanical design, is desired. Traditional nanopositioning designs employ relatively *flexible* piezoactuators and flexure-based mechanisms. In these designs, the lowest mechanical

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_4, © Springer International Publishing Switzerland 2014

resonance is typically less than 1 kHz for a lateral travel range of 10–100 μm. The first mechanical resonance is one of the major limiting factors in speed (Ando et al. 2002; Schitter et al. 2007). Command signals such as triangle waves at 1/10 to 1/100th the first mechanical resonance can excite dynamics that cause significant output oscillation and distortion. A positioning stage's resonance is related to its effective mass, m_{eff}, and stiffness, k_{eff}. Although the effective mass can be reduced to achieve the same effect, due to robustness issues this is not a recommended approach. In particular, the design must be able to accommodate variations in the mass of a sample tray, for example, without significant affect on the mechanical resonance. In designs involving a mechanical amplification factor A_f, the stiffness k_{eff} is given by

$$k_{\text{eff}} = \frac{k_p}{A_f^2}, \tag{4.1}$$

where k_p is the stiffness of the piezoactuator. Reasonable k_{eff} is achieve when A_f is less than five (Hicks et al. 1997).

High-speed nanopositioning is needed in many applications, including video-rate SPM. For instance, the dynamic behavior of micro- and nano-scale processes, such as the movement of biological cells, DNA, and molecules, occur at time scales much faster than the scanning capabilities of conventional SPMs, for example an AFM. Therefore, AFMs capable of high-speed operation are required to observe these processes in real-time (Guthold et al. 1999). High-throughput, probe-based nanomanufacturing is also another area where high-speed positioning of the probe tip is needed. Primarily a serial technique, the total process time of probe-based fabrication is proportional to the number of desired features for a given linear scan rate (Snow et al. 1997). In this respect, a high-throughput positioning stage can drastically reduce manufacturing time.

The mechanical design process first begins by considering the environment in which the stage will be operated. To illustrate the design process, the steps taken to design an example serial-kinematic high-speed multiaxis nanopositioner is presented.

4.2 Operating Environment

At the macroscopic level, small changes in the surrounding environment, such as temperature, humidity, and floor motion, usually go undetected. However, at the micron to nanoscale, the effects may be significant. Micro- and nanopositioning stages can be found in many environments, including research laboratories, ultra-high vacuum chambers, precision machine shops, and environmental scanning electron microscopes (Muller et al. 2007; Samara-Ratna et al. 2007). Certain locations may be well-controlled in terms of temperature, humidity, and external mechanical vibrations. In such areas, there is minimal concern that variations in the environmental conditions

will cause a deterioration in the operating performance of the stage. Instead, performance degradation will likely to occur due to mechanical fatigue and thermal issues, the former being a slow process. For example, one major concern is the self-heating of the piezoelectric actuator at high operating frequencies. The high temperature can affect the repeatability, precision, and life of the stage. Thermally induced stress can cause mechanical failure in flexure mechanisms, joints, and glue layers. The heat generation in a piezoelectric material is attributed to hysteretic losses in the material (Devos et al. 2008). An estimate of the thermal active power, P_a, generated in the actuator due to a sinusoidal input signal is given by Physik Instrumente (2009)

$$P_a \approx \frac{\pi}{4} \tan(\delta) f C V_{pp}^2, \tag{4.2}$$

where f is the frequency of the input signal (in Hz), C is actuator's nominal capacitance, V_{pp} is the peak-to-peak voltage of the input signal, and $\tan(\delta)$ is the dielectric (loss) factor. Under large signal conditions, as much as 12 % of the electrical power used to drive the actuator is converted to heat. The generated heat can limit the actuator's performance, and if not properly isolated, nearby samples and components can be affected. Therefore, best cooling practices should be employed for the actuator and drive electronics. Furthermore, it has been shown that the optimal operating frequency for minimal heat dissipation is close to the resonance frequency for standard piezoelectric materials (Devos et al. 2008).

But for environments where significant fluctuations in operating conditions exist, special considerations should be taken during the design process. For instance, a well-sealed and water-resistant enclosure is recommended for devices, which operate in areas prone to contact with liquids such as coolants, water, oils, and corrosive elements. An enclosure also prevents conductive particles such as fine metal shavings from degrading the piezoceramic and causing short circuits. Stainless steel is commonly used as an enclosure material (Physik Instrumente 2009). In some cases, a protective coating can be sprayed over the actuator and stage assembly to provide additional protection from the environment.

Systems which operate in areas prone to large temperature fluctuations and high temperatures should be closely monitored and protected against. Thermocouple sensors can be used to measure the temperature of critical or nearby components to ensure that excessive heating does not occur. An environment chamber in which the temperature can be closely controlled may be required.

Positioning stages used in areas where a significant level of external mechanical vibration exists should be properly isolated. The lack of isolation will allow the transmission of mechanical disturbances, which can excite the resonances of the positioning system, therefore affecting accuracy. The frequency of vibration of a tall building due to wind is on the order of 1 to 50 Hz. Nearby machinery and equipment can vibrate up to several hundred Hertz. A survey of the vibration level of the environment should be done to determine whether specialized foundations and vibration isolation platforms are required. Additionally, acoustic vibrations in the air should also be considered and protected against. A simple acoustic chamber can

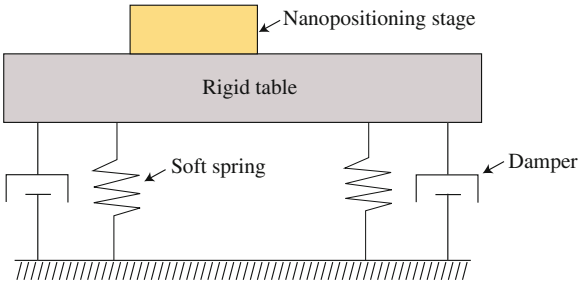


Fig. 4.1 Vibration isolation table consisting of a rigid table supported by soft springs and dampers

be constructed using readily available acoustic foam with a sufficiently high noise reduction coefficient (NRC).

The goal of using a vibration isolation platform is to minimize the effect of relative motion between any two or more components. Building vibration and vibration from other sources such as a fan motor in a nearby desktop computer can transmit through workbench mounting points to the positioning stage. The isolation table basically consists of a rigid table supported by relatively soft spring legs with damping as shown in Fig. 4.1. The *quality* of a vibration isolation platform (optic table) that is schematically depicted in Fig. 4.1 is quantified in terms of its compliance $\mathcal{C}$, the ratio of the excited vibrational amplitude x to the magnitude of the forcing vibration F. The lower the compliance, the better the table upon which the nanopositioning stage rests. In other words, the table should have zero or minimal response due to an applied force or vibration. An ideal table is a rigid body, which does not resonate, and it exhibits a compliance given by

$$\mathcal{C} \triangleq \frac{x}{F} = \frac{1}{f^2}, \tag{4.3}$$

where f is the frequency. Therefore, the quality of any table should be compared to the compliance of the ideal case (4.3). To understand the effects of table mass, leg stiffness, and damping, consider the following one degree-of-freedom compliance relationship for the vibration table shown in Fig. 4.1,

$$\mathcal{C} = \frac{1/k}{\sqrt{(1 - f^2/f_n^2) + (2\zeta f/fn)^2}}, \tag{4.4}$$

where k is the stiffness, ζ is the damping coefficient, and f_n is the natural resonance frequency. At low frequencies the compliance is determined by the stiffness of the table, therefore, the higher the stiffness the lower the compliance. At resonance, however, the compliance is dominated by the amount of damping, thus damping is critical in reducing the transmission of vibration from the environment to the table near the table's resonance. Finally, the mass of the table only makes a significant

contribution at high frequencies. Ideally, the resonance frequency of the table itself should be made as low as possible to avoid excitation caused by building vibration.

4.3 Methods for Actuation

Aside from using piezoelectric actuators for creating displacement, other possible candidates for include electrostrictive and magnetostrictive actuators. Like piezoelectric materials, electrostrictives (Damjanovic and Newnham 1992) can convert electrical to mechanical energy and vice versa. The strain to voltage relationship for an electrostrictive actuator is governed by

$$\epsilon = cV^2, \tag{4.5}$$

where c is a constant. The achievable strain can be as much as 0.15 %, and one major advantage is they exhibit much lower hysteresis compared to standard PZT. However, optimum performance can only be achieved over a narrow electric field and temperature range.

Magnetostrictive materials (Stillesjo et al. 1998; Tan and Baras 2004), which convert magnetic to mechanical energy, offer relatively linear behavior within the range of 0.1 % strain. The governing equations for these materials are similar to those for piezoelectrics. These materials have been applied to the development of micropositioning systems (Tsodikov and Rakhovsky 1998).

Shape memory alloy (SMA), for example nickel-titanium, is a *active material* whereby a change in temperature causes a change in the atomic crystal structure of the alloy. As a result, the material undergoes shape change with achievable strain as high as 8 % when the material transforms between the martensite phase (monoclinic at low temperature) and the austenite phase (cubic at high temperature) (Waram 1993). This unique behavior can be exploited to create SMA-based actuators (or positioners), and compared to piezoelectric actuators, SMAs offer relatively large strain and high strength-to-weight ratio (e.g., recovery stress $>$ 500 MPa). Unfortunately, their slow response times and significant hysteresis behavior limit their application in high speed nanopositioning.

Compliant microelectromechanical systems (MEMS) for microscale positioning can be achieved using electrostatic, thermal, piezoelectric, pneumatic, as well as magnetostrictive and electromagnetic actuation (Liu 2006). Such devices are created using standard or specialized MEMS fabrication techniques. A graphical performance chart has been developed to provide a quantitative comparison of MEMS-based actuators in terms of maximum force, displacement capability, resolution, and natural frequency (Bell et al. 2005). Additionally, a detailed review of actuators for micro- and nano-positioners can be found in Hubbard et al. (2006), Sahu et al. (2010). Performance of the actuators is delineated based on range, resolution, footprint, output force, speed of response (bandwidth), and electrical drive considerations. It is worth noting that electrothermal and electrostatic actuators are the most widely used

actuators for nanoscale applications. This is because of their straightforward integration with standard MEMS-based fabrication processes, relatively small footprint (<1 mm^2) and design simplicity. More specifically, electrothermal actuators operate on the principle of Joule heating and differential thermal expansion (Liu 2006; Bechtold et al. 2005). In particular, an electrical closed-loop is formed by designing the actuating mechanism to consist of a 'hot' and 'cold' arm. The difference in the heating of each arm induces strain, and thus mechanical deformation. Typically, electrothermal actuators are suitable for large deflection (up to 20 μm), with output force in the micro- to milli-Newton range (10 μN to 10 mN), and operating voltage well below 15 V. These actuators exhibit the smallest footprints (<1 mm^2) making them suitable for a wide variety of nanoscale applications. However, the high temperature (200–600 °C) may be undesirable for certain temperature sensitive applications. An extensive review of electrothermal actuators can be found in (Geisberger and Sarkar 2006). MEMS-based electrostatic actuators operate on the principle of Coulomb attraction due to application of a bias voltage between two plates (moving and fixed) (Hubbard et al. 2006; Geisberger and Sarkar 2006). For the simplest parallel-plate configuration, the capacitance C gives a measure of the stored energy, which is a function of the plate area A, permittivity of the medium ϵ_o, and distance between the plates d. In general, the output force is a nonlinear function of the gap between the plates. The operating voltage ranges from 20–100 V. Electrostatic microactuators provide higher positioning resolution (<1.5 nm) and faster response (micro-second range) as compared to electrothermal actuators. Because of their straightforward fabrication, small footprint (≈ 1 mm^2), and low power consumption they find potential use at the nanoscale. However, they are not preferable for applications such as in-situ manipulation in electron microscopes as electric fields due to high voltage may interfere with the imaging electron beam.

4.4 Flexure Hinges

4.4.1 Introduction

Although piezoactuators are capable of sub-nanometer positioning resolution, they provide limited travel range. A modest 10-mm long piezo-stack actuator (Noliac SCMA-P7) at full drive voltage of 200 V extends a maximum of 11 μm (unconstrained). Larger displacements can be achieved with longer actuators or mechanical amplifiers. However, it is pointed out that these options come at a cost of lower mechanical bandwidth (that is, resonance frequency) due to the reduction of effective stiffness. In fact, the first resonance frequency is inversely proportional to an actuator's maximum stroke. To illustrate, consider a fixed-free plate-stack piezoactuator with constant rectangular cross-section. The extension of the actuator along its length (longitudinal displacement) is given by

Fig. 4.2 A gripper with four flexure hinges and four rigid links

$$\delta \approx d_{31} L U, \tag{4.6}$$

where d_{31} is the strain coefficient perpendicular to the polarization direction, L is the length of the actuator, and U is the electric field. From vibrations, the frequency of the first longitudinal mode of the stack actuator can be expressed as Inman (2001)

$$f = \frac{\pi}{L}\sqrt{\frac{E}{\rho}}, \tag{4.7}$$

where E is the elastic modulus and ρ is the density. By eliminating the dependence on the actuator length L in Eq. (4.7) using (4.6), the frequency of the first longitudinal mode is

$$f = \frac{\pi d_{33} U}{\delta_z}\sqrt{\frac{E}{\rho}} \propto \frac{1}{\delta_z}. \tag{4.8}$$

Therefore, the first resonance frequency is inversely proportional to the actuator's maximum displacement δ. Higher bandwidth is achieved by using more compact piezoactuators, but the achievable travel range is reduced.

Flexure hinges are commonly employed in the design of effective mechanical amplifiers for macro as well as MEMS-based devices. Figure 4.2 shows an application of flexure hinges. The gripper mechanism consists of four flexures connecting four rigid links. The use of flexure hinges over traditional rotational joints enables the gripper to be easily manufactured as one part. For nanopositioning, as illustrated in Fig. 4.3a1,a2, for a given actuator stroke δ, a flexure-based mechanical amplifier provides a scaled output of $a\delta$, where a is primarily a function of the geometry of the mechanical amplifier. Flexure hinges are also commonly employed to guide the motion of sample stages to minimize parasitic motion, as well as to increase the stiffness of an actuator along off-axis or out-of-plane directions to improve mechanical resonances. Figure 4.3b1,b2 show an example of a single-axis positioning stage for displacing a mass m. The flexure hinges, one on each side of the mass, serve to guide the motion of the mass along a straight path.

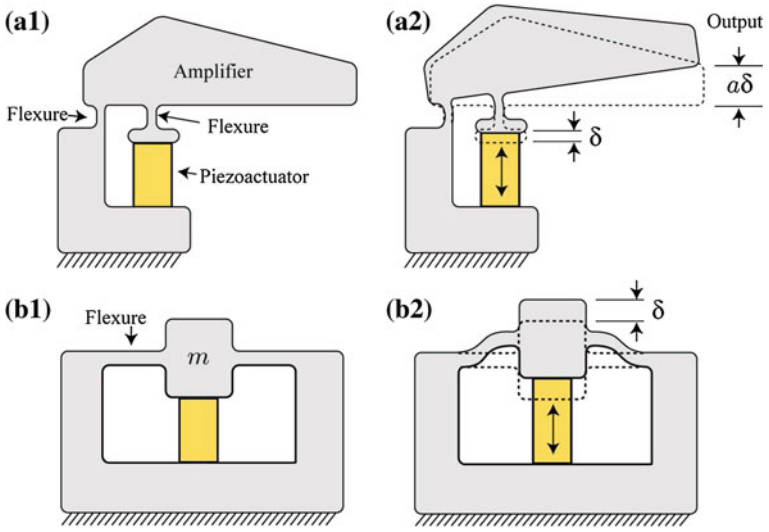


Fig. 4.3 Application of flexure hinges: **a1** and **a2** mechanical amplifier; **b1** and **b2** motion guiding to minimize runout

As indicated by the examples in Fig. 4.3, a flexure hinge is simply a thin elastic member that connects two rigid bodies and provides limited relative rotational motion through bending or flexing. The major distinction between a classical mechanical joint, such as a rotational bearing, and a flexure hinge is that the center of rotation for the two connected members in the former are collocated, whereas for the flexure hinge the rotation is noncollocated. Interested readers will find detailed discussions on flexure design and compliant mechanisms in Smith (2000), Howell (2001), Lobontiu (2003). For nanopositioning systems, flexure hinges are far more compact compared to traditional mechanical hinges. They are invaluable because there is no friction loss, need for lubrication, or hysteresis effects.

4.4.2 Types of Flexures

Flexure hinges can be designed for one to multiple degrees-of-freedom motion as illustrated in Fig. 4.4. For example a single axis flexure hinge (Fig. 4.4a) is used for planar motion, where as two- and multi-axis flexure hinges (Fig. 4.4b, c) are ideally suited for three-dimensional motion. Standard milling and electrical discharge machining are used to create the one and two degrees-of-freedom flexures, where as a turning operation is used to create the multiaxis flexure hinge shown in Fig. 4.4c.

Commonly used flexure hinge designs are shown in Fig. 4.5. The corner-filleted design offers more evenly distributed stresses compared to the basic design of Fig. 4.5a.

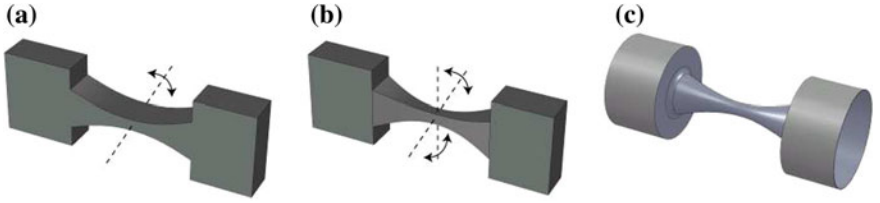


Fig. 4.4 Flexure degrees of freedom: **a** one, **b** two, and **c** multiple degrees-of-freedom

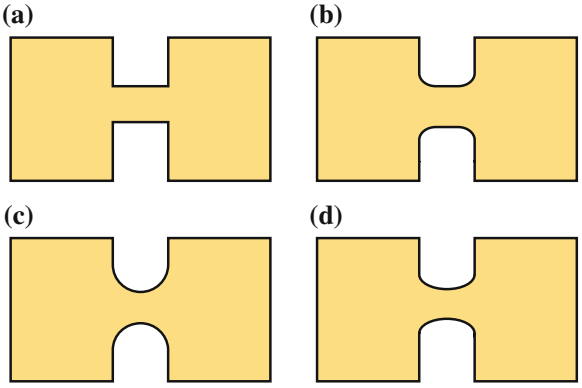


Fig. 4.5 Different types of flexure hinges: **a** basic, **b** filleted, **c** circular, and **d** elliptical

The flexure geometries shown in Fig. 4.4 are often of a monolithic design; that is, milled, turned, or specially machined from a solid block of material. However, flexure hinges can be made by assembling thin members with rigid members using fasteners or through bonding. Examples of multiaxis nanopositioning stages made from these two types of flexures are shown in Fig. 4.6. The main advantage of making a flexure from individual parts is it can be fabricated with standard milling and turning processes. The disadvantage of assembled flexures is performance. Because fasteners and adhesives are used, inconsistencies in the assembly process and boundary conditions can have a drastic effect on the flexures static and dynamic performance. On the other hand, monolithic designs offer more predictable and repeatable performance. However, monolithic designs, especially for flexure with extremely thin dimensions, may require specialized machining processes such as wire electrical discharge machining or MEMS fabrication techniques.

4.4.3 Flexure Hinge Compliance Equations

Flexure hinges are frequently designed to operate over small displacements and angles of rotation. For homogenous linear elastic and isotropic materials, the closed-form solution for the deformation of a flexure hinge can be derived using, for example,

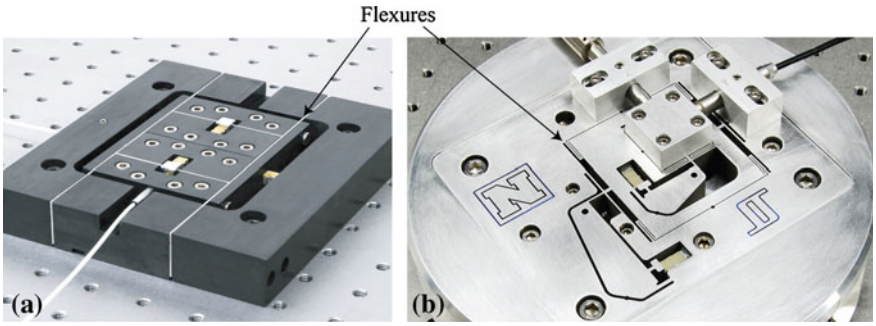


Fig. 4.6 Examples of flexures made by **a** assembling thin members with rigid blocks (Leang and Fleming 2009) and **b** monolithic design fabricated by wire electrical discharge machining

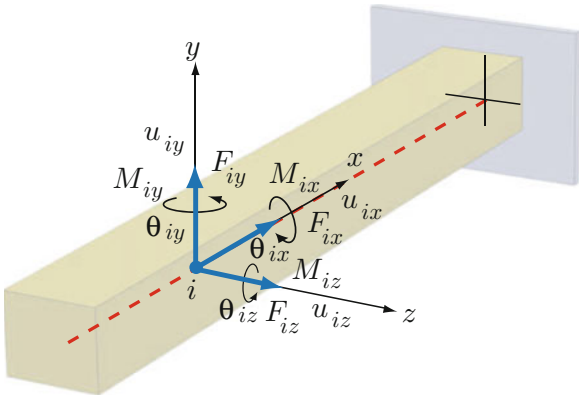


Fig. 4.7 Generic flexure member

Castigliano's displacement (second) theorem (Lobontiu 2003). Consider as an example, the generic flexure member, which is shown as a long slender beam, in Fig. 4.7.

Castigliano's displacement theorem enables the calculation of deformations (or rotations) of elastic bodies at a specific point i under external loading, moments, or support reactions acting at that location. The linear and angular deformations at location i due to force F_i and moment M_i are

$$u_i = \frac{\partial E_\epsilon}{\partial F_i}, \tag{4.9}$$

$$\theta_i = \frac{\partial E_\epsilon}{\partial M_i}, \tag{4.10}$$

where E_ϵ is the strain energy. In the event that a deformation at location i is sought where there are no loads or reactions applied at that location, fictitious loads, $\hat{F}_i$ and $\hat{M}_i$, are used and the deformations are determined using Eqs. (4.9) and (4.10).

The example long slender beam in Fig. 4.7 is subjected to bending, shearing, axial load, and torsion. Therefore, the strain energy can be expressed as Lobontiu (2003)

$$E_\epsilon = E_{\epsilon,\text{ bending}} + E_{\epsilon,\text{ shearing}} + E_{\epsilon,\text{ axial}} + E_{\epsilon,\text{ torsion}}. \tag{4.11}$$

Specifically,

$$\begin{aligned} E_{\epsilon,\text{ bending}} &= E_{\epsilon,\text{ bending},y} + E_{\epsilon,\text{ bending},z}, \\ &= \int_L \frac{M_y^2}{2EI_y}\,\mathrm{d}s + \int_L \frac{M_z^2}{2EI_z}\,\mathrm{d}s; \end{aligned} \tag{4.12}$$

$$\begin{aligned} E_{\epsilon,\text{ shearing}} &= E_{\epsilon,\text{ shearing},y} + E_{\epsilon,\text{ shearing},z}, \\ &= \int_L \frac{\alpha V_y^2}{2GA}\,\mathrm{d}s + \int_L \frac{\alpha V_z^2}{2GA}\,\mathrm{d}s; \end{aligned} \tag{4.13}$$

$$\begin{aligned} E_{\epsilon,\text{ axial}} &= E_{\epsilon,\text{ axial},x}, \\ &= \int_L \frac{P_x^2}{2EA}\,\mathrm{d}s; \end{aligned} \tag{4.14}$$

$$\begin{aligned} E_{\epsilon,\text{ torsion}} &= E_{\epsilon,\text{ torsion},x}, \\ &= \int_L \frac{M_x^2}{2GJ}\,\mathrm{d}s; \end{aligned} \tag{4.15}$$

where α is a constant based on the cross-section. Equations (4.9–4.15) can be combined into the following matrix form

$$\delta_i = \mathcal{C}_i \mathbf{P}_i, \tag{4.16}$$

where $\delta_i = [u_i\ \theta_i]^T$, $\mathcal{C}_i$ is the compliance (flexibility) matrix, and $\mathbf{P}_i$ represents all the loads and moments acting at point i.

Closed-form capacity and precision for rotation solutions for many flexure hinge geometries can be found in Lobontiu (2003). In the following, a brief summary of the results for the capacity for rotation of single-axis flexure hinges with constant width and vertical profiles as shown in Fig. 4.8 are presented for convenience. These flexures are ideally suited for planar motion, and they are used extensively in the design of parallel (Schitter et al. 2007) and serial-kinematics (Leang and Fleming 2009) nanopositioning stages.

First, it is the assumed that the flexure is relatively long compared to the dimensions of its cross-section. For shorter flexure design where shearing effects must be taken into account, see results in Lobontiu (2003). Let L denote the length of the flexure and $t(x)$ be its thickness as a function of the location x. The minimum thickness of the flexure over L is given by t.

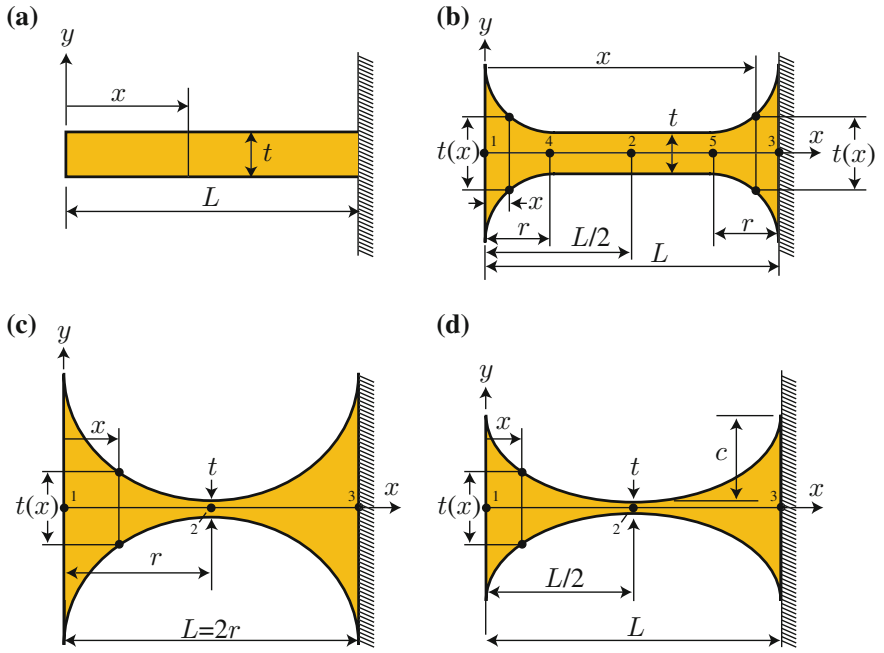


Fig. 4.8 Commonly used flexure hinges and their dimensions: **a** basic, **b** corner-filleted, **c** circular, and **d** ellipse

In these types of flexures, the loading at the free end (location 1) has the following six components:

- two bending moments, M_{1y} and M_{1z};
- two shearing forces, F_{1y} and F_{1z};
- one axial load, F_{1x}; and
- one torsional moment, M_{1x}.

The in-plane components, M_{1z}, F_{1y}, and F_{1x} are the most significant. The out-of-plane components M_{1y}, F_{1z}, and M_{1x} are generally lower in magnitude and appear due to manufacturing and assembly issues. The torsional component can be neglected. Taking these into account, the deformation equation (4.9) can be written as

$$\begin{bmatrix} \mathbf{u}_1^{ip} \\ \mathbf{u}_1^{op} \end{bmatrix} = \begin{bmatrix} \mathcal{C}_1^{ip} & 0 \\ 0 & \mathcal{C}_1^{op} \end{bmatrix} \begin{bmatrix} \mathbf{P}_1^{ip} \\ \mathbf{P}_1^{op} \end{bmatrix}, \tag{4.17}$$

where the displacement and load vectors $\mathbf{u}$ and $\mathbf{P}$, respectively, have been divided into in-plane (superscript ip) and out-of-plane (superscript op) subvectors. The subvectors are

$$\mathbf{u}_1^{ip} = \begin{bmatrix} u_{1x} \\ u_{1y} \\ \theta_{1z} \end{bmatrix}; \quad \mathbf{u}_1^{op} = \begin{bmatrix} u_{1z} \\ \theta_{1y} \end{bmatrix}; \quad \mathbf{P}_1^{ip} = \begin{bmatrix} F_{1x} \\ F_{1y} \\ M_{1z} \end{bmatrix}; \quad \mathbf{P}_1^{op} = \begin{bmatrix} F_{1z} \\ M_{1y} \end{bmatrix}. \tag{4.18}$$

The in- and out-of-plane submatrices are

$$\mathcal{C}_1^{ip} = \begin{bmatrix} C_{1,x-F_x} & 0 & 0 \\ 0 & C_{1,y-F_y} & C_{1,y-M_z} \\ 0 & C_{1,\theta_z-F_y} & C_{1,\theta_z-M_z} \end{bmatrix}; \quad \mathcal{C}_1^{op} = \begin{bmatrix} \mathcal{C}_{1,z-F_z} & \mathcal{C}_{1,z-M_y} \\ \mathcal{C}_{1,\theta_y-F_z} & \mathcal{C}_{1,\theta_y-M_y} \end{bmatrix}, \tag{4.19}$$

where $C_{1,y-M_z} = C_{1,\theta_z-F_y}$ and $C_{1,z-M_y} = C_{1,\theta_y-F_z}$. The in-plane compliance equations are

$$\begin{aligned} C_{1,x-F_x} &= \frac{1}{Ew} I_1; \quad C_{1,y-F_y} = \frac{12}{Ew} I_1 2; \\ C_{1,y-M_x} &= \frac{1}{Ew} I_3; \quad C_{1,\theta_z-M_z} = \frac{12}{Ew} I_4. \end{aligned} \tag{4.20}$$

The out-of-plane compliance equations are

$$\begin{aligned} C_{1,z-F_z} &= \frac{12}{Ew^3} I_5; \quad C_{1,z-M_y} = \frac{12}{Ew^3} I_6; \\ C_{1,\theta_y-M_y} &= \frac{12}{Ew^3} I_1 = \frac{12}{w^2} C_{1,x-F_x}. \end{aligned} \tag{4.21}$$

And finally, the integrals above are

$$\begin{aligned} I_1 &= \int_0^L \frac{1}{t(x)}\, \mathrm{d}x; \quad I_2 = \int_0^L \frac{x^2}{t^3(x)}\, \mathrm{d}x; \quad I_3 = \int_0^L \frac{x}{t^3(x)}\, \mathrm{d}x; \\ I_4 &= \int_0^L \frac{1}{t^3(x)}\, \mathrm{d}x; \quad I_5 = \int_0^L \frac{x^2}{t(x)}\, \mathrm{d}x; \quad I_6 = \int_0^L \frac{x}{t(x)}\, \mathrm{d}x. \end{aligned} \tag{4.22}$$

For a constant rectangular cross-section flexure hinge as shown in Fig. 4.8a, where the thickness is $t(x) = t$ for $0 \leq x \leq L$, the in-plane compliances are founded by solving the integrals in Eq. (4.22) and substituting the results into Eq. (4.20). The final results are

$$\begin{aligned} \mathcal{C}_{1,x-F_x} &= \frac{L}{Ewt}; \quad \mathcal{C}_{1,y-F_y} = \frac{4L^3}{Ewt^3}; \\ \mathcal{C}_{1,y-M_z} &= \frac{6L^2}{Ewt^3}; \quad \mathcal{C}_{1,\theta_z-M_z} = \frac{12L}{Ewt^3}; \\ \mathcal{C}_{1,z-F_z} &= \frac{4L^3}{Ew^3t}; \quad \mathcal{C}_{1,z-M_y} = \frac{6L^2}{Ew^3t}; \end{aligned}$$

$$\mathcal{C}_{1,\theta_y-M_y} = \frac{12L^3}{Ew^3t}. \tag{4.23}$$

For a corner-filleted flexure hinge as shown in Fig. 4.8b, where the thickness is

$$t(x) = \begin{cases} t + 2\left[r - \sqrt{x(2r-x)}\right], & x \in [0, r] \\ t, \quad x \in [r, L-r] \\ t + 2\{r - \sqrt{(L-x)[2r-(L-x)]}\}, & x \in [L-r, r], \end{cases} \tag{4.24}$$

the in-plane compliances are

$$\begin{aligned}
\mathcal{C}_{1,x-F_x} &= \frac{1}{Ew}\left[\frac{L-2r}{t} + \frac{2(2r+t)}{\sqrt{t(4r+t)}}\arctan\sqrt{1+\frac{4r}{t}} - \frac{\pi}{2}\right]; \\
\mathcal{C}_{1,y-F_y} &= \frac{3}{Ew}\left\{\frac{4(L-2r)(L^2-Lr+r^2)}{3t^3}\right. \\
&\quad + \frac{\sqrt{t(4r+t)}[-80r^4+24r^3t+8(3+2\pi)r^2t^2]}{4\sqrt{t^5(4r+t)^5}} \\
&\quad + \frac{\sqrt{t(4r+t)}[4(1+2\pi)rt^3+\pi t^4]}{4\sqrt{t^5(4r+t)^5}} \\
&\quad + \frac{(2r+t)^3(6r^2-4rt-t^2)\arctan\sqrt{1+\frac{4r}{t}}}{\sqrt{t^5(4r+t)^5}} \\
&\quad + \frac{-40r^4+8Lr^2(2r-t)+12r^3t+4(3+2\pi)r^2t^2}{2t^2(4r+t)^2} \\
&\quad + \frac{2(l+2\pi)rt^3+\frac{\pi t^4}{2}}{2t^2(4r+t)^2} + \frac{4L^2r(6r^2+4rt+t^2)}{t^2(2r+t)(4r+t)^2} \\
&\quad - \frac{(2r+t)[-24(L-r)^2r^2-8r^3t+14r^2t^2+8rt^3+t^4]}{\sqrt{t^5(4r+t)^5}} \\
&\quad \left.\times \arctan\sqrt{1+\frac{4r}{t}}\right\}; \\
\mathcal{C}_{1,y-M_z} &= -\frac{6L}{Ewt^3(2r+t)(4r+t)^2}\left\{(4r+t)[L(2r_t)(4r+t)^2\right. \\
&\quad - 4r^2(16r^2+13rt+3t^2)] + 12r^2(2r+t)^2\sqrt{t(4r+t)} \\
&\quad \left.\times \arctan\sqrt{1+\frac{4r}{t}}\right\}; \\
\mathcal{C}_{1,\theta_z-M_z} &= \frac{12}{Ewt^3}\left\{L-2r+\frac{2r}{(2r+t)(4r+t)^3}\left[t(4r+t)(6r^2+4rt+t^2)\right.\right. \\
&\quad \left.\left. + 6r(2r+t)^2\sqrt{t(4r+t)}\arctan\sqrt{1+\frac{4r}{t}}\right]\right\}.
\end{aligned} \tag{4.25}$$

The out-of-plane compliances are

$$\begin{aligned}
\mathcal{C}_{1,z-F_z} &= \frac{12}{Ew^3}\left\{\frac{(L-2r)(L^2-Lr+r^2)}{3t} + Lr\left[\log\frac{t}{2r+t} - \frac{2(L-2r)}{\sqrt{t(4r+t)}}\right.\right. \\
&\quad \left.\left. \times \arctan\sqrt{1+\frac{4r}{t}}\right]\right\}; \\
\mathcal{C}_{1,z-M_y} &= \frac{6}{Ew^3t}\left\{L(L-2r) + 2r\left[t\log\frac{t}{2r+t} - 2(L-2r)\sqrt{\frac{t}{4r+t}}\right.\right. \\
&\quad \left.\left. \times \arctan\sqrt{1+\frac{4r}{t}}\right]\right\}; \\
\mathcal{C}_{1,\theta_y-M_y} &= \frac{6}{Ew^3t}\left[2L - 4r = \pi t + 4(2r+t)\sqrt{\frac{t}{4r+t}}\arctan\sqrt{\frac{t}{4r+t}}\right]. \quad (4.26)
\end{aligned}$$

For a circular flexure hinge as shown in Fig. 4.8c, where the thickness is $t(x) = t + 2[r - \sqrt{x(2r-x)}]$ for $0 \le x \le L$, the final results for the in-plane compliances are

$$\begin{aligned}
\mathcal{C}_{1,x-F_x} &= \frac{1}{Ew}\left[\frac{2(2r+t)}{\sqrt{t(4r+t)}}\arctan\sqrt{1+\frac{4r}{t}-\frac{\pi}{2}}\right]; \\
\mathcal{C}_{1,y-F_y} &= \frac{3}{4Ew(2r+t)}\left\{2(2+\pi)r + \pi t + \frac{8r^3(44r^2+28rt+5t^2)}{t^2(4r+t)}\right. \\
&\quad + (2r+t)\sqrt{t(4r+t)} \\
&\quad \times \frac{-80r^4+24r^3t+8(3+2\pi)r^2t^2+4(1+2\pi)rt^3+\pi t^4}{\sqrt{t^5(4r+t)^5}} \\
&\quad \left. - \frac{8(2r+t)^4(-6r^2+4rt+t^2)}{\sqrt{t^5(4r+t)^5}}\arctan\sqrt{1+\frac{4r}{t}}\right\}; \\
\mathcal{C}_{1,y-M_z} &= \frac{24r^2}{Ewt^3(2r+t)(4r+t)^3}\left[t(4r+t)(6r^2+4rt+t^2)\right. \\
&\quad \left. + 6r(2r+t)^2\sqrt{t(4r+t)}\arctan\sqrt{1+\frac{4r}{t}}\right] \\
\mathcal{C}_{1,\theta_z-M_z} &= \frac{\mathcal{C}_{1,y-M_z}}{r}. \quad (4.27)
\end{aligned}$$

The out-of-plane compliances are

$$\mathcal{C}_{1,z-F_z} = \frac{24r^2}{Ew^3}\log\frac{t}{2r+t};\quad \mathcal{C}_{1,z-M_y} = \frac{\mathcal{C}_{1,z-F_z}}{2r};$$
$$\mathcal{C}_{1,\theta_y-M_y} = \frac{6}{Ew^3t}\left[4(2r+t)\sqrt{\frac{t}{4r+t}}\arctan\sqrt{\frac{t}{4r+t}} - \pi t\right]. \tag{4.28}$$

Finally, for an ellipse flexure hinge as shown in Fig. 4.8d, with the thickness given by

$$t(x) = t + 2c\left[1 - \sqrt{1 - \left(1 - \frac{2x}{c}\right)^2}\right], \tag{4.29}$$

where c is a constant, the in-plane compliances are

$$\mathcal{C}_{1,x-F_x} = \frac{1}{4Ewc}\left[\frac{4(2c+t)}{\sqrt{t(4c+t)}}\arctan\sqrt{1+\frac{4c}{t}} - \pi\right];$$
$$\mathcal{C}_{1,y-F_y} = \frac{3L^3}{16Ewt^3c^3(2c+t)(4c+t)^2}\Big\{t[96c^5 + 96c^4t + 8(11+4\pi)c^3t^2$$
$$+ 32(1+\pi)c^2t^3 + 2(2+5\pi)ct^4 + \pi t^5] - 4\sqrt{\frac{t}{4c+t}}(2c+t)^4$$
$$(-6c^2 + 4ct + t^2)\arctan\sqrt{1+\frac{4c}{t}}\Big\};$$
$$\mathcal{C}_{1,y-M_z} = \frac{6L^2}{Ewt^2(2c+t)(8c^2+t^2)}$$
$$\times\left[6c^2 + 4ct + t^2 + \frac{6c(2c+t)^2}{\sqrt{t(4c+t)}}\arctan\sqrt{1+\frac{4c}{t}}\right]. \tag{4.30}$$

The out-of-plane compliances are

$$\mathcal{C}_{1,z-F_z} = \frac{3L^3}{16Ew^3c^3}\left\{2(4-\pi)c^2 + 4(1+\pi)ct + \frac{(2c+t)(-4c^2+4ct+t^2)}{\sqrt{t(4c+t)}}\right.$$
$$\left.\times\left[2\arctan\frac{2c}{\sqrt{t(4c+t)}} - \pi\right]\right\};$$
$$\mathcal{C}_{1,z-M_y} = \frac{3L^2}{2Ew^3c}\left\{\frac{2c+t}{\sqrt{t(4c+t)}}\left[2\arctan\frac{2c}{\sqrt{t(4c+t)}} + \pi\right] - \pi\right\}. \tag{4.31}$$

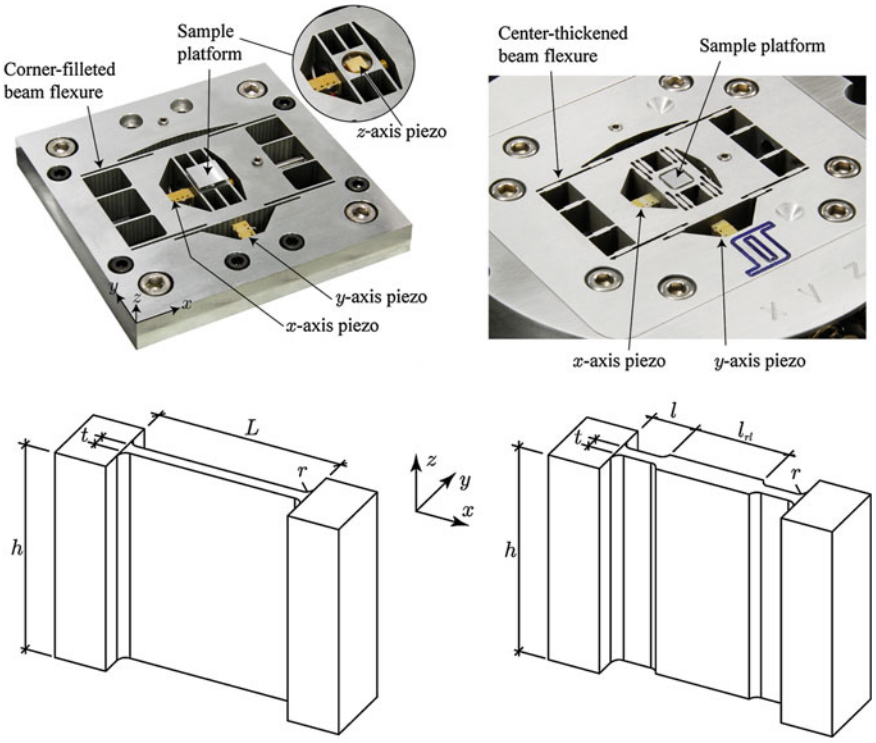


Fig. 4.9 Flexure design for increased out-of-plane stiffness. *Left* conventional corner-filleted beam flexure; and *right* serial-compliant double-hinged flexure with thickened center section

4.4.4 Stiff Out-of-Plane Flexure Designs

The majority of multiaxis nanopositioning stages for applications such as SPMs employ flexures hinges to guide the motion of the sample platform. The main objective is to limit parasitic (i.e., out-of-plane and off-axis) motion so that the stage only moves in the direction of actuation. For scanning at low speed, parasitic motion of the sample platform can be minimized using a simple beam flexure to guide the motion of the platform. As actuation frequencies increase, in- and out-of-plane resonance modes can be excited, thus limiting the positioning speed. However, dominant resonances occurring in the actuation direction are tolerable compared to out-of-plane modes; preferably, if the actuation modes precede the out-of-plane or off-axis modes. To ensure this, flexures should be designed to maximize the out-of-plane stiffness, yet be sufficiently *soft* to avoid affecting the achievable stroke of the actuator. Take the lateral scanning motion for example, where along the x and y-axis the sample platform is connected to simple beam flexures as shown in Fig. 4.9, top and bottom left photograph and sketch. The vertical stiffness of the sample platform can be maximized by (1) increasing the number of flexures in x and y, (2) utilizing shorter

(effective length) flexures, and (3) converting the flexures from constant rectangular cross section beam flexures to a serial-compliant double-hinged flexure with a “rigid" center connecting link as shown in Fig. 4.9 (Kenton and Leang 2012).

It is pointed out that the limiting factor of decreasing the flexure thickness is stress. Shorter thinner beam flexure will have higher stress concentration than a longer thicker beam flexure of equal stiffness. When a corner-filleted beam flexure (Fig. 4.9) is displaced in the vertical direction, the majority of the vertical displacement is caused by shear deformation of the center section. Thus, an effective way to increase the out-of-plane stiffness of a beam flexure is by thickening the center section, effectively converting the beam flexure into a double-hinged serial flexure as shown in Fig. 4.9, top and bottom right photograph and sketch. A serial-compliant flexure is one such that there are more than one flexure or flexure hinge in series with each other separated by a rigid link. By increasing the number of flexures, decreasing the flexure length, and thickening the center section of a beam flexure to create a serial-compliant double-hinged flexure, the effective vertical stiffness can be increased significantly (Kenton and Leang 2012).

4.4.5 Failure Considerations

In the design of flexure hinges for nanopositioning systems, failure due to yield and fatigue must be considered. Failure due to yield occurs when the deformation of the flexure exceeds that of the proportionality limit. Ductile materials are often chosen for flexure design. For ductile and isotropic materials, two most frequently used failure criteria includes the maximum shear stress theory (Tresca) and the maximum energy of deformation theory (von Mises) (Beer and Johnston 1992).

The maximum shear stress criterion is based on the idea that failure in ductile materials is caused by shearing stresses. A given structural component is deemed safe as long as the maximum shear stress $\tau_{\max}$ in the component is less than the shearing stress of the component at yield under a tensile test. Specifically, if the principal stresses σ_a and σ_b have the same sign, the maximum shear stress criterion gives

$$|\sigma_a| < \sigma_y \quad |\sigma_b| < \sigma_y, \tag{4.32}$$

where σ_y is the yield stress of the material. If the principal stresses σ_a and σ_b have opposite signs, the maximum shear stress criterion gives

$$|\sigma_a - \sigma_b| < \sigma_y. \tag{4.33}$$

The maximum energy of deformation criterion states that a given structural component is deemed safe as long as the maximum value of the distortion energy per unit volume of that material is less than the distortion energy per unit volume required to cause yield in a tensile-test specimen of the same material. Under plane stress, the distortion energy per unit volume in an isotropic material is

$$u_d = \frac{1}{6G}(\sigma_a^2 - \sigma_a\sigma_b + \sigma_b^2), \tag{4.34}$$

where G is the modulus of rigidity. A tensile-test specimen when it starts to yield has $\sigma_a = \sigma_y$ and $\sigma_b = 0$, therefore $(u_d)_y = \sigma_y^2/6G$. As a result, as long as $u_d < (u_d)_y$ or

$$\sigma_a^2 - \sigma_a\sigma_b + \sigma_b^2 < \sigma_y^2, \tag{4.35}$$

then the structural component is safe.

Commonly used failure criteria for brittle materials include the maximum normal stress criterion and Mohr's criterion and can be found in Beer and Johnston (1992). One of the most practical approaches uring the design process to ensure that a given flexure is design within the failure tolerances is to employ finite element programs such as ANSYS (Canonsburg, PA, USA) and Solidworks with COSMOSWorks (Concord, MA, USA).

4.4.6 Finite Element Approach for Flexure Design

The finite element analysis (FEA) method is a power numerical technique for solving engineering and mathematical physics problems that include structural analysis, heat transfer, electro-mechanical coupling, fluid flows, and mass transport. It is particularly useful for problems with complicated geometries, materials properties, loadings, and boundary conditions. Popular commercially available programs includes ANSYS (Canonsburg, PA, USA) and Solidworks with COSMOSWorks (Concord, MA, USA). Figure 4.10 shows FEA results that compare the stress distribution between the basic geometry to that of a corner-filleted flexure (Fig. 4.5b). The ability for generating quick and relatively accurate results using FEA software has made it a popular choice for mechanical design. The FEA programs can also be used to estimate frequency response functions for nanopositioning designs, where results can show the effects of cross-couplings and out-of-plane behaviors. Users can expect accuracy of less than 10 % compared to experimentally measured results (see example in Fig. 4.11).

4.5 Material Considerations

4.5.1 Materials for Flexure and Platform Design

A basic nanopositioning stage consists of a rigid frame upon which an actuator rests and pushes off against to displace a mass or flexure member. The material for the frame as well as any required flexure hinges must be carefully selected for optimum

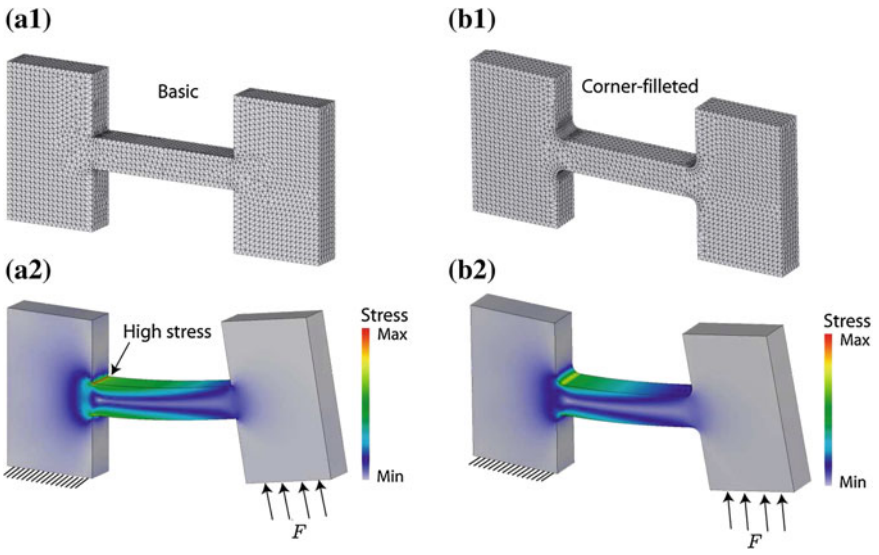


Fig. 4.10 Finite element results comparing the stress distribution between a **a** basic and **b** corner-filleted flexure hinge

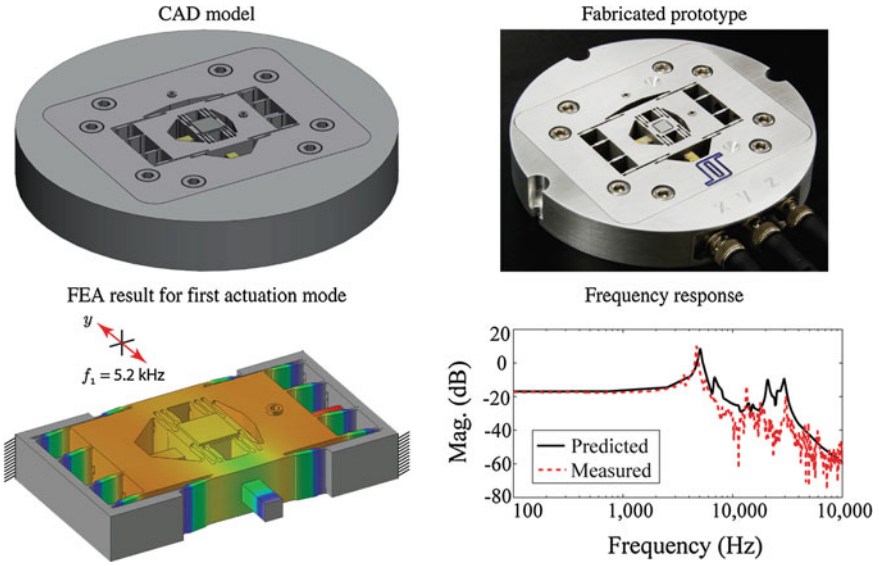


Fig. 4.11 Comparison between FEA predicted and measured dynamic performance

Table 4.1 Properties of various materials

Material	Density (kg/m^3)	Young's Mod. (GPa)	Thermal Cond. (W/m-C)	CTE (10^{-6}/°C)
6061 Aluminum	2710	70	167	23.6
7075 Aluminum	2800	72	130	23.6
Stainless steel	7920	190	16.3	17.3
Titanium	4730	115	15.6	9.5
Invar (Invar 36)	8130	144	13.8	1.6
Super Invar	8137	144	10.5	0.3

static and dynamic performance. Some popular materials are listed in Table 4.1. Aluminum alloys, such as the 7075 grade is the most commonly used due to its machinability and favorable density-to-stiffness ratio. Materials which exhibit extremely low thermal coefficient of expansion include Invar and Super Invar (Schilfgaarde et al. 1999). Stainless steel is often used to create the rigid base due to its high elastic modulus and resistance to corrosion. Likewise, AISA A2 steel, which is easily machinable, has an equivalent elastic modulus to stainless steel following a heat-treating process. The heat-treating process involves heating the material to 850 °C, followed by cooling in the furnace at 10 °C per hour to 650 °C. Finally, the material is cooled freely in air. The other components that require a high stiffness to density ratio, such as the flexure hinges and sample platforms, are constructed from aluminum.

4.5.2 Thermal Stability of Materials

Thermal expansion is the dimensional change of a material due to a change in temperature, and it is generally inversely proportional to the melting point of a material. The effect can severely limit the precision, repeatability, and overall performance of a nanopositioning system, such as causing temperature-dependent drift in motion and thermal stresses which ultimately lead to cracking, warping, or loosening of components.

The change in length (from l_0 to l_f) for a solid material for a given change in temperature (from T_0 to T_f) is given by

$$\frac{l_f - l_0}{l_0} = \alpha(T_f - T_0), \tag{4.36}$$

where α is the linear thermal coefficient of thermal expansion and has units of $(°C)^{-1}$ or K^{-1}. For nanometer motion, thermal effects can not be ignored. Careful material selection and design are effective methods for minimizing thermal effects. Table 4.1 lists the mechanical and thermal properties of commonly-used materials for the design of nanopositioning stages. For example, the coefficient of thermal expansion (CTE) for aluminum is 23×10^{-6} /°C, while for Super Invar alloy it

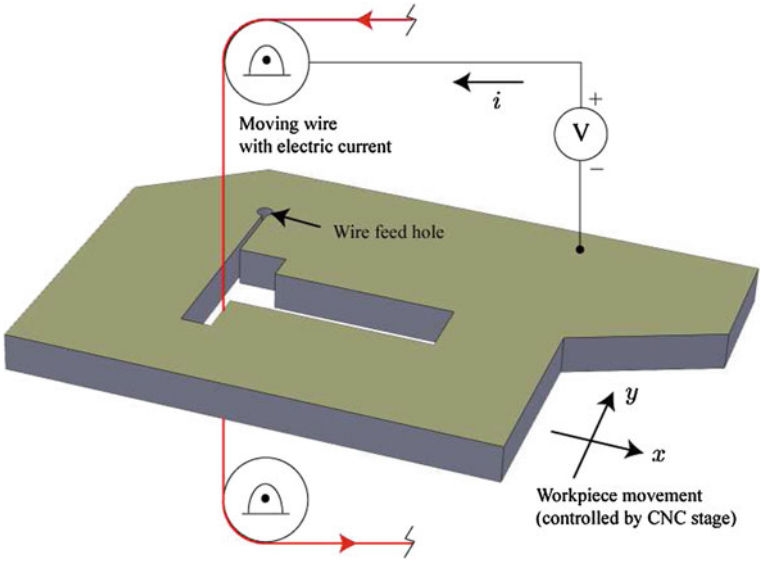


Fig. 4.12 Wire EDM machining process

is only 0.3×10^{-6} /°C, over 70 times lower. Best practices also include carefully matching the stage material with the materials of surround support structures. Also, materials with high thermal conductivity quickly reach thermal equilibrium, thus minimizing transient distortion during thermal expansion.

4.6 Manufacturing Techniques

The use of flexure mechanisms with stock piezoactuators require the ability to manufacture them. In general, the appropriate manufacturing technique for a given design and application depends on the scale of the positioning stage and the selected materials.

Standard milling, turning, and drilling techniques are best suited for metals, such as aluminum, titanium, and steel. These techniques are best for feature sizes above 1 mm, and they can achieve tolerances on the order of $\pm$ 0.001 in ($\approx$25.4 μm).

Monolithic flexures, especially those with complex shapes and intricate dimensions, are best created using wire electrical discharge machining (EDM). 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 an electrode tool (wire) and the workpiece. As shown in Fig. 4.12, a basic EDM system consists of a wire electrode, a wire-feeding mechanism, and a conductive part that is moved relative to the wire in the lateral (x and y) directions using a numeric computer controlled

(CNC) stage. To create a cutout, a wire feed hole is first created where the wire is passed through the hole. During machining, debris is often flushed away from the cutting surface using an appropriate liquid. Wire diameters of approximately 100 μm are often used. Traditional machining techniques are often used to remove the bulk of the stock before performing EDM. Dimensional accuracy on the order of ±12 μm (± 0.0005 inch) can be achieved using the EDM process.

4.7 Design Example: A High-Speed Serial-Kinematic Nanopositioner

The design of a high-speed serial kinematic nanopositioner is described as an illustrative example. The nanopositioner is created for high-bandwidth applications that include video-rate scanning probe microscopy and high-throughput probe-based nanofabrication. The design offers approximately 9 × 9 × 1 μm range of motion and kHz bandwidth. Vertically stiff, double-hinged serial flexures are employed to guide the motion of the sample platform to minimize parasitic motion (runout) and off-axis effects (refer to Sect. 4.4.4). Additionally, the stage's out-of-plane stiffness is further improved by increasing the quantity of flexures n, decreasing the length L of each flexure, and thickening each flexure's center cross section. The effects of varying these parameters are examined in some detail. Along the vertical axis (z), a novel plate flexure guides the motion of the sample stage to minimize the effects of bending modes. Bending modes can significantly limit the positioning speed by causing the sample platform to rock side-to-side. It is pointed out for scanning-type applications, one lateral axis operates much faster than the other, and thus the serial-kinematic configuration is well-suited for these types of applications (Ando et al. 2008; Picco et al. 2007; Leang and Fleming 2009). Finally, the stage is integrated with a commercial scan-by-probe atomic force microscope and imaging and tracking results up to a line rate of 7 kHz are presented. At this line rate, 70 frames per second AFM video (100 × 100 pixels resolution) can be achieved.

4.7.1 State-of-the-Art Designs

A summary of existing multiaxis nanopositioning designs is listed in Table 4.2. One of the simplest and most effective way to achieve three-axis motion is to employ sectored tube-shaped piezoelectric actuators (Schitter and Stemmer 2004). However, the mechanical resonance of piezoelectric tube scanners is typically less than 1 kHz in the lateral scan directions, thus limiting the scan speed (Schitter and Stemmer 2004; Schitter et al. 2008; Fleming 2009; Rifai and Youcef-Toumi 2001). Additionally, the mechanical cross coupling causes undesirable SPM image distortion (Rifai and Youcef-Toumi 2001). In general, the maximum open-loop (without compensation)

Table 4.2 Summary of nanopositioner designs

Configuration	Range (μm)	Dominant Res. (kHz)	Imaging/line rate (range)
Tube scanner (Schitter and Stemmer 2004)	125 (x/y)	0.71 (x)	122 lines/s
		0.70 (y)	(13.5 × 13.5 μm)
Tube scanner	n/a	6.35 (x/y)	3 lines/s
Dual stage (z) (Schitter et al. 2008)		80 (z)	(25 μm)
Tube scanner	100 (x/y)	0.68 (x/y)	6.25 lines/s
Dual stage (z) (Fleming 2009)	10 (z)	23 (z)	(25 × 25 μm)
Shear piezo (Rost et al. 2005)	0.3 (x/y)	~64	80 frames/s
	0.20 (z)	>100	(128 × 128 px)
Flexure guided (Ando et al. 2008)	1 (x)	45	33 frames/s
	3 (y)		(100 × 100 px)
	2 (z)	360 ("self")	
Tuning fork (x)	<1 (x)	100	1000 frames/s
Flexure guided (y) (Picco et al. 2007)	2 (y)	40	(100 × 100 px)
Flexure guided (Schitter and Rost 2008)	13 (x/y)	>20	7810 lines/s
	4.3 (z)	33	(n/a)
Flexure guided (Yong and Aphale 2009)	25 (x/y)	2.73	n/a

positioning bandwidth is 1/100 to 1/10th of the dominant resonance (Clayton et al. 2009).

One of the earliest works on stiff mechanical nanopositioners was by Ando and co-workers (Ando et al. 2002), where stiff piezo-stack actuators were used to create a high-speed scanner. The researchers demonstrated imaging at 12.5 frames/s (100 × 100 pixels per image), and they used the system to capture real-time video of biological specimens (Ando et al. 2005). Shortly after, Schitter and co-workers also developed a scanner based on piezo-stack actuators, but in their design the actuators were arranged in a push-pull configuration and mechanical flexures were used to decouple the lateral and transverse motions (Schitter et al. 2007). They employed FEA to optimize the performance of the mechanical structure (Kindt et al. 2004). The reported AFM imaging rate is 8 frames/s (256 × 256 pixels). By exploiting the stiffness of shear piezos and a compact design, a scanner was created for imaging up to 80 frames/s (128 × 128 pixels) with a line rate of 10.2 kHz (Rost et al. 2005). The achievable range of motion is 300 × 300 nm. Another unique approach for high-speed scanning involves a piezo-stack actuator combined with a tuning fork as reported in Humphris et al. (2003). The tuning fork operated at resonance and AFM images were acquired at 100 frames/s (128 × 128 pixels). Likewise, a combined flexure-based scanner and tuning fork achieved imaging rate in excess of 1000 frames/s in Picco et al. (2007). Although the tip motion was fast, the range was limited and the trajectory was sinusoidal.

Flexure-guided piezoactuated scanning stages (Scire and Teague 1978), both direct drive serial-kinematic (Ando et al. 2008; Leang and Fleming 2009) and parallel-kinematic (Schitter et al. 2008) configurations, have been developed for

high-speed purposes. The advantages of flexure-guided scanners are high mechanical resonances and low cross-coupling. Multiple piezoactuators per degree-of-freedom have been used to increase range and bandwidth, but at the cost of increased power to drive the piezoactuators at high frequencies (Ando et al. 2008; Schitter et al. 2008). Designs which involve mechanical amplification have been implemented to increase range without having to increase the actuator's length (Scire and Teague 1978; Yong and Aphale 2009). However, the added mass of the mechanical amplifier along with the flexible linkages lowers the mechanical resonance. In general, a tradeoff must be made between range and bandwidth.

4.7.2 Tradeoffs and Limitations in Speed

The major tasks to design a high-speed nanopositioner include: (1) identifying relevant design parameters and tradeoffs, such as range of motion and maximum scanning bandwidth, (2) using FEA tools to optimize the mechanical structure, and (3) developing the necessary electronic hardware for the scanner. The design process is often iterative.

First, it is worth noting that range of motion conflicts directly with the achievable mechanical resonance. For example, large range requires large mechanical amplification A_f, which lowers the effective stiffness of the scanner (see Eq. (4.1)), and therefore the mechanical resonance (Kindt et al. 2004). One can expect that a piezo-driven nanopositioner with range of 1 μm or less will have a dominant mechanical in the hundreds of kHz range. For a positioner with a range between 1 and 5 μm, the mechanical resonance is often in the tens of kHz range. When the range is between 5 to 10 μm, the resonance falls to the kHz to tens of kHz range. Finally, ranges above 10 μm drop the mechanical resonance to the kHz and hundreds of Hz range. Most high-speed nanopostioners have operating range of less than 10 μm. This range of motion is still practical as it enables a scanner used in AFM to observe a wide spectrum of specimens and samples, from micron-size cells to submicron-size subjects such as DNA.

Video-rate SPM imaging requires a modest 30 frames per second, where each frame is at least 100 × 100 pixels. At this frame rate, the required linear scan rate is 3 kHz for the fast scan axis (along the x-direction, for example), and 30 Hz for the slow (y) axis. Faster scanning will increase frame rate, and/or frame resolution. The desired linear scan rate establishes a target for the dominant mechanical resonance in both axes. Assuming that the frequency of the command signals to drive the actuators must be at least 1/10th of the lowest resonance along each axis to avoid dynamic effects, the lowest mechanical resonance should be 300 and 30 kHz, for the y and x axis, respectively.

For comparison, the relationship between range and resonance frequency for a variety of commercial and custom nanopositioners is shown in Fig. 4.13 (Kenton 2010). The range is plotted with respect to the resonance frequency for each stage when provided. When full details are not provided for multiaxis positioners,

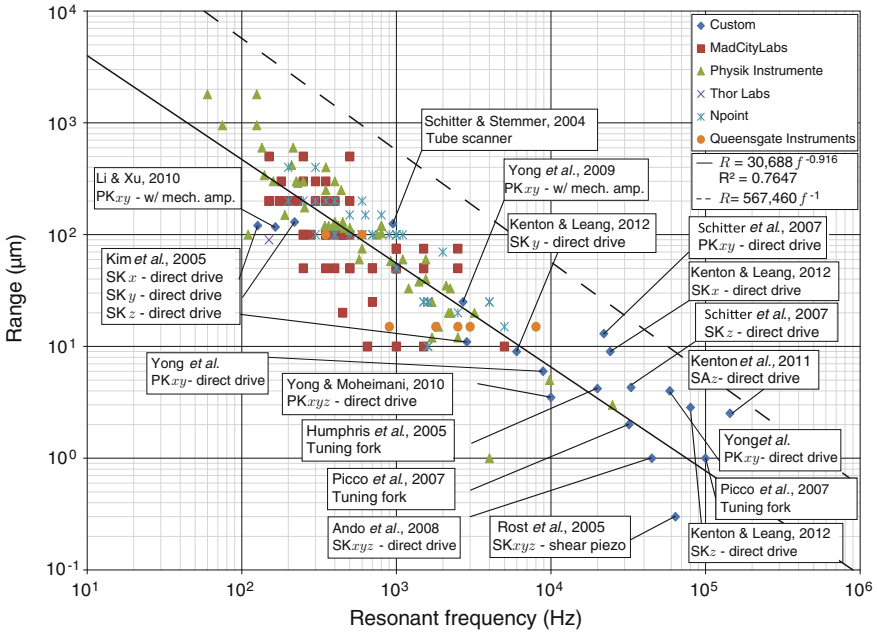


Fig. 4.13 High-performance commercial and custom nanopositioners plotted as range with respect to resonance frequency (adapted from Ref. Kenton (2010)). The *solid line* represents a linear least-square-error line fit to the data points. The *dashed line* represents the theoretical first mechanical resonance in the actuation mode for a fixed-free piezoactuator (assuming 1 μm of travel per 1 mm length). SK = serial-kinematic, PK = parallel-kinematic, SA = single-axis, x, y, z refers to axis being referenced

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. The dashed line in Fig. 4.13 marks the theoretical limit for a fixed-free piezoactuator with a modulus of elasticity of 33.9 GPa and a density of 8,000 kg/m^3 assuming 1 μm of travel per mm of piezo length (Kenton 2010). The commercial and custom nanopositioners in Fig. 4.13 are well below this theoretical limit. The solid line represents a fit to the data for commercial and custom nanopositioners.

The required power to drive the subject scanner must also be considered. The available power restricts the amount of voltage and current that can be delivered to the actuator. In turn, this restricts the type and dimensions of the piezoactuator that can be used for positioning. Larger piezoelectric actuators can provide greater stroke, but have higher capacitance and require more power at high frequencies.

Finally, cost and manufacturability must also be factored into the design. The scanner fabrication should not utilize any exotic materials or processes and should be tolerant of typical machining tolerances.

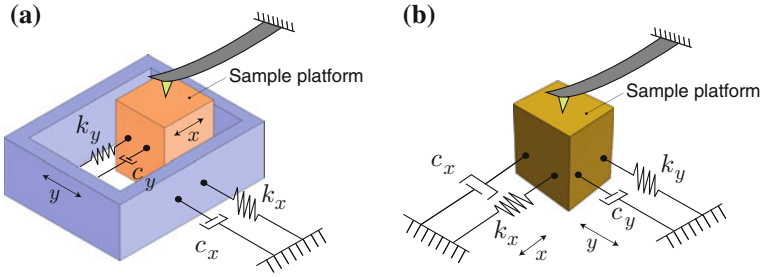


Fig. 4.14 Simplified models for two-axis scanning: **a** serial- and **b** parallel-kinematic configuration. The spring and damping constants include the effects of the piezoactuators and added flexures in each direction

4.7.3 Serial- Versus Parallel-Kinematic Configurations

For scanning in two directions, there are two basic configurations: serial and parallel kinematics as illustrated in Fig. 4.14. In a serial-kinematic system, for example the design used by several commercial vendors of scanning stages and in Ando et al. (2005), Kenton and Leang (2012), there is exactly one actuator (and sensor) for each degree of freedom (see Fig. 4.14a). One disadvantage of this design is the inability to measure (and correct for) parasitic motion such as runout or guiding error. Although the serial configuration is simple to design, a penalty is that high resonance frequency can only be achieved in one axis.

In a parallel-kinematic scanning stage, e.g., Schitter et al.'s work (Schitter 2007), all actuators are connected in parallel to the sample platform (see Fig. 4.14b). This arrangement enables rotation of the image, i.e., the fast scanning axis can be chosen arbitrarily. An advantage of this configuration is that parasitic motion due to runout and guiding error can easily be measured and corrected. However, since the mechanical dynamics of both the lateral and transverse axes are similar, high-bandwidth control hardware is required for both directions. In contrast, for the serial-kinematic configuration only the high-speed axis requires high power and wide bandwidth performance, reducing overall cost.

In the simplified model shown in Fig. 4.14, the effective stiffnesses and damping effect for both the serial and parallel kinematic configurations include the flexures and the actuators along each direction. To achieve high resonance frequencies, the effective stiffnesses should be a high as possible while achieving the desired range of motion. The effects of inertial force generated by the sample platform during scanning must also be taken into account. The flexures must provide enough preload to avoid exposing the stack actuator to damaging tensional forces.

While the resonance frequency of the fast axis is of primary concern, the slow-axis resonance frequency can essentially be ignored. For example, the scan rate of the slow axis is one-hundredth the scan rate of the fast-axis when acquiring a 100×100 pixel image. Therefore, the fast scan axis can be designed independently without any

Table 4.3 Comparison of plate-stack piezoactuators

Size (mm)	Free stroke (μm)	Cap. (nF)	k_a (N/μm)	k_z* (N/μm)	k_z/k_a
3 × 3 × 10	11.15	114	30.5	2.0	0.066
5 × 5 × 10	11.78	387	84.8	10.7	0.126
7 × 7 × 10	12.09	835	166.1	28.0	0.169
10 × 10 × 10	12.13	1673	339.0	69.2	0.204

* Stiffness for fixed-guided beam accounting for shear (see Sect. 4.7.6)

significant consideration for performance implications on the slow-scan axis. As previously stated, for scanning-type applications, one lateral axis operates much faster than the other, and thus the serial-kinematic configuration is well-suited for these types of applications (Ando et al. 2008; Picco et al. 2007; Leang and Fleming 2009).

4.7.4 Piezoactuator Considerations

The actuating mechanism of choice for scanning at high speed is the piezoactuator (Kenton and Leang 2012), particular piezo-stack actuators. Although thin shear-piezos offer higher mechanical resonances, their range is rather limited (sub-micron level) (Rost et al. 2005). Piezo-stack actuators are stiff and compact. A comparison of four plate-stack piezoactuators (Noliac) of varying cross-sectional areas is shown in Table 4.3. Each actuator in this comparison is 10-mm long and meets the desired free stroke of 11 μm. (A small percentage of the free stroke will be lost due to flexure stiffness and boundary conditions associated with gluing the piezo-stack to the stage during assembly.) The capacitance increase is nearly proportional to the cross-sectional area with an average of 15.5 nF/mm^2 (for a 10-mm long piezo-stack actuator). The Young's modulus is calculated from the blocking force and free stroke. For instance, the Young's modulus of a 5 × 5 × 10 mm piezoactuator is determined to be 33.9 GPa (Leang and Fleming 2009). As shown in Table 4.3, higher actuation and out-of-plane stiffness can be obtained by using larger (cross-section) piezo-stacks. The cost, however, is higher capacitance which increases the net power to drive the actuators, especially at high frequencies.

To achieve the desired scan range of 10 × 10 μm, a piezoelectric stack actuator with dimensions of 5 × 5 × 10 mm and capacitance of 387 nF is chosen (e.g., Noliac SCMA-P7). Figure 4.15 shows a photograph of the piezo-stack actuator to drive the high- and low-speed stages. The actuator stroke is 11.8 μm, with an unloaded resonance frequency of 220 kHz, stiffness of 283 N/μm, blocking force of 1000 N, and maximum drive voltage of 200 V. The elastic modulus calculated from the blocking pressure P_B and strain ϵ is $E = \frac{P_B}{\epsilon} = 33$ GPa.

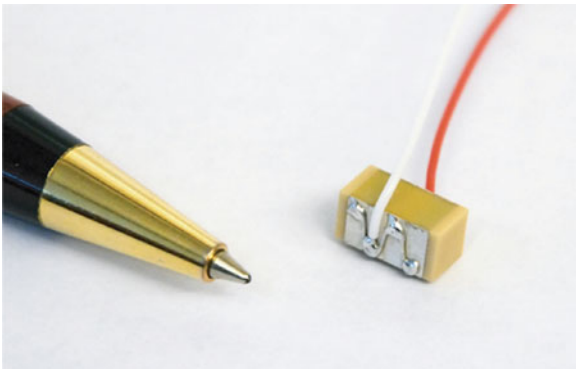


Fig. 4.15 A photograph of the piezo-stack actuator used to drive the x and y stage

4.7.5 Preloading Piezo-Stack Actuators

As discussed in Sect. 2.5.3, piezo-stacks are intolerant to tensile (as well as shear) stresses. Because stacks are constructed of glued (or fused) piezoelectric layers, tensional loads can cause the actuators to fail at the interface (glue) layers. Manufacturers often specify a tensile load limit less than 10 % of the compressive load limit. During high-speed operation, inertial forces due to the sample mass must be taken into account to avoid excessive tensile stresses. A preload force must be incorporated to eliminate the possibility of the actuator being exposed to excessive tensile forces. The preload must be applied in such a way that full surface contact is achieved to assure good load distribution (see Fig. 2.9). Recommended preload force is 20 % of the compressive load limit of the actuator, and the preload spring stiffness should be at most 10 % of the actuator stiffness.

Flexures can be used to apply the appropriate preload on the piezo-stack actuator to compensate for the inertial force during dynamic operation. The flexures serve two purposes: to eliminate tensile stress and to guide the extension/contraction of the actuator so that parasitic motion is minimized. The required preload is estimated from Newton's Second Law by computing the maximum sample platform acceleration during maximum excursion and scan frequency. In particular, the magnitude of the expected dynamic (inertial) force on a piezoactuator assuming sinusoidal motion at frequency f (in Hz) is given by

$$F_i = 4\pi^2 m_{\text{eff}} \left(\frac{\Delta x}{2} \right) f^2, \tag{4.37}$$

where m_{eff} is the effective mass and Δx is the total stroke length. For example, a 5 g sample positioned over a 10 μm range at 3,000 Hz, the minimum preload requirement is 8.9 N. Considering a safety factor of 2, the required preload is at least 18 N.

4.7.6 Flexure Design for Lateral Positioning

The basic layout for a serial-kinematic design is shown in Fig. 4.16, where the high-speed (x-axis) stage is nested inside of the low-speed (y-axis) stage (Fig. 4.16a, b). The stage body is manufactured from 7075 aluminum using the wire EDM process to create a monolithic platform. The sample platform is located on the x-stage body, and vertical motion is achieved with a piezo-stack actuator embedded into the x-stage body (Ando et al. 2008) (see details in Fig. 4.16c). Compliant flexures with improved vertical-stiffness to minimize out-of-plane motion are used to guide the motion of the sample platform. The flexures are strategically placed to minimize the sample platform's tendency to rotate (θ_x, θ_y, θ_z) at high frequencies. Also, the stage is designed to ensure that the first resonance in all three axes are axial (piston) modes, rather than off-axis modes, which can severely limit performance.

For translational motion, u_i ($i = x,\ y,\ z$), the single degree-of-freedom mechanical resonance is given by $f_{u_i,0} = \frac{1}{2\pi}\sqrt{\frac{k_i}{m_i}}$, where m_i and k_i are the effective translational mass and stiffness, respectively. Likewise for rotational motion, θ_i ($i = x,\ y,\ z$), the first resonance is $f_{\theta_i,0} = \frac{1}{2\pi}\sqrt{\frac{k_{\theta_i}}{J_i}}$, where J_i and k_{θ_i} are the effective mass moment of inertia and rotational stiffness, respectively. To insure that actuation modes occur before the out-of-plane modes, the strategy taken is to optimize the stage geometry and flexure configuration so 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 the actuation stiffness-to-mass ratio k_x/m_x. Figure 4.17 shows the simplification of a high-speed x-stage into single degree-of-freedom systems to model four of the dominating resonance modes. The top and side views are broken down to show the effective springs and masses affecting the body for (d) actuation u_x, (e) and (f) rotation θ_z and θ_y, and (g) vertical u_z modes. Damping is omitted for convenience.

The vertical stiffness of the x- and y-stages is increased by (1) increasing the number of flexures, (2) utilizing shorter (effective length) flexures, and (3) converting the flexures from constant rectangular cross section beam flexures to a serial-compliant double-hinged flexure with a "rigid" center connecting link (see Fig. 4.9). The first step taken to increase the flexure stiffness in the vertical direction is studying how the total number of flexures n used in parallel, flexure thickness t, and length L affect the vertical stiffness k_z for a given actuation stiffness k_a. This comparison is done analytically and using finite element analysis (COSMOSWorks FEA).

The stiffness of a flexure is defined as the ratio of a load F and the resulting displacement u. The displacements and loads are: translational displacement u_i, rotational displacement θ_i, translational force F_i acting on a point in the i direction, and moment M_i (torque T) acting about the i axis (θ_i), respectively, where $i = x, y, z$. Figure 4.7 illustrates the corresponding directions of the displacements and loads acting on the free end of a fixed/free cantilever beam which models a beam flexure. The in- and out-of-plane compliances for a fixed/free beam is derived using Castigliano's second theorem (Timoshenko 1953; Lobontiu 2003; Craig 2000). The compliance equations are then used to derive equations for the actuation and vertical

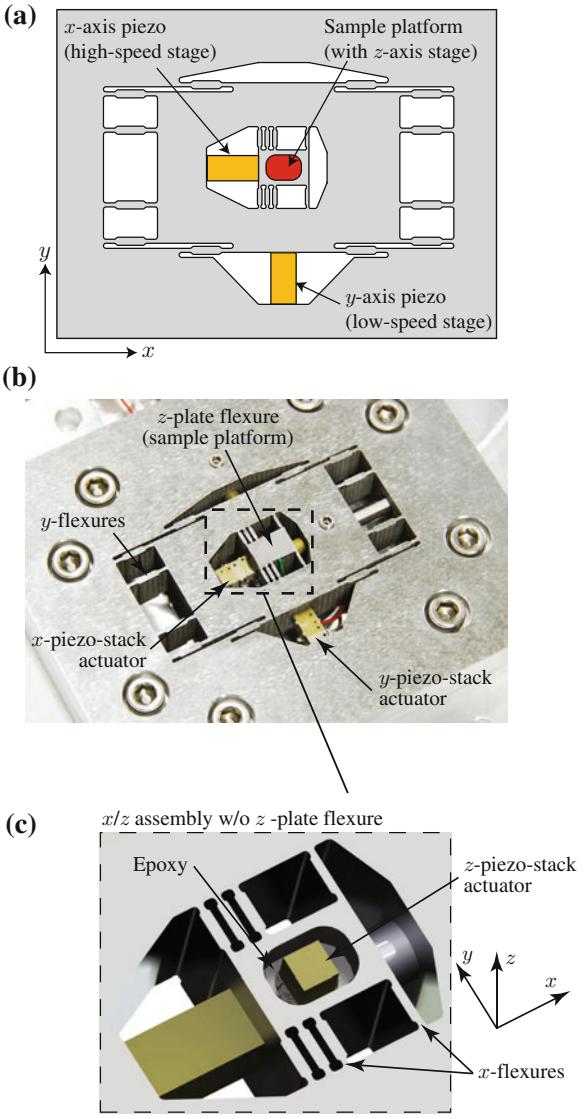


Fig. 4.16 A serial-kinematic nanopositioner: **a** *top view*, **b** fabricated stage, and **c** details of the sample platform and z-stage

stiffness k_i of a fixed/roller guided beam shown in Fig. 4.18a1 through a3. It is pointed out that the fillet radius is considerably smaller compared to the flexure length and therefore has minimal effect on the flexure stiffness. For this reason, to simplify the flexure stiffness equations in this initial analysis, the compliance equations are derived for a beam with a constant cross sectional thickness.

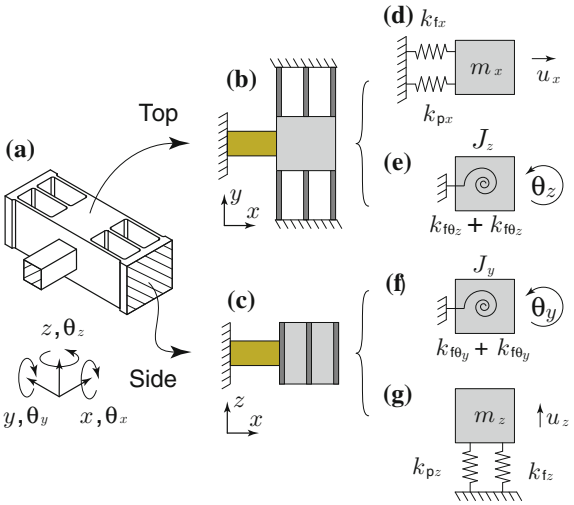


Fig. 4.17 Generic flexure-guided stage simplified to single degree-of-freedom systems modeling the dominant modes

For a fixed/free beam of rectangular cross section, the total strain energy is

$$\begin{aligned} U &= U_{\text{axial}} + U_{\text{torsion}} + U_{\text{bending}} + U_{\text{shear}} \\ &= \int_0^L \left[\frac{F^2}{2AE} + \frac{T^2}{2GJ} + \frac{M^2}{2EI} + \frac{\alpha V^2}{2GA} \right] \mathrm{d}x, \end{aligned} \tag{4.38}$$

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, $G = \frac{E}{2(1+\nu)}$ is the shear modulus, ν is Poisson's ratio, $J = ht^3 \left[\frac{1}{3} - 0.21 \frac{t}{h} \left(1 - \frac{t^4}{12h^4} \right) \right]$ is the approximate torsional moment of inertia (Young and Budynas 2002), $I = \frac{ht^3}{12}$ is the second moment of inertia about the vertical z axis, V is the shear force, and α is a shape factor for the cross section used in the shear equation (for a rectangular cross section $\alpha = 6/5$) (Craig 2000; Young and Budynas 2002; Park 2005).

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

$$u_i = \frac{\partial U}{\partial F_i}; \qquad \theta_i = \frac{\partial U}{\partial M_i}. \tag{4.39}$$

From here the compliance is simply found by dividing the displacement by the applied load, i.e.,

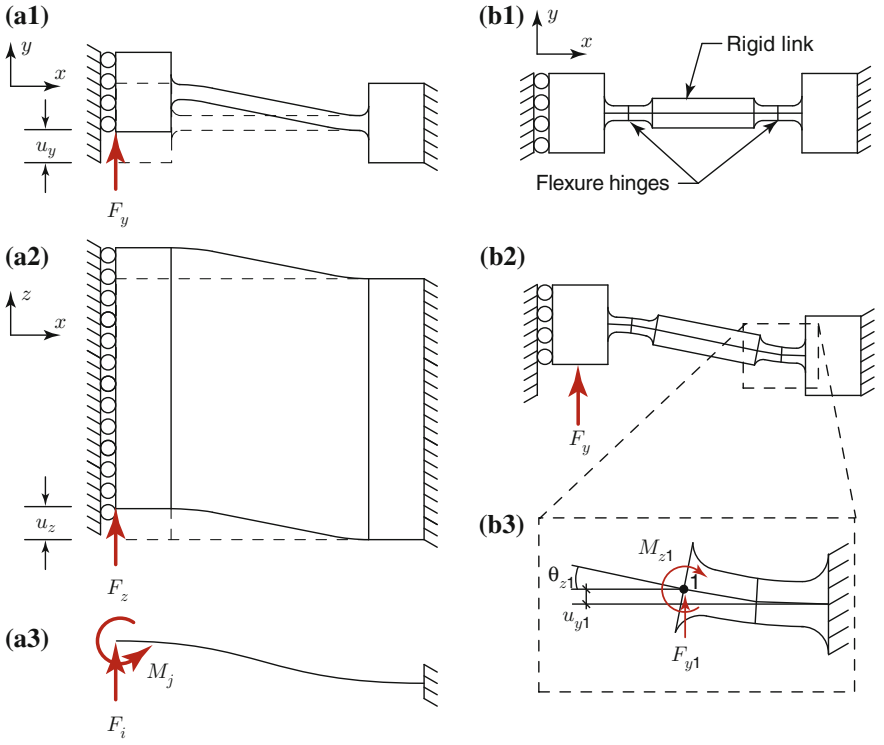


Fig. 4.18 Corner-filleted and center-thickened flexures showing loads and deformations: **a1** *top* and **a3** *side views* showing displacement caused by force F_i, for $i = x,\ y$, in a fixed/guided end configuration, and **a3** loads acting on the free end of a fixed/free beam for a corner-filleted. **b1** *top view*, **b2** *top view* with applied load, and **b3** expanded view of corner-filleted flexure hinge

$$C_{u_i,F_j} = \frac{u_i}{F_j}; \qquad C_{\theta_i,M_j} = \frac{\theta_i}{M_j}. \tag{4.40}$$

For example, the compliance of the rectangular cross section fixed-free beam in Fig. 4.7a due to a point load in the y direction starts with the total strain energy

$$U = \int_0^L \frac{M(x)^2}{2EI(x)}\,\mathrm{d}x + \int_0^L \frac{\alpha V(x)^2}{2GA(x)}\,\mathrm{d}x, \tag{4.41}$$

where $A(x)$ and $I(x)$ are constant. The coordinate system is placed on the free end of the flexure as shown in (Lobontiu 2003) where the shear is $V(x) = F_y$ and moment is $M(x) = F_y x$. The total strain energy for the applied load is

$$U = \frac{F_y^2}{2EI}\int_0^L x^2\,\mathrm{d}x + \frac{\alpha F_y^2}{2GA}\int_0^L \mathrm{d}x = \frac{F_y^2 L^3}{6EI} + \frac{\alpha F_y^2 L}{2GA}. \tag{4.42}$$

Therefore, the resultant displacement is

$$u_y = \frac{\partial U}{\partial F_y} = \frac{F_y L^3}{3EI} + \frac{\alpha L F_y}{GA}, \tag{4.43}$$

and the compliance is

$$C_{22} = \frac{u_y}{F_y} = \frac{L^3}{3EI} + \frac{\alpha L}{GA}. \tag{4.44}$$

The compliances are then used to form the compliance matrix $\mathbf{C}$ which is defined as the ratio of the displacement $\mathbf{U} = \begin{bmatrix} x & y & \theta_z & z & \theta_y & \theta_x \end{bmatrix}^T$ for a given load $\mathbf{L} = \begin{bmatrix} F_x & F_y & M_z & F_z & M_y & M_x \end{bmatrix}^T$, hence the displacement vector is

$$\begin{Bmatrix} u_x \\ u_y \\ \theta_z \\ u_z \\ \theta_y \\ \theta_x \end{Bmatrix} = \begin{bmatrix} C_{11} & 0 & 0 & 0 & 0 & 0 \\ 0 & C_{22} & C_{23} & 0 & 0 & 0 \\ 0 & C_{23} & C_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & C_{44} & C_{45} & 0 \\ 0 & 0 & 0 & C_{45} & C_{55} & 0 \\ 0 & 0 & 0 & 0 & 0 & C_{66} \end{bmatrix} \begin{Bmatrix} F_x \\ F_y \\ M_z \\ F_z \\ M_y \\ M_x \end{Bmatrix}. \tag{4.45}$$

For a constant cross section fixed/free beam the compliances are $C_{11} = \frac{L}{AE}$, $C_{22} = \frac{L^3}{3EI} + \frac{\alpha L}{GA}$, $C_{23} = \frac{L^2}{2EI}$, $C_{33} = \frac{L}{EI}$, $C_{44} = \frac{4L^3}{Eh^3t} + \frac{\alpha L}{GA}$, $C_{45} = \frac{6L^2}{Eh^3t}$, $C_{55} = \frac{12L}{Eh^3t}$, and $C_{66} = \frac{L}{GJ}$. For a long slender beam, shear strain has little effect and therefore can be ignored in C_{22}. For a short beam with a significant height-to-length aspect ratio, such as the vertical displacement of the flexure shown in Fig. 4.18a2, much of the deflection is in shear, and therefore can not be ignored.

The displacement vector equation presented above is used to solve for the actuation stiffness k_y and vertical stiffness k_z of a fixed/guided flexure beam, i.e., $F_i/u_i = k_i$. Torsional stiffness is not investigated because the θ_x rotational mode is largely dependant upon the vertical flexure stiffness when the flexures are placed at the corners of the stage body. Figure 4.18a3 shows the applied load and the expected deflection curve of the flexure in both the (a1) actuation direction and (a2) vertical direction. The active load being applied to the flexure is the in-plane force F_i. The resultant moment $M_j = -F_i L/2$ is caused by the roller-guided end constraint. Therefore, the flexure displacement in the actuation direction u_y due to the applied force F_y and moment $M_z = -F_y L/2$ is

$$u_y = C_{22}F_y + C_{23}M_z = C_{22}F_y - C_{23}F_yL/2$$
$$= F_y\left[\frac{L^3}{3EI} + \frac{\alpha L}{Ght} - \frac{L}{2}\frac{L^2}{2EI}\right]. \tag{4.46}$$

Taking the ratio of the applied load to the displacement, the actuation stiffness (neglecting shear) is

$$k_y = \frac{F_y}{y} = \left[\frac{L^3}{12EI} + \frac{\alpha L}{Ght}\right]^{-1} \cong \frac{12EI}{L^3}. \tag{4.47}$$

Using the same method, the displacement of the flexure in the vertical direction u_z is

$$u_z = C_{44}F_z + C_{45}M_y = C_{44}F_z - C_{45}F_zL/2$$
$$= F_z\left[\frac{4L^3}{Eh^3t} + \frac{\alpha L}{Ght} - \frac{L}{2}\frac{6L^2}{Eh^3t}\right]. \tag{4.48}$$

Similarly, the vertical stiffness is

$$k_z = \left[\frac{L^3}{Eh^3t} + \frac{\alpha L}{Ght}\right]^{-1}. \tag{4.49}$$

Because of the high aspect ratio in the vertical direction, shear cannot be ignored.

Equations (4.47) and (4.49) are used to study the effect of the quantity of flexures n and flexure thickness t on the effective vertical out-of-plane stiffness $k_{z\,\text{eff}}$. To do this, the desired actuation stiffness $k_{y\,\text{eff}} = 10$ N/μm is divided amongst the number of flexures n to give the actuation stiffness for an individual flexure $k_{y\,i}$. From there, Eq. (4.47) is used to determine the length L for $t \in [0.3, 1]$ mm. The individual vertical stiffness $k_{z\,i}$ is then calculated using Eq. (4.49). The effective vertical stiffness is $k_{z\,\text{eff}} = \sum^n k_{z\,i}$. By increasing the number of flexures from 2 to 12 (1-mm thick) the vertical stiffness is increased from 76 to 226 N/μm (197 % increase). For $n = 2$, decreasing the flexure thickness from 1 to 0.3-mm thick (which effectively decreases the flexure length) increased the vertical stiffness from 76 to 79.5 N/μm (4.6 % increase). Increasing the number of flexures from 2 to 12 and decreasing the flexure thickness from 1 to 0.3-mm thick produces a vertical stiffness of 260 N/μm (242 % increase). In Fig. 4.19, the circles denote the $k_{z\,\text{eff}}$ values obtained using FEA. The FEA results follow the trend of the analytical results with the only variance being an increase in effective stiffness (average increase = 3.15 %). Increasing flexure height h also contributes to increasing vertical stiffness but at the cost of a taller stage body, which increases the mass m thus reducing the actuation resonance.

The most dramatic increase in vertical stiffness for a beam flexure is observed by increasing the number of flexures n. Decreasing the flexure thickness (and as a result, the flexure length) increases the vertical stiffness as well. But the limiting factor of

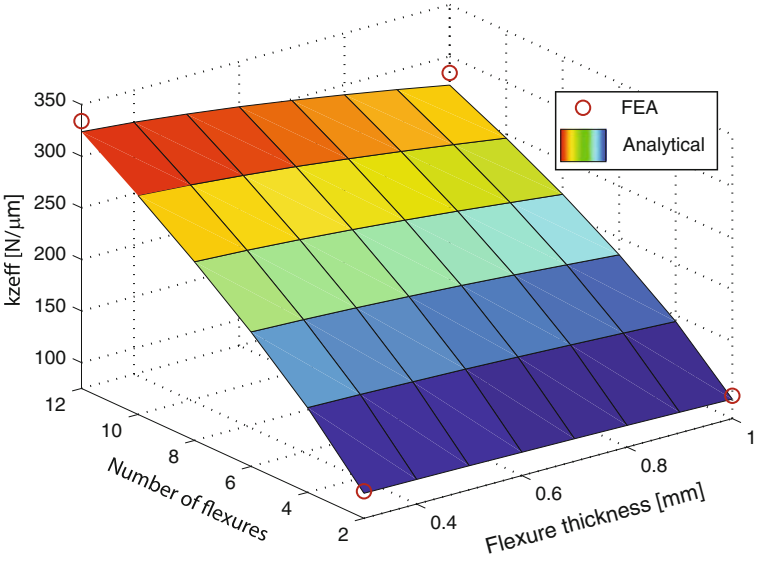


Fig. 4.19 FEA and analytical results showing effective vertical flexure stiffness $k_{z\ \mathrm{eff}}$ with respect to flexure thickness t and quantity of flexures n. Effective actuation stiffness $k_{y\ \mathrm{eff}}$ is held constant at 10 N/μm

decreasing the flexure thickness is stress. A shorter thinner beam flexure will have higher stress concentration than a longer thicker beam flexure of equal stiffness.

When a corner-filleted beam flexure, as studied above, is displaced in the actuation direction, the majority of the strain is located at the flexure ends near the fillets. Additionally, when the same flexure is displaced in the vertical direction, the majority of the vertical displacement is in shear strain located at the center cross-section. An effective way to further increase the out-of-plane stiffness of a beam flexure is to increase the thickness of the center section of the flexure, thus converting the beam flexure into a double-hinged serial flexure as shown in Fig. 4.9. Both analytical and FEA methods are used to study the vertical stiffness of the 'thickened' flexures. The cross-sectional area and second moment of inertia values in Eq. (4.41) are replaced with $A(x) = ht(x)$ and $I(x) = ht(x)^3/12$, respectively. For example, the thickness of the flexure in Fig. 4.9 is

$$t(x) = \begin{cases} t + 2\left[r - \sqrt{x(2r-x)}\right], & x \in [0, a] \\ t, & x \in [a, b] \\ t + 2\left[r - \sqrt{(l-x)(2r-l+x)}\right], & x \in [b, c] \\ t + 2r, & x \in [c, d] \\ t + 2\left[r - \sqrt{(l-g)(2r-l+g)}\right], & x \in [d, e] \\ t, & x \in [e, f] \\ t + 2\left[r - \sqrt{g(2r-g)}\right], & x \in [f, L] \end{cases} \tag{4.50}$$

Table 4.4 y-axis flexure stiffness comparison

Type	$k_{y\ \text{eff}}$ (N/μm)		$k_{z\ \text{eff}}$ (N/μm)	
	Analytical	FEA	Analytical	FEA
Filleted beam	5.82	6.00	196.6	213.9
Thickened center	5.84	5.32	237.6	238.2

where $a = r$, $b = l - r$, $c = l$, $d = L - l$, $e = d + r$, $f = L - r$, $g = L - x$, t and l are thickness and length of the thin section of the flexure, r is the fillet radius, $t + 2r = T$ is the thickness of the thickened section, and L is the length of the entire flexure. For this case, the compliance is determined by first determining the total strain energy (Eq. 4.41) while using the thickness function $t(x)$ in the area $A(x)$ and second moment of inertia $I(x)$ expressions. Again, the coordinate system is placed on the free end for simplification and to allow for direct integration as shown in (Lobontiu 2003). For instance, the total strain energy for bending due to a point load is

$$\begin{aligned} U &= \int_0^L \frac{M(x)^2}{2E\frac{ht(x)^3}{12}}\,\mathrm{d}x + \int_0^L \frac{\alpha V(x)^2}{2Ght(x)}\,\mathrm{d}x, \\ &= \frac{12F_y^2}{2Eh}\int_0^L \frac{x^2}{t(x)^3}\,\mathrm{d}x + \frac{\alpha F_y^2}{2Gh}\int_0^L \frac{1}{t(x)}\,\mathrm{d}x. \end{aligned} \tag{4.51}$$

Taking the partial derivative with respect to the applied force F_y gives the displacement

$$u_y = \frac{\partial U}{\partial F_y} = \frac{12F_y}{Eh}\int_0^L \frac{x^2}{t(x)^3}\,\mathrm{d}x + \frac{\alpha F_y}{Gh}\int_0^L \frac{1}{t(x)}\,\mathrm{d}x. \tag{4.52}$$

The in-plane (and out-of-plane) stiffness is then calculated numerically by taking the ratio of the force to deflection. Table 4.4 compares the actuation and vertical stiffness of a standard filleted flexure beam to a thickened flexure beam obtained analytically and using FEA. This comparison shows how the vertical stiffness of beam flexures similar to the ones used on the y-stage can be increased an additional 19.3 % by simply increasing the thickness of the center section. To keep the actuation stiffness $k_{y\ \text{eff}}$ constant, the length L of the thickened flexure is increased from 9.75 mm to 10.70 mm.

In summary, the effective vertical stiffness can be improved to increase the out-of-plane stiffness by (1) increasing the number of flexures n, (2) decreasing the flexure length L, and (3) thickening the center section of a beam flexure to create a serial-compliant double-hinged flexure.

Flexure placement is important to help increase rotational stiffness. Increasing the length (and width) of a stage and placing flexures at the corners of the moving

platform increase rotational stiffness of the platform. However, the cost of increasing the size of the platform is increasing overall mass, thus lowering the mechanical resonance.

The first five modes for the x- y- and z-stages are predicted using the *frequency* tool in COSMOSWorks (FEA). (Detailed discussion of the z-stage design is presented below.) It is assumed that the resonances of the y-stage would not be excited by the dynamic motion of the inner nested x-stage. This allows the design shown in Fig. 4.20a to be broken down into the low-speed y-stage (Fig. 4.20b1–b5), high-speed x-stage (Fig. 4.20c1–c5), and vertical z-stage (Fig. 4.20d1–d5). The boundary faces of each stage (shown hatched) have a fixed boundary condition. All contacting components are bonded together with compatible mesh. The meshing is done at "high quality" with refined meshing at the flexure fillets and pivot points (0.25 mm minimum element size on surfaces). The materials used and their corresponding mechanical properties are as follows:

- Aluminum: $E = 72$ GPa, $\nu = 0.33$, $\rho = 2700$ Kg/m^3;
- Steel: $E = 200$ GPa, $\nu = 0.28$, $\rho = 7800$ Kg/m^3;
- Piezo-stack: $E = 33.9$ GPa, $\nu = 0.30$, $\rho = 8000$ Kg/m^3;
- Alumina: $E = 300$ GPa, $\nu = 0.21$, $\rho = 3960$ Kg/m^3,

where the modulus for the piezo-stack was calculated from the stiffness and blocking force. The predicted first mechanical resonance for the y-, x-, and z-stage are 5.96, 25.9, and 113 kHz, respectively, all of which are in the corresponding stage actuation direction as preferred. Simulated FEA frequency response is done using the *Linear Dynamic (Harmonic)* tool in COSMOSWorks. A constant amplitude sinusoidal force is applied in the actuation direction at the corners of the piezoactuator/stage interfaces. The force generated is assumed proportional to the applied voltage. A global modal damping ratio of 0.025 is applied to simulate the damping of aluminum alloy and to produce a gain of 20 dB. Figure 4.21a1, b1 show the predicted frequency response plots for the x- and y-axis with the resonant peaks occurring at 25.9 and 5.96 kHz, respectively.

4.7.7 Design of Vertical Stage

The quick movements of the z-stage when tracking sample features such as steps, may excite the resonance modes of the nesting x-stage. To minimize impulsive forces along the vertical direction, a dual counterbalance configuration is utilized. Ando et al. (2008) describe four configurations which include face mounting, mounting both faces of the actuator to flexures, and inserting the piezoactuator in a hole and allowing the end faces to be free. Dual face-mounted z-piezoactuators are a simple and effective method for counterbalancing. However, the disadvantage is the first resonance mode for a slender piezoactuator is bending as shown in Fig. 4.22a1, instead of the desired actuation mode as illustrated in Fig. 4.22a2. Inserting the piezoactuator into a hole in

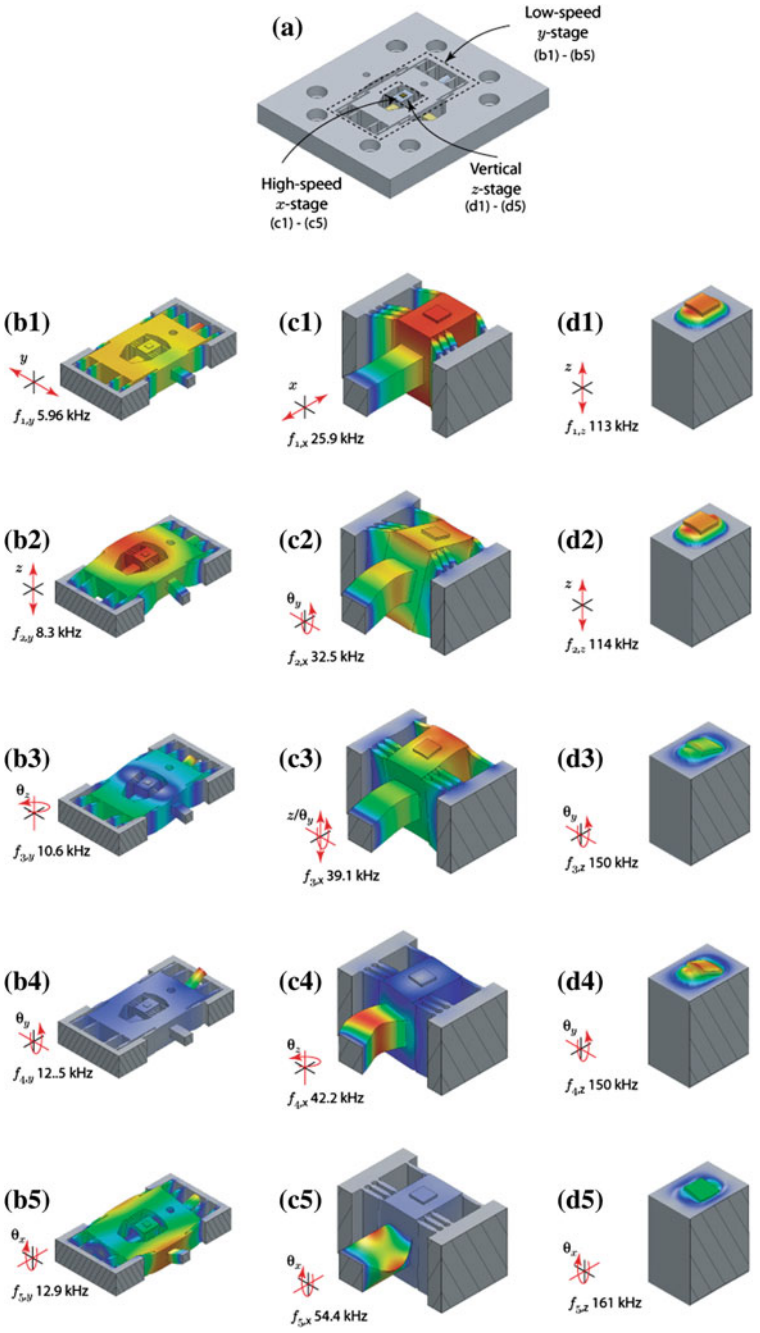


Fig. 4.20 Finite element analysis results showing first five modes: **a** high-speed scanning stage; **b1**–**b5** low-speed *y*-stage; **c1**–**c5** high-speed *x*-stage; and **d1**–**d5** *vertical z*-stage. Each stage section is designed to have the first mechanical resonance to occur in the actuation direction

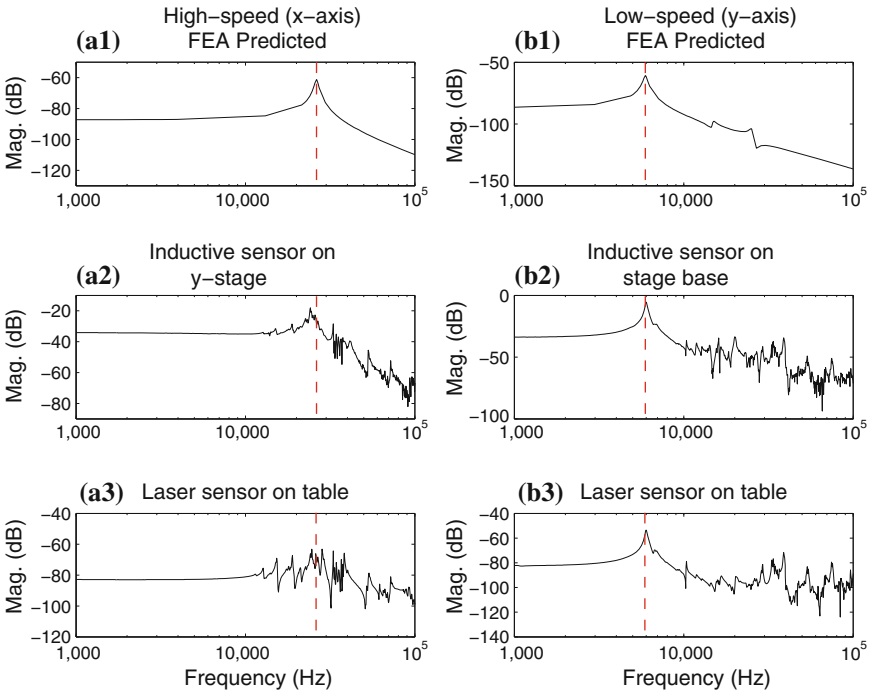


Fig. 4.21 **a1**–**a3** Comparison of predicted and measured frequency response functions for the high-speed stage (x-axis), **b1**–**b3** the low-speed stage (y-axis). The *vertical dashed line* is used to compare the experimentally measured results to the FEA predicted first resonance peak

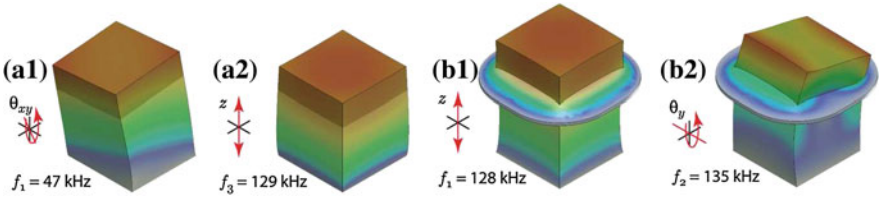


Fig. 4.22 FEA results for z-piezo with 1-mm thick sample, **a** without and **b** with flexure

the x-stage is tested, but unfortunately the design requires a long piezoactuator and did not constrain the end faces well.

A new configuration as shown in Fig. 4.23 is proposed in which a dual face-mounted piezo arrangement is combined with a compliant end plate flexure. The piezoactuators are first recessed within the nesting stage so that the free face is flush with the top surface of the stage body. The plate flexure is glued to the free end of the piezoactuator and the surrounding surface of the stage. Figure 4.22b1, b2) show how by using a plate flexure, the bending (and torsional) modes can be shifted above the frequency of the actuation mode.

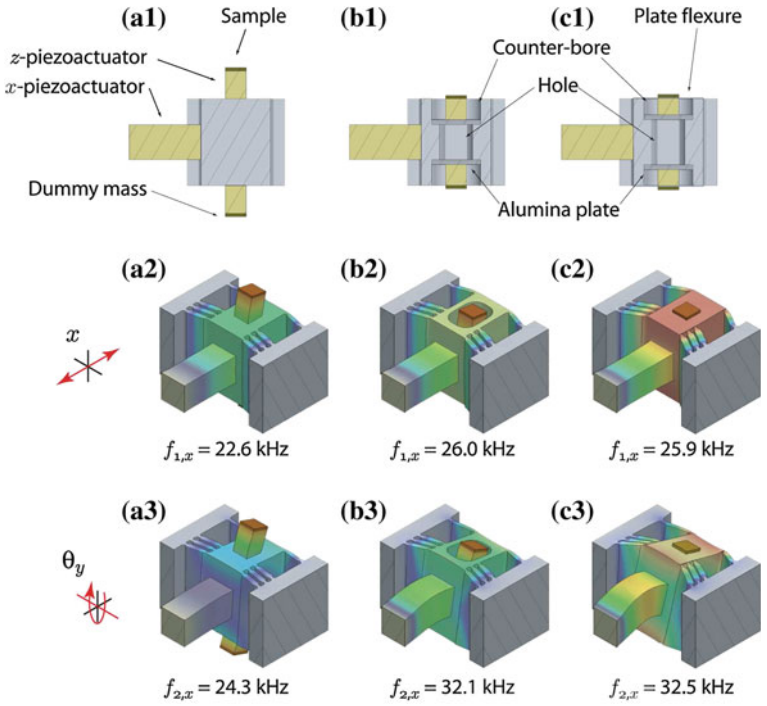


Fig. 4.23 Three configurations for the z-piezoactuator and corresponding first two modes. The dynamic characteristics of the face-mounted configuration in **a** are improved by **b** recessing the z-piezoactuator into the x-stage body and **c** adding a plate flexure to the free face of the z-actuator **c**

4.7.8 Fabrication and Assembly

The main stage body of the scanner is constructed from a single block of 7075 aluminum alloy, where the features are machined using traditional milling and wire EDM processes. The x- and y-stages are displaced with $5 \times 5 \times 10$ mm Noliac SCMAP07 piezo-stack actuators, where the motion is guided by compliant, center-thickened flexures described above. The x-flexures are designed to have a pivot point thickness of 0.5 ± 0.03 mm to produce an effective axial stiffness of 14 ± 2 N/μm. When assembling the x- and y-stages, it is important to preload the piezoactuators. Failure to preload will result in lower mechanical resonances that resemble the predicted free stage resonance (stage without piezoactuator). Preloading is accomplished by initially displacing the stages in the actuation direction and sliding the piezoactuators in place, then applying shims and glue. The tension on the stage is then released onto the piezoactuator resulting in preload. Other preload mechanisms include set-screws and spring-based mechanisms.

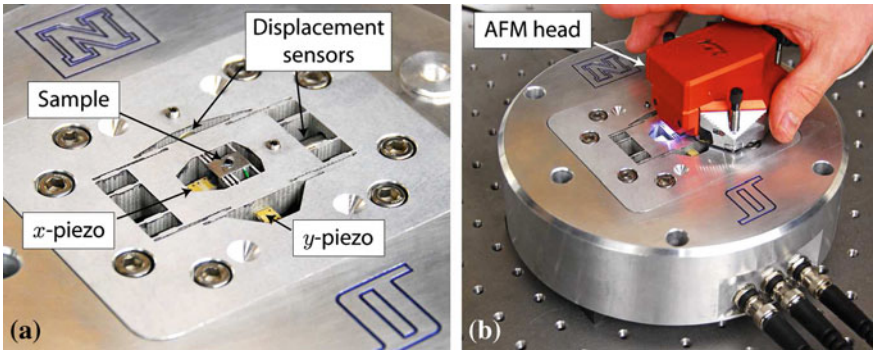


Fig. 4.24 Assembled positioner: **a** stage with sample glued to sample platform. Holes were machined into the stage body to accommodate inductive or capacitive sensors. The sensors are held in place with set screws as shown. **b** An AFM head [Nanosurf, easyScan 2 (www.nanoscience.com)] coupled with the positioning stage for high-speed AFM imaging experiments

The z-stage is designed using two 3×3 mm Noliac SCMAP06 piezo-stack actuators. The actuators are recessed within the nested x-stage. The base of each actuator is glued to an alumina plate while the free end is constrained using plate flexures. To increase the stiffness of the piezoactuators, the plate-stacks are used without the stock 1-mm thick ceramic insulating end-plates. Instead, the mounting face is insulated by the alumina base plate, while the top surface is insulated from the plate flexure with a thin sheet of mica. The experimental prototype is shown in Fig. 4.24 with the scanner body bolted to an aluminum base.

4.7.9 Drive Electronics

Due to the capacitive nature of piezoelectric transducers, high-speed operation requires large current and power dissipation. For example, the fast scanning axes, x and z, require drive electronics capable of supplying sufficient power to drive the capacitive piezoelectric loads at high frequency. If the maximum driving voltage, trajectory, and frequency are known, the current and power dissipation are easily computed by conservatively approximating the transducer as a purely capacitive load. For example, the current

$$I_p = CsV_p, \tag{4.53}$$

where s is the Laplace variable, and C and V_p are the transducer capacitance (380 nF) and load voltage, respectively. The power dissipation in a linear amplifier is

$$P_d = I_p(V_s - V_p), \tag{4.54}$$

where V_s is the supply voltage (200 V).

For the example design, the nominal capacitances for the x- and z-axis actuators are 380 and 100 nF, respectively. The piezo-amplifiers are built around the Power Amp Design (www.powerampdesign.net) PAD129 power op-amp, with a gain bandwidth product of 1 MHz. A 200 V DC power supply is constructed from two linear regulated 100 V, 3 A DC power supplies (Acopian A100HT300) connected in series. A 35 kHz low pass filter is used to smooth the input signal to the power op-amp. Commercially available high-bandwidth amplifiers are available from suppliers such as PiezoDrive, Australia; Trek Inc., Japan; and PiezoMechanik, Germany. A more detailed discussion of electrical considerations is presented in Chap. 14.

4.7.10 Experimental Results

The fabricated scanner shown in Fig. 4.24 is tested to determine the stiffness, maximum range, and dynamic characteristics. Prior to assembly, the effective stiffnesses of the x- and y-stage are determined by taking the ratio of the measured displacement due to an applied load. Static loads are applied to the stages by mounting the scanner vertically to a fixture (z-axis perpendicular to ground), running a cable through the hole in the x-stage and hanging masses from the cable. A total of 15 Lbf (66.7 N) is applied in the positive and negative direction in 2.5 Lbf (11.1 N) increments. Displacement is measured using a Kaman inductive sensor (SMU9000-15N). The analytical, FEA predicted, and measured stiffnesses are 7.82, 7.42, and 3.81 N/μm, respectively, for the x-stage and 4.28, 4.04, and 5.10 N/μm, respectively, for the y-stage. The discrepancy between the predicted and measured values are attributed to machining tolerances.

Application of 180 V peak-to-peak sine input at 10 Hz to the x and y piezoactuators resulted in 8.19 and 8.34 μm travel, respectively. Since the x and y axes can tolerate a maximum of 200 V, the maximum lateral range of the stage is approximately 9×9 μm. Application of 200 V peak-to-peak sine input at 10 Hz to the vertical z piezoactuators will give approximately 1 μm of travel. Over these ranges, the measured x/y cross coupling is 75 nm peak-to-peak (1.83 % or −34.75 dB) in y caused by actuating the x piezo and 24 nm peak-to-peak (0.6 % or −44.44 dB) in x caused by actuating the y piezo. The measured vertical runouts are 27.6 nm peak-to-peak (0.35 % or −49.2 dB) caused by actuating the x piezo and 81.4 nm peak-to-peak (0.97 % or −40.3 dB) caused by actuating the y piezo. It is noted that the lateral x-to-y and y-to-x cross coupling may be caused by the y-stage's compliance. For example, x-actuation may cause slight deformation in the y-stage body leading to measured cross coupling. Similarly, lateral-to-vertical cross coupling (x-to-z and y-to-z) may be caused by a tilted sample or misalignment of the displacement sensor, e.g., when the tilted sample translates laterally, the tilted surface may appear to move vertically relative to a fixed sensor.

Frequency response functions are measured using a dynamic signal analyzer (Stanford Research Systems SRT785). Small inputs (<70 mV) are applied to the piezo amplifiers during the test to minimize the effect of nonlinearity such as

hysteresis. Measurements for the x- and y-stages are taken with both the stage mounted sensors (inductive sensor for x, and ADE Capacitive sensor for y) and again with a single point laser vibrometer (Polytec CLV-1000 with CLV-800-vf40 laser unit) mounted to the vibration isolation table (for both x and y). The measured responses are shown in Fig. 4.21 along with the FEA predictions. When measured relative to the y-stage body, the x-stage has a dominant first resonance peak at 24.2 kHz (a2) which matches well with the predicted value of 25.9 kHz (a1). Several small pole/zero pairs appear before the dominant peak. However, when measured using a laser vibrometer relative to an outside body such as the vibration isolation table (a3), the response shows additional unexpected resonances. These peaks are thought to be due to modes in the y-stage being excited by the x-stage. Unfortunately, these modes are not detectable when the sensor is attached to the y-stage body. The measured dominant resonance for the y-stage at 6.0 kHz, both measured using the capacitive displacement sensor attached to the stage body (a2) and the laser vibrometer on table (a3) matches the predicted FEA value (a1) at 5.96 kHz very well. Not only do the dominant resonances agree with the FEA results, they are also piston modes relative to their mounting point as predicted by FEA. The frequency response for the z-axis is measured using the deflection of a 360 kHz tapping-mode AFM cantilever (Vista Probes T300 www.vistaprobes.com) in contact-mode over the sample surface. The dominant resonance is approximately 70 kHz in the actuation (piston) mode.

4.8 Chapter Summary

This chapter described the design considerations for high-speed nanopositioning. An example three-axis, serial-kinematic high-speed scanner based on piezo-stack actuators is used to illustrate the design process. Important considerations include:

- balancing the tradeoff between scanning range and achievable mechanical resonance,
- taking advantage of FEA tools to optimize mechanical resonances, and
- designing drive electronics which considers the capacitive nature of piezoactuators.

The example scanner achieved a range of approximately $9 \times 9 \times 1$ μm, where the fast scanning axis is optimized for speed. Experimental results showed a good correlation with simulation, where finite element analysis predicted the dominant resonances along the fast (x-axis) and slow (y-axis) scanning axes at 25.9 and 6.0 kHz, respectively. The measured dominant resonances of the prototype stage in the fast and slow scanning directions were measured at 24.2 and 6.0 kHz, respectively, which were in good agreement with the FEA predictions. In the z-direction, the measured dominant resonance was measured at approximately 70 kHz. This is sufficient to achieve SPM line rates in excess of 3 kHz.

References

Ando T, Kodera N, Uchihashi T, Miyagi A, Nakakita R, Yamashita H, Matada K (2005) High-speed atomic force microscopy for capturing dynamic behavior of protein molecules at work. E-J Surf Sci Nanotechnol 3:384–392

Ando T, Uchihashi T, Fukuma T (2008) High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes. Prog Surf Sci 83(7–9):337–437

Ando T, Kodera N, Maruyama D, Takai E, Saito K, Toda A (2002) A high-speed atomic force microscope for studying biological macromolecules in action. Jpn J Appl Phys Part 1. 41(7B): 4851–4856

Bechtold R, Rudnyi E, Korvink J (2005) Dynamic electro-thermal simulation of microsystems-a review. J Micromech Microeng 15(11):R17–R31

Beer FP, Johnston ER (1992) Mechanics of materials, 2nd edn. McGraw Hill, New York

Bell DJ, Lu TJ, Fleck NA, Spearing SM (2005) MEMS actuators and sensors: observations on their performance and selection for purpose. J Micromech Microeng 15(7):S153–S164

Clayton GM, Tien S, Leang KK, Zou Q, Devasia S (2009) A review of feedforward control approaches in nanopositioning for high-speed SPM. J Dyn Syst Meas Contr 131:061 101(1–19)

Craig RR (2000) Mechanics of materials, 2nd edn. John Wiley & Sons, New York

Damjanovic D, Newnham RE (1992) Electrostrictive and piezoelectric materials for actuator applications. J Intell Mater Syst Struct 3(2):190–208

Devos S, Reynaerts D, Brussel HV (2008) Minimising heat dissipation in ultrasonic piezomotors by working in a resonant mode. Precis Eng 32:114–125

Fleming AJ (2009) High-speed vertical positioning for contact-mode atomic force microscopy. In: Proceedings of IEEE/ASME international conference on advanced intelligent mechatronics, Singapore, pp 522–527

Geisberger A, Sarkar N (2006) MEMS/NEMS: techniques in microelectrothermal actuator and their applications. Springer, New York

Guthold M, Zhu X, Rivetti C, Yang G, Thomson NH, Kasas S, Hansma HG, Smith B, Hansma PK, Bustamante C (1999) Real-time imaging of one-dimensional diffusion and transcription by escherichia coli rna polymerase. Biophys J 77(4):2284–2294

Hicks TR, Atherton PD, Xu Y, McConnell M (1997) The nanopositioning book. Queensgate Intstruments Ltd., Berkshire

Howell LL (2001) Compliant mechanisms. John Wiley & Sons, New York

Hubbard NB, Culpepper ML, Howell LL (2006) Actuators for micropositioners and nanopositioners. Appl Mech Rev 59(6):324–334

Humphris ADL, Hobbs JK, Miles MJ (2003) Ultrahigh-speed scanning near-field optical microscopy capable of over 100 frames per second. Appl Phys Lett 83(1):6–8

Inman D (2001) Engineering vibration, 2nd edn. Prentice Hall, Upper Saddle River

Kenton BJ (2010) Design, characterization, and control of a high-bandwidth serial-kinematic nanopositioning stage for scanning probe microscopy applications. Ph.D. dissertation, Mechanical Engineering

Kenton BJ, Leang KK (2012) Design and control of a three-axis serial-kinematic high-bandwidth nanopositioner. IEEE/ASME Trans Mech 17(2):356–369

Kindt JH, Fantner GE, Cutroni JA, Hansma PK (2004) Rigid design of fast scanning probe microscopes using finite element analysis. Ultramicroscopy 100(3–4):259–265

Leang KK, Fleming AJ (2009) High-speed serial-kinematic AFM scanner: design and drive considerations. Asian J Control Spec Issue Adv Control Methods Scan Probe Microsc Res Tech 11(2):144–153

Liu C (2006) Foundations of MEMS. Prentice Hall, Upper Saddle River

Lobontiu N (2003) Compliant mechanisms design of flexure hinges. CRC Press LLC, Boca Raton

Muller K-D, Marth H, Pertsch P, Stiebel C, Zhao X (2007) Piezo based long travel actuators in special environmental conditions. In: 12th European space mechanisms and tribology symposium (ESMATS), pp SP–653

Park SR (2005) A mathematical approach for analyzing ultra precision positioning system with compliant mechanism. J Mater Res Process Technol 164–165:1584–1589
Physik Instrumente (2009) Piezo nano positionoing: Inspirations 2009
Picco LM, Bozec L, Ulcinas A, Engledew DJ, Antognozzi M, Horton M, Miles MJ (2007) Breaking the speed limit with atomic force microscopy. Nanotechnology 18(4):044 030(1–4)
Rifai OME, Youcef-Toumi K (2001) Coupling in piezoelectric tube scanners used in scanning probe microscopes. In: American Control Conference, vol 4, pp 3251–3255
Rost MJ, Crama L, Schakel P, van Tol E, van Velzen-Williams GBEM, Overgauw CF, ter Horst H, Dekker H, Okhuijsen B, Seynen M, Vijftigschild A, Han P, Katan AJ, Schoots K, Schumm R, van Loo W, Oosterkamp TH, Frenken JWM (2005) Scanning probe microscopes go video rate and beyond. Rev Sci Instrum 76(5):053 710–1–053 710–9
Sahu B, Taylor CR, Leang KK (2010) Emerging challenges of microactuators for nanoscale positioning, assembly, and manipulation", ASME J Manuf Sci Eng Special Issue Nanomanuf 132(3):030917 (16 pages)
Samara-Ratna P, Atkinson H, Stevenson T, Hainsworth SV, Sykes J (2007) Design of a micromanipulation system for high temperature operation in an environmental scanning electron microscope (ESEM). J Micromech Microeng 17(1):104–114
Schilfgaarde MV, Abrikosov IA, Johansson B (1999) Origin of the Invar effect in iron-nickel alloys. Nature 400(6739):46–49
Schitter G (2007) Advanced mechanical design and control methods for atomic force microscopy in real-time. In: American Control Conference, 2007, pp 3503–3508
Schitter G, Stemmer A (2004) Identification and open-loop tracking control of a piezoelectric tube scanner for high-speed scanning-probe microscopy. IEEE Trans Control Syst Technol 12(3): 449–454
Schitter G, Rost MJ (2008) Scanning probe microscopy at video-rate. Mater Today 11(1):40–48
Schitter G, Thurner PJ, Hansma PK (2008) Design and input-shaping control of a novel scanner for high-speed atomic force microscopy. Mechatronics 18(5–6):282–288
Schitter G, Åström KJ, DeMartini BE, Thurner PJ, Turner KL, Hansma PK (2007) Design and modeling of a high-speed AFM-scanner. IEEE Trans Control Syst Technol 15(5):906–915
Schitter G, Rijkee WF, Phan N (2008) Dual actuation for highbandwidth nanopositioning. In: IEEE conference on decision and control, 2008, pp 5176–5181
Scire FE, Teague EC (1978) Piezodriven 50-μm range stage with subnanometer resolution. Rev Sci Instr 49(12):1735–1740
Smith T (2000) Flexures: elements of elastic mechanisms. Gordon and Breach, Amsterdan
Snow ES, Campbell PM, Perkins FK (1997) Nanofabrication with proximal probes. Proc IEEE 85(4):601–611
Stillesjo F, Engdahl G, Bergqvist A (1998) A design technique for magnetostrictive actuators with laminated active material. IEEE Trans Magn 34(4):2141–2143
Tan X, Baras JS (2004) Modeling and control of hysteresis in magnetostrictive actuators. Automatica 40:1469–1480
Timoshenko SP (1953) History of strength of materials. McGraw-Hill Book Company, New York
Tsodikov SF, Rakhovsky VI (1998) Magnetostrictive force actuators for superprecise positioning. In: International symposium on discharges and electrical insulation in vacuum (ISDEIV), vol 2, pp 713–719
Waram T (1993) Actuator design using shape memory alloys, 2nd edn. T. C. Waram, Ontario
Yong YK, Aphale SS, Moheimani SOR (2009) Design, identification, and control of a flexure-based xy stage for fast nanoscale positioning. IEEE Trans Nanotechnol 8(1):46–54
Yong Y, Moheimani SOR, Kenton BJ, Leang KK (2012) Invited review: high-speed flexure-guided nanopositioning: mechanical design and control issues. Rev Sci Instrum 83(12):121101
Young WC, Budynas RG (2002) Roark's formula for stress and strain, 7th edn. McGraw-Hill, New York

Chapter 5
Position Sensors

Position sensors with nanometer resolution are a key component of many precision imaging and fabrication machines. Since the sensor characteristics can define the linearity, resolution and speed of a nanopositioner, the sensor performance is a foremost consideration. The first goal of this chapter is to define concise performance metrics and to provide exact and approximate expressions for error sources including nonlinearity, drift, and noise. The second goal is to review current position sensor technologies and to compare their performance. The sensors considered include: resistive, piezoelectric and piezoresistive strain sensors; capacitive sensors; electrothermal sensors; eddy current sensors; linear variable displacement transformers; interferometers and linear encoders.

5.1 Introduction

The sensor requirements of a nanopositioning system are among the most demanding of any control system. The sensors must be compact, high-speed, immune to environmental variation, and able to resolve position down to the atomic scale. In many applications, such as Atomic Force Microscopy (Abramovitch et al. 2007; Salapaka and Salapaka 2008) or nanofabrication (Tseng 2008; Vicary and Miles 2008), the performance of the machine or process is primarily dependent on the performance of the position sensor, thus, sensor optimization is a foremost consideration.

In order to define the performance of a position sensor, it is necessary to have strict definitions for the characteristics of interest. At present, terms such as accuracy, precision, nonlinearity, and resolution are defined loosely and often vary between manufacturers and researchers. The lack of a universal standard makes it difficult to predict the performance of a particular sensor from a set of specifications. Furthermore, specifications may not be in a form that permits the prediction of closed-loop performance.

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_5, © Springer International Publishing Switzerland 2014

This chapter provides concise definitions for the linearity, drift, bandwidth, and resolution of position sensors. The measurement errors resulting from each source are then quantified and bounded to permit a straightforward comparison between sensors. An emphasis is placed on specifications that allow the prediction of closed-loop performance as a function of the controller bandwidth.

Although there are presently no international standards for the measurement or reporting of position sensor performance, this chapter is aligned with the definitions and methods reported in the ISO/IEC 98:1993 Guide to the Expression of Uncertainty in Measurement (ISO/IEC 1994), and the ISO 5723 Standard on Accuracy (Trueness and Precision) of Measurement Methods and Results (ISO 1994).

The noise and resolution of a position sensor is potentially one of the most misreported sensor characteristics. The resolution is commonly reported without mention of the bandwidth or statistical definition and thus has little practical value.

To improve the understanding of this issue, the relevant theory of stochastic processes is reviewed in Sect. 5.2. The variance is then utilized to define a concise statistical description of the resolution, which is a straightforward function of the noise density, bandwidth, and $1/f$ corner frequency.

The second goal of this chapter is to provide a tutorial introduction and comparison of sensor technologies suitable for nanopositioning applications. To be eligible for inclusion, a sensor must be capable of a 6σ-resolution better than 10 nm with a bandwidth greater than 10 Hz. The sensor cannot introduce friction or contact forces between the reference and moving target, or exhibit hysteresis or other characteristics that limit repeatability.

The simplest sensor considered is the metal foil strain gauge discussed in Sect. 5.3.1. These devices are often used for closed-loop control of piezoelectric actuators but are limited by temperature dependence and low sensitivity (Schitter et al. 2002). Piezoresistive and piezoelectric strain sensors provide improved sensitivity but at the cost of stability and DC performance.

The most commonly used sensors in nanopositioning systems (Devasia et al. 2007) are the capacitive and eddy-current sensors discussed in Sects. 5.3.4 and 5.3.6. Capacitive and eddy-current sensors are more complex than strain sensors but can be designed with subnanometer resolution, albeit with comparably small range and low bandwidth. They are used extensively in applications such as atomic force microscopy (Salapaka and Salapaka 2008; Leang et al. 2009; Fleming et al. 2010a, b) and nanofabrication (Tseng et al. 2008; Vicary and Miles 2008). The Linear Variable Displacement Transformer (LVDT) described in Sect. 5.3.7 is a similar technology that is intrinsically linear. However, this type of sensor is larger than a capacitive sensor and due to the larger range, is not as sensitive.

To achieve high absolute accuracy over a large range, the reference standard is the laser heterodyne interferometer discussed in Sect. 5.3.8. Although bulky and costly, the interferometer has been the sensor of choice for applications such as IC wafer steppers (Butler 2011; Mishra et al. 2007) and metrological systems (Merry et al. 2009). New fiber interferometers are also discussed that are extremely compact and ideal for extreme environments.

Aside from the cost and size, the foremost difficulties associated with an interferometer are the susceptibility to beam interference, variation in the optical medium, and alignment error. Since an interferometer is an incremental position sensor, if the beam is broken or the maximum traversing speed is exceeded, the system must be returned to a known reference before continuing. These difficulties are somewhat alleviated by the absolute position encoders described in Sect. 5.3.9. A position encoder has a read-head that is sensitive to a geometric pattern encoded on a reference scale. Reference scales operating on the principle of optical interference can have periods of 128 nm and a resolution of a few nanometers.

Other sensor technologies that were considered but did not fully satisfy the eligibility criteria include optical triangulation sensors (Shan et al. 2008), hall effect sensors, and magnetoresistive sensors. In general, optical triangulation sensors are available in ranges from 0.5 mm to 1 m with a maximum resolution of approximately 100 nm. Hall effect sensors are sensitive to magnetic field strength and hence the distance from a known magnetic source. These sensors have a high resolution, large range, and wide bandwidth but are sensitive to external magnetic fields and exhibit hysteresis of up to 0.5 % which degrades the repeatability. The magnetoresistive sensor is similar except that the resistance, rather than the induced voltage, is sensitive to magnetic field. Although typical anisotropic magnetoresistive (AMR) sensors offer similar characteristics to the Hall effect sensor, recent advances stimulated by the hard disk industry have provided major improvements (Parkin et al. 2003). In particular, the giant magnetoresistive effect (GMR) can exhibit two orders of magnitude greater sensitivity than the AMR effect which equates to a resistance change of up to 70 % at saturation. Such devices can also be miniaturized and are compatible with lithographic processes. Packaged GMR sensors in a full-bridge configuration are now available from NVE Corporation, NXP Semiconductor, Siemens, and Sony. Aside from the inherent nonlinearities associated with the magnetic field, the major remaining drawback is the hysteresis of up to 4 % which can severely impact the performance in nanopositioning applications. Despite this, miniature GMR sensors have shown promise in nanopositioning applications by keeping the changes in magnetic field small (Sahoo et al. 2011; Kartik et al. 2012). However, to date, the linearity and hysteresis of this approach has not been reported.

5.2 Sensor Characteristics

5.2.1 Calibration and Nonlinearity

Position sensors are designed to produce an output that is directly proportional to the measured position. However, in reality, all position sensors have an unknown offset, sensitivity, and nonlinearity. These effects must be measured and accounted for in order to minimize the uncertainty in position.

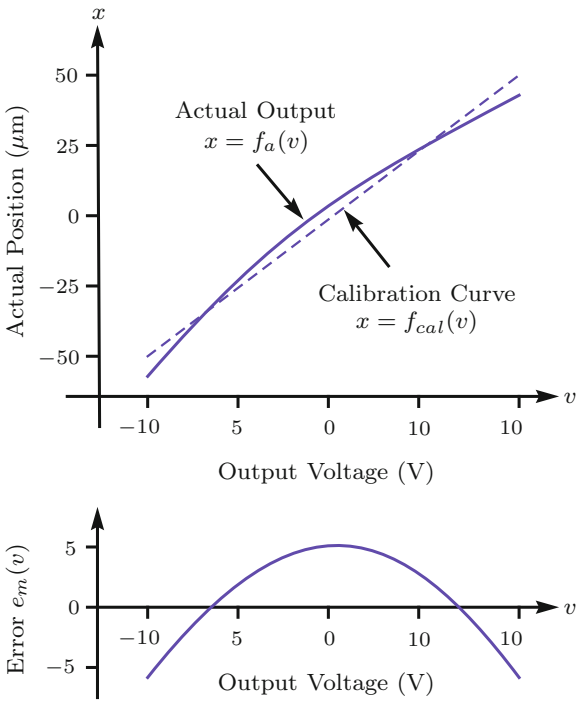


Fig. 5.1 The actual position versus the output voltage of a position sensor. The calibration function $f_{cal}(v)$ is an approximation of the sensor mapping function $f_a(v)$ where v is the voltage resulting from a displacement x. $e_m(v)$ is the residual error

The typical output voltage curve for a capacitive position sensor is illustrated in Fig. 5.1. A nonlinear function $f_a(v)$ maps the output voltage v to the actual position x. The calibration process involves finding a curve $f_{cal}(v)$ that minimizes the mean-square error, known as the least-squares fit, defined by

$$\theta^* = \arg\min \sum_{i=1}^{N} [x_i - f_{cal}(\theta, v_i)]^2 , \tag{5.1}$$

where v_i and x_i are the data points and θ^* is the vector of optimal parameters for $f_{cal}(\theta, v)$. The simplest calibration curve, as shown in Fig. 5.1, is a straight line of best fit,

$$f_{cal}(v) = \theta_0 + \theta_1 v. \tag{5.2}$$

In the above equation, the sensor offset is θ_0 and the sensitivity is θ_1 μm/V. More complex mapping functions are also commonly used, including the higher order polynomials

$$f_{cal}(v) = \theta_0 + \theta_1 v + \theta_2 v^2 + \theta_3 v^3 \cdots \tag{5.3}$$

Once the calibration function $f_{cal}(v)$ is determined, the actual position can be estimated from the measured sensor voltage. Since the calibration function does not perfectly describe the actual mapping function $f_a(v)$, a mapping error results. The mapping error $e_m(v)$ is the residual of (5.1), that is

$$e_m(v) = f_a(v) - f_{cal}(\theta^*, v). \tag{5.4}$$

If $e_m(v)$ is positive, the true position is greater than the estimated value and vice-versa. Although the mapping error has previously been defined as the peak-to-peak variation of $e_m(v)$ (Hicks et al. 1997), this may underestimate the positioning error if $e_m(v)$ is not symmetric. A more conservative definition of the mapping error (e_m) is

$$e_m = \pm \max |e_m(v)| \tag{5.5}$$

It is also possible to specify an unsymmetrical mapping error such as $+\max e_m(v)$, $-\min e_m(v)$ however, this is more complicated. For the sake of comparison, the maximum mapping error (nonlinearity) is often quoted as a percentage of the full-scale range (FSR), for example

$$\text{Mapping Error } (\%) = \pm 100 \frac{\max |e_m(v)|}{\text{FSR}}. \tag{5.6}$$

Since there is no exact consensus on the reporting of nonlinearity, it is important to know how the mapping error is defined when evaluating the specifications of a position sensor. A less conservative definition than that stated above may exaggerate the accuracy of a sensor and lead to unexplainable position errors. It may also be necessary to consider other types of nonlinearity such as hysteresis (Nyce 2004). However, sensors that exhibit hysteresis have poor repeatability and are generally not considered for precision sensing applications.

5.2.2 Drift and Stability

In addition to the nonlinearity error discussed above, the accuracy of a positioning sensor can also be severely affected by changes in the mapping function $f_a(v)$. The parameters of $f_a(v)$ may drift over time, or be dependent on environmental conditions such as temperature, humidity, dust, or gas composition. Although, the actual parametric changes in $f_a(v)$ can be complicated, it is possible to bound the variations by an uncertainty in the sensitivity and offset. That is,

$$f_a(v) = (1 + k_s) f_a^*(v) + k_o, \tag{5.7}$$

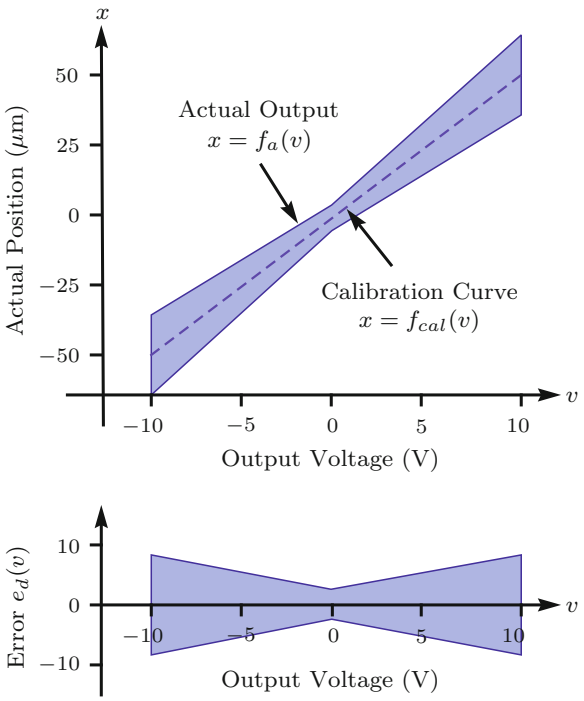


Fig. 5.2 The worst-case range of a linear mapping function $f_a(v)$ for a given error in sensitivity and offset. In this example the greatest error occurs at the maximum and minimum of the range

where k_s is the sensitivity variation usually expressed as a percentage, k_o is the offset variation, and $f_a^*(v)$ is the nominal mapping function at the time of calibration. With the inclusion of sensitivity variation and offset drift, the mapping error is

$$e_d(v) = (1 + k_s) f_a^*(v) + k_o - f_{cal}(v). \tag{5.8}$$

Equations (5.7) and (5.8) are illustrated graphically in Fig. 5.2. If the nominal mapping error is assumed to be small, the expression for error can be simplified to

$$e_d(v) = k_s f_{cal}(v) + k_o. \tag{5.9}$$

That is, the maximum error due to drift is

$$e_d = \pm \left(k_s \max |f_{cal}(v)| + k_o\right). \tag{5.10}$$

Alternatively, if the nominal calibration cannot be neglected or if the shape of the mapping function actually varies with time, the maximum error due to drift must be evaluated by finding the worst-case mapping error defined in (5.5).

5.2.3 Bandwidth

The bandwidth of a position sensor is the frequency at which the magnitude of the transfer function $v(s)/x(s)$ drops by 3 dB. Although the bandwidth specification is useful for predicting the resolution of a sensor, it reveals very little about the measurement errors caused by sensor dynamics. For example, a sensor phase-lag of only 12 degrees causes a measurement error of 10 % FSR.

If the sensitivity and offset have been accounted for, the frequency domain position error is

$$e_{bw}(s) = x(s) - v(s), \tag{5.11}$$

which is equal to

$$e_{bw}(s) = x(s)\left(1 - P(s)\right), \tag{5.12}$$

where $P(s)$ is the sensor transfer function and $(1 - P(s))$ is the multiplicative error. If the actual position is a sine wave of peak amplitude A, the maximum error is

$$e_{bw} = \pm A\left|1 - P(s)\right|. \tag{5.13}$$

The worst-case error occurs when $A = \text{FSR}/2$, in this case,

$$e_{bw} = \pm\frac{\text{FSR}}{2}\left|1 - P(s)\right|. \tag{5.14}$$

The error resulting from a Butterworth response is plotted against normalized frequency in Fig. 5.3. Counter to intuition, the higher order filters produce more error, which is surprising because these filters have faster roll-off, however, they also contribute more phase-lag. If the poles of the filter are assumed to be equal to the cut-off frequency, the low-frequency magnitude of $|1 - P(s)|$ is approximately

$$|1 - P(s)| \approx n\frac{f}{f_c}, \tag{5.15}$$

where n is the filter order and f_c is the bandwidth. The resulting error is approximately

$$e_{bw} \approx \pm A\, n\frac{f}{f_c}. \tag{5.16}$$

That is, the error is proportional to the magnitude of the signal, filter order, and normalized frequency. This is significant because the sensor bandwidth must be significantly higher than the operating frequency if dynamic errors are to be avoided. For example, if an absolute accuracy of 10 nm is required when measuring a signal with an amplitude of 100 μm, the sensor bandwidth must be ten-thousand times greater than the signal frequency.

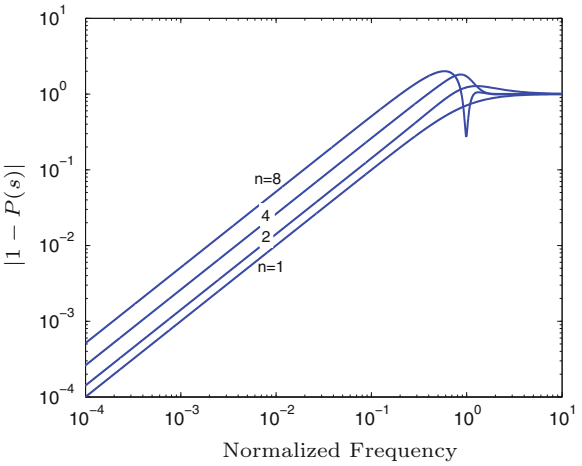


Fig. 5.3 The magnitude of error caused by the sensor dynamics $P(s)$. The frequency axis is normalized to the sensor 3 dB bandwidth. Lower order sensor dynamics result in lower error but typically result in significantly lesser bandwidths. In this example the dynamics are assumed to be n^{th} order Butterworth

In the above derivation, the position signal was assumed to be sinusoidal, for different trajectories, the maximum error must be found by simulating Eq. (5.12). Although the RMS error can be found analytically by applying Parseval's equality, there is no straightforward method for determining the peak error, aside from numerical simulation. In general, signals that contain high-frequency components, such as square and triangle waves cause the greatest peak error.

5.2.4 Noise

In addition to the actual position signal, all sensors produce some additive measurement noise. In many types of sensors, the main source of noise is from the thermal noise of resistors and the voltage and current noise in conditioning circuit transistors. As these noise processes can be approximated by Gaussian random processes, the total measurement noise can also be approximated by a Gaussian random process.

A Gaussian random process produces a signal with normally distributed values that are correlated between instances of time. We also assume that the noise process is zero-mean and that the statistical properties do not change with time, that is, the noise process is stationary. A Gaussian noise process can be described by either the autocorrelation function or the power spectral density. The autocorrelation function of a random process $\mathcal{X}$ is

$$R_{\mathcal{X}}(\tau) = E\left[\mathcal{X}(t)\mathcal{X}(t+\tau)\right], \tag{5.17}$$

where E is the expected value operator. The autocorrelation function describes the correlation between two samples separated in time by τ. Of special interest is $R_{\mathcal{X}}(0)$ which is the variance of the process. The variance of a signal is the expected value of the varying part squared. That is,

$$\text{Var } \mathcal{X} = E\left[\left(\mathcal{X} - E\left[\mathcal{X}\right]\right)^2\right]. \tag{5.18}$$

Another term used to quantify the dispersion of a random process is the standard deviation σ which is the square-root of variance,

$$\sigma_{\mathcal{X}} = \text{Standard deviation of } \mathcal{X} = \sqrt{\text{Var } \mathcal{X}} \tag{5.19}$$

The standard deviation is also the Root-Mean-Square (RMS) value of a zero-mean random process. Further properties of the variance and standard deviation can be found in Chap. 13.

The power spectral density $S_{\mathcal{X}}(f)$ of a random process represents the distribution of power or variance across frequency f. For example, if the random process under consideration was measured in Volts, the power spectral density would have the units of V^2/Hz. The power spectral density can be found by either the averaged periodogram technique or from the autocorrelation function. The periodogram technique involves averaging a large number of Fourier transforms of a random process,

$$2 \times E\left[\frac{1}{T}\left|\mathscr{F}\left\{\mathcal{X}_T(t)\right\}\right|^2\right] \Rightarrow S_{\mathcal{X}}(f) \;\; as \; T \Rightarrow \infty. \tag{5.20}$$

This approximation becomes more accurate as T becomes larger and more records are used to compute the expectation. In practice, $S_{\mathcal{X}}(f)$ is best measured using a Spectrum or Network Analyzer, these devices compute the approximation progressively so that large time records are not required. Practical techniques for the measurement of power spectral density are discussed in Sect. 13.7. The power spectral density can also be computed from the autocorrelation function. The relationship between the autocorrelation function and power spectral density is known as the Wiener-Khinchin relations, given by

$$S_{\mathcal{X}}(f) = 2\mathscr{F}\left\{R_{\mathcal{X}}(\tau)\right\} = 2\int_{-\infty}^{\infty} R_{\mathcal{X}}(\tau)e^{-j2\pi f\tau}\, d\tau, \text{ and} \tag{5.21}$$

$$R_{\mathcal{X}}(\tau) = \frac{1}{2}\mathscr{F}^{-1}\left\{S_{\mathcal{X}}(f)\right\} = \frac{1}{2}\int_{-\infty}^{\infty} S_{\mathcal{X}}(f)e^{j2\pi f\tau}\, df, \tag{5.22}$$

If the power spectral density is known, the variance of the generating process can be found from the area under the curve, that is

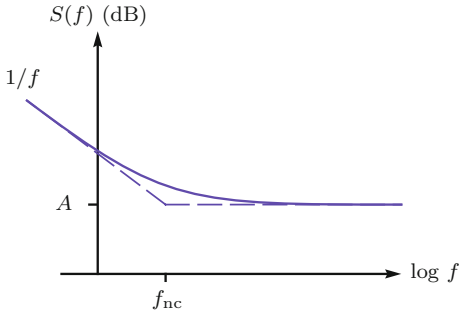


Fig. 5.4 A constant power spectral density that exhibits $1/f$ noise at low frequencies. The *dashed lines* indicate the asymptotes

$$\sigma_{\mathcal{X}}^2 = E\left[\mathcal{X}^2(t)\right] = R_{\mathcal{X}}(0) = \int_0^\infty S_{\mathcal{X}}(f)\, df, \tag{5.23}$$

Rather than plotting the frequency distribution of power or variance, it is often convenient to plot the frequency distribution of the standard deviation, which is referred to as the spectral density. It is related to the standard power spectral density function by a square-root, that is,

$$\text{Spectral density} = \sqrt{S_{\mathcal{X}}(f)}. \tag{5.24}$$

The units of $\sqrt{S_{\mathcal{X}}(f)}$ are units/$\sqrt{\text{Hz}}$ rather than units^2/Hz. The spectral density is preferred in the electronics literature as the RMS value of a noise process can be determined directly from the noise density and effective bandwidth. For example, if the noise density is a constant c $V/\sqrt{\text{Hz}}$ and the process is perfectly band limited to f_c Hz, the RMS value or standard deviation of the resulting signal is $c\sqrt{f_c}$. To distinguish between power spectral density and noise density, A is used for power spectral density and $\sqrt{A}$ is used for noise density. An advantage of the spectral density is that a gain k applied to a signal $u(t)$ also scales the spectral density by k. This differs from the standard power spectral density function that must be scaled by k^2.

Since the noise in position sensors is primarily due to thermal noise and $1/f$ (flicker) noise, the power spectral density can be approximated by

$$S(f) = A\frac{f_{\text{nc}}}{|f|} + A, \tag{5.25}$$

where A is power spectral density and f_{nc} is the noise corner frequency illustrated in Fig. 5.4. The variance of this process can be found by evaluating Eq. (5.23). That is,

$$\sigma^2 = \int_{f_l}^{f_h} A\frac{f_{\text{nc}}}{|f|} + A\, df. \tag{5.26}$$

where f_l and f_h define the bandwidth of interest. Extremely low-frequency noise components are considered to be drift. In positioning applications, f_l is typically chosen between 0.01 and 0.1 Hz. By solving Eq. (5.26), the variance is

$$\sigma^2 = A f_{\text{nc}} \ln \frac{f_h}{f_l} + A(f_h - f_l). \tag{5.27}$$

If the upper frequency limit is due to a linear filter and $f_h >> f_l$, the variance can be modified to account for the finite roll-off of the filter, that is

$$\sigma^2 = A f_{\text{nc}} \ln \frac{f_h}{f_l} + A k_e f_h. \tag{5.28}$$

where k_e is a correction factor that accounts for the finite roll-off. For a first-, second-, third-, and fourth-order response k_e is equal to 1.57, 1.11, 1.05, and 1.03, respectively (van Etten 2005).

5.2.5 Resolution

The random noise of a position sensor causes an uncertainty in the measured position. If the distance between two measured locations is smaller than the uncertainty, it is possible to mistake one point for the other. In fabrication and imaging applications, this can cause manufacturing faults or imaging artifacts. To avoid these eventualities, it is critical to know the minimum distance between two adjacent but unique locations.

Since the random noise of a position sensor has a potentially large dispersion, it is impractically conservative to specify a resolution where adjacent locations never overlap. Instead, it is preferable to state the probability that the measured value is within a certain error bound. Consider the plot of three noisy measurements in Fig. 5.5 where the resolution δ_y is shaded in gray. The majority of sample points in y_2 fall within the bound $y_2 \pm \delta_y/2$. However, not all of the samples of y_2 lie within the resolution bound, as illustrated by the overlap of the probability density functions. To find the maximum measurement error, the resolution is added to other error sources as described in Sect. 5.2.6.

If the measurement noise is approximately Gaussian distributed, the resolution can be quantified by the standard deviation σ (RMS value) of the noise. The empirical rule (Brown and Hwang 1997) states that there is a 99.7 % probability that a sample of a Gaussian random process lie within $\pm 3\sigma$. Thus, if we define the resolution as $\delta = 6\sigma$ there is only a 0.3 % probability that a sample lies outside of the specified range. To be precise, this definition of resolution is referred to as the 6σ-resolution. Beneficially, no statistical measurements are required to obtain the 6σ-resolution if the noise is Gaussian distributed.

In other applications where more or less overlap between points is tolerable, another definition of resolution may be more appropriate. For example, the 4σ

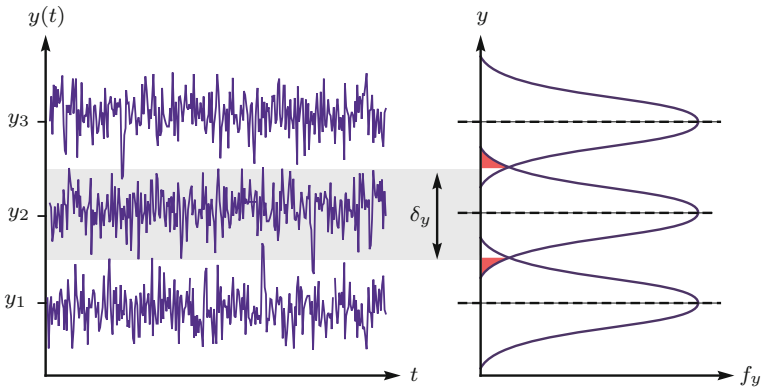


Fig. 5.5 The time-domain recording $y(t)$ of a position sensor at three discrete positions y_1, y_2, and y_3. The *large- shaded* area represents the resolution of the sensor and the approximate peak-to-peak noise of the sensor. The probability density function f_y of each signal is shown on the *right*

resolution would result in an overlap 4.5 % of the time, while the 10σ resolution would almost eliminate the probability of an overlap. Thus, it is not the exact definition that is important; rather, it is the necessity of quoting the resolution together with its statistical definition.

Although there is no international standard for the measurement or reporting of resolution in a positioning system, the ISO 5725 Standard on Accuracy (Trueness and Precision) of Measurement Methods and Results (ISO 1994) defines precision as the standard deviation (RMS Value) of a measurement. Thus, the 6σ-resolution is equivalent to six times the ISO definition for precision.

If the noise is not Gaussian distributed, the resolution can be measured by obtaining the 99.7 percentile bound directly from a time-domain recording. To obtain a statistically valid estimate of the resolution, the recommended recording length is 100 s with a sampling rate 15 × the sensor bandwidth (Fleming 2012), see Sect. 13.9.3. An anti-aliasing filter is required with a cut-off frequency 7.5 × the bandwidth. Since the signal is likely to have a small amplitude and large offset, an AC coupled preamplifier is required with a high-pass cut-off of 0.03 Hz or lower (Fleming 2012), see Sect. 13.9.3.

Another important parameter that must be specified when quoting resolution is the sensor bandwidth. In Eq. (5.28), the variance of a noise process is shown to be approximately proportional to the bandwidth f_h. By combining Eq. (5.28) with the above definition of resolution, the 6σ-resolution can be found as a function of the bandwidth f_h, noise density $\sqrt{A}$, and $1/f$ corner frequency f_{nc},

$$6\sigma\text{-resolution} = 6\sqrt{A}\sqrt{f_{\text{nc}} \ln \frac{f_h}{f_l} + k_e f_h}. \tag{5.29}$$

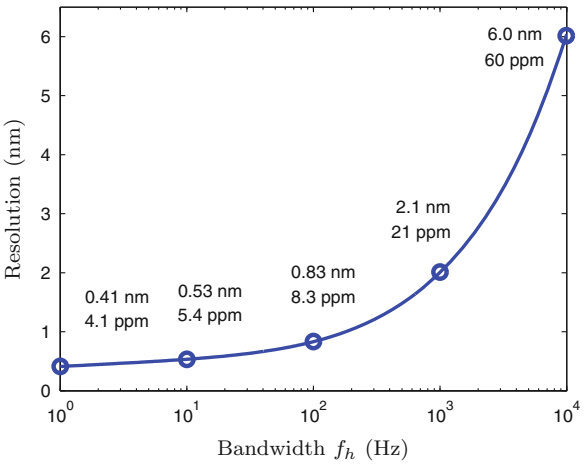


Fig. 5.6 The resolution versus bandwidth of a position sensor with a noise density of 10 pm/$\sqrt{\text{Hz}}$ and a $1/f$ corner frequency of 10 Hz. ($f_l = 0.01$ Hz and $k_e = 1$). At low frequencies, the noise is dominated by $1/f$ noise; however, at high frequencies, the noise increases by a factor of 3.16 for every decade of bandwidth

From Eq. (5.29), it can be observed that the resolution is approximately proportional to the square-root of bandwidth when $f_h >> f_{nc}$. It is also clear that the $1/f$ corner frequency limits the improvement that can be achieved by reducing the bandwidth. Note that Eq. (5.29) relies on a noise spectrum of the form (5.25) which may not adequately represent some sensors. The resolution of sensors with irregular spectrum's can be found by solving (5.23) numerically. Alternatively, the resolution can be evaluated from time-domain data, as discussed above.

The trade-off between resolution and bandwidth can be illustrated by considering a typical position sensor with a range of 100 μm, a noise density of 10 pm/$\sqrt{\text{Hz}}$, and a $1/f$ corner frequency of 10 Hz. The resolution is plotted against bandwidth in Fig. 5.6. When the bandwidth is below 100 Hz, the resolution is dominated by $1/f$ noise. For example, the resolution is only improved by a factor of two when the bandwidth is reduced by a factor of 100. Above 1 kHz, the resolution is dominated by the flat part of the power spectral density, thus a ten times increase in bandwidth from 1 to 10 kHz causes an approximately $\sqrt{10}$ reduction in resolution.

Many types of position sensors have a limited full-scale range (FSR); examples include strain sensors, capacitive sensors, and inductive sensors. In this class of sensor, sensors of the same type and construction tend to have an approximately proportional relationship between the resolution and range. As a result, it is convenient to consider the ratio of resolution to the full-scale range, or equivalently, the dynamic range (DNR). This figure can be used to quickly estimate the resolution from a given range, or conversely, to determine the maximum range given a certain resolution. A convenient method for reporting this ratio is in parts per million (ppm), that is

Table 5.1 Summary of the exact and simplified worst-case measurement errors

Error source	Exact	Simplified bound
Mapping error e_m	$f_a(v) - f_{cal}(\theta^*, v)$	$\pm \max \lvert e_m(v) \rvert$
Drift e_d	$(1 + k_s) f_a^*(v) + k_o - f_{cal}(v)$	$\pm (k_s \max \lvert f_{cal}(v) \rvert + k_o)$
Bandwidth e_{bw}	$\mathscr{F}^{-1}\{x(s)(1 - P(s))\}$	$\pm \frac{Anf}{f_c}$ (sine-wave)
Noise δ	NA	$6\sqrt{A}\sqrt{f_{\text{nc}} \ln \frac{f_h}{f_l} + k_e f_h}$

$$\text{DNR}_{\text{ppm}} = 10^6 \frac{6\sigma\text{-resolution}}{\text{Full-scale range}}. \tag{5.30}$$

This measure is equivalent to the resolution in nanometers of a sensor with a range of 1 mm. In Fig. 5.6 the resolution is reported in terms of both absolute distance and the dynamic range in ppm. The dynamic range can also be stated in decibels,

$$\text{DNR}_{\text{db}} = 20 \log_{10} \frac{\text{Full-scale range}}{6\sigma\text{-resolution}}. \tag{5.31}$$

Due to the strong dependence of resolution and dynamic range on the bandwidth of interest, it is clear that these parameters cannot be reported without the frequency limits f_l and f_h, to do so would be meaningless. Even if the resolution is reported correctly, it is only relevant for a single operating condition. A better alternative is to report the noise density and $1/f$ corner frequency, which allows the resolution and dynamic range to be calculated for any operating condition. These parameters are also sufficient to predict the closed-loop noise of a positioning system that incorporates the sensor (Fleming 2012). If the sensor noise is not approximately Gaussian or the spectrum is irregular, the resolution is measured using the process described above for a range of logarithmically spaced bandwidths.

5.2.6 Combining Errors

The exact and worst-case errors described in Sect. 5.2 are summarized in Table 5.1. In many circumstances, it is not practical to consider the exact error as this is dependent on the position. Rather, it is preferable to consider only the simplified worst-case error. An exception to the use of worst-case error is the drift error e_d. In this case, it may be unnecessarily conservative to consider the maximum error since the exact error is easily related to the sensor output by the uncertainty in sensitivity and offset.

To calculate the worst-case error e_t, the individual worst-case errors are summed, that is

$$e_t = e_m + e_d + e_{bw} + \delta/2 \tag{5.32}$$

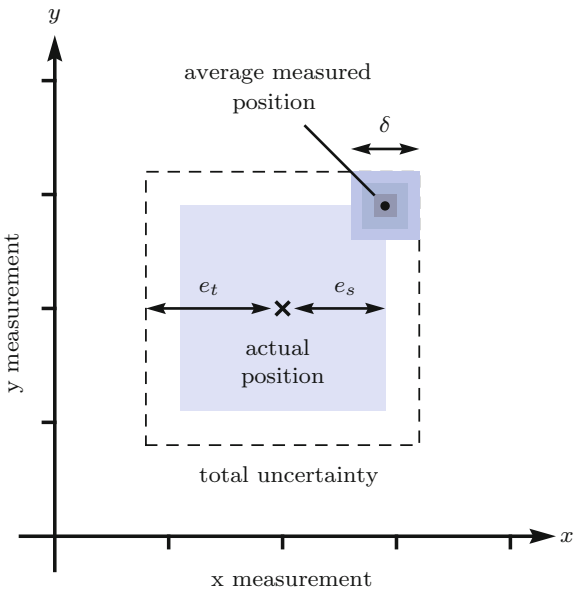


Fig. 5.7 The total uncertainty of a two-dimensional position measurement is illustrated by the *dashed box*. The total uncertainty e_t is due to both the static trueness error e_s and the noise δ

where e_m, e_d, e_{bw}, $\delta/2$ are the mapping error, the drift error, the error due to finite bandwidth, and the error due to noise whose maximum is half the resolution δ. The sum of the mapping and drift error can be referred to as the static trueness error e_s which is the maximum error in a static position measurement when the noise is effectively eliminated by a slow averaging filter. The total error and the static trueness error are illustrated graphically in Fig. 5.7.

5.2.7 Metrological Traceability

The error of a position sensor has been evaluated with respect to the true position. However, in practice, the "true" position is obtained from a reference sensor that may also be subject to calibration errors, nonlinearity and drift. If the tolerance of the calibration instrument is significant, this error must be included when evaluating the position sensor accuracy. However, such consideration is usually unnecessary as the tolerance of the calibration instrument is typically negligible compared to the position sensor being calibrated. To quantify the tolerance of a calibration instrument, it must be compared to a metrological reference for distance. Once the tolerance is known, measurements produced by the instrument can then be related directly to the reference, such measurements are said to be metrologically traceable.

Metrological traceability is defined as "the property of a measurement result whereby the result can be related to a reference through a documented unbroken chain of calibrations, each contributing to the measurement uncertainty" (JCGM200 2008). The reference for a distance measurement is the meter standard, defined by the distance traveled by light in vacuum over 1/299 792 458 seconds. Laser interferometers are readily calibrated to this standard since the laser frequency can be compared to the time standard which is known to an even higher accuracy than the speed of light.

Metrological traceability has little meaning by itself and must be quoted with an associated uncertainty to be valid (JCGM200 2008). If a position sensor is calibrated by an instrument that is metrologically traceable, subsequent measurements made by the position sensor are also metrologically traceable to within the bounds of the uncertainty for a specified operating environment (ISO/IEC 1994).

To obtain metrologically traceable measurements with the least uncertainty, an instrument should be linked to the reference standard through the least number of intervening instruments or measurements. All countries have a national organization that maintains reference standards for the calibration instruments. It should be noted that these organizations have individual policies for the reporting of traceability if their name is quoted. For example, to report that a measurement is NIST Traceable, the policy of the National Institute of Standards and Technology (USA), must be adhered to. Examples of measurement standards organizations include:

- National Measurement Institute (NIM), Australia
- Bureau International des Poids et Mesures (BIPM), France
- Physikalisch-Technische Bundesanstalt (PTB), Germany
- National Metrology Institute of Japan (NMIJ), Japan
- British Standards Institution (BMI), United Kingdom
- National Institute of Standards and Technology (NIST), USA.

5.3 Nanometer Position Sensors

5.3.1 Resistive Strain Sensors

Due to their simplicity and low-cost, resistive strain gauges are widely used for position control of piezoelectric actuators. Resistive strain gauges can be integrated into the actuator or bonded to the actuator surface. An example of a piezoelectric actuator and resistive strain gauge is pictured in Fig. 5.14a. Other application examples can be found in Lu et al. (2004), Dong et al. (2007), Schitter et al. (2008), Fleming and Leang (2010).

Resistive strain gauges are constructed from a thin layer of conducting foil laminated between two insulating layers. With a zig-zag conductor pattern, strain gauges can be designed for high sensitivity in only one direction, for example, elongation. When a strain gauge is elongated, the resistance increases proportionally. The change

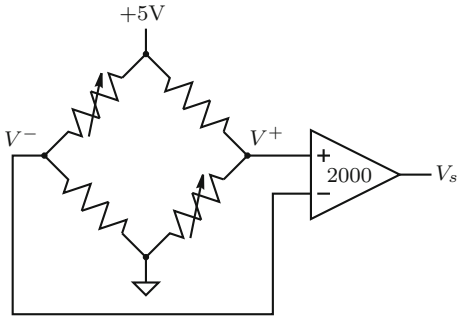


Fig. 5.8 A two-varying-element bridge circuit that contains two fixed resistors and two strain-dependent resistors. All of the nominal resistance values are equal. A simultaneous change in the two-varying-elements produces a differential voltage across the bridge

in resistance per unit strain is known as the gauge factor GF defined by

$$\text{GF} = \frac{\Delta R/R_G}{\epsilon}, \tag{5.33}$$

where ΔR is the change in resistance from the nominal value R_G for a strain ϵ. As the gauge factor is typically in the order of 1 or 2, the change is resistance is similar in magnitude to the percentage of strain. For a piezoelectric transducer with a maximum strain of approximately 0.1 %, the change in resistance is around 0.1 %. This small variation requires a bridge circuit for accurate measurement.

In Fig. 5.14b, a 10 mm Noliac SCMAP07 piezoelectric actuator is pictured with a strain gauge bonded to each of the two nonelectrode sides. The strain gauges are Omega SGD-3/350-LY13 gauges, with a nominal resistance of 350 Ohms and package dimensions of 7×4 mm. The electrical wiring of the strain gauges is illustrated in Fig. 5.8. The two-varying-element bridge circuit is completed by two dummy 350 Ohm wire wound resistors and excited by a 5 Volt DC source. The differential bridge voltage ($V^+ - V^-$) is acquired and amplified by a Vishay Micro-Measurements 2120B strain gauge amplifier. The developed voltage from a two-varying-element bridge is

$$V_s = \frac{A_v V_b}{2}\left(\frac{\Delta R}{R_G + \Delta R/2}\right), \tag{5.34}$$

where A_v=2000 is the differential gain and V_b=5 V is the excitation voltage. By substituting (5.33) into (5.34) and neglecting the small bridge nonlinearity[1], the measured voltage is proportional to the strain ϵ and displacement d by

[1] In a two-varying-element bridge circuit, the nonlinearity due to $\Delta R/2$ in Eq. (5.34) is 0.5 % nonlinearity per percent of strain (Kester 2002). Since the maximum strain of a piezoelectric actuator is 0.1 %, the maximum nonlinearity is only 0.05 % and can be neglected. If this magnitude of nonlinearity is not tolerable, compensating circuits are available (Kester 2002)

$$V_s = \frac{1}{2} A_v V_b \mathrm{GF} \epsilon \tag{5.35}$$

$$V_s = \frac{1}{2L} A_v V_b \mathrm{GF} d, \tag{5.36}$$

where L is the actuator length. With a gauge factor of 1, the position sensitivity of the amplified strain sensor is predicted to be 0.5 V/μm which implies a full-scale voltage of 5 V from a displacement of 10 μm. The actual sensitivity was found to be 0.3633 V/μm (Fleming and Leang 2010).

The bridge configuration shown in Fig. 5.8 is known as the two-varying-element bridge. It has twice the sensitivity of a single-element bridge but is also slightly nonlinear and sensitive to temperature variations between the gauge and bridge resistances. A detailed review of bridge circuits and their associated instrumentation can be found in Ref. Kester (2002). The best configuration is the four-varying-element differential bridge. This arrangement requires four strain gauges, two of which experience negative strain and another two that experience positive strain. Since the bridge is made entirely from the same elements, the four-varying-element bridge is insensitive to temperature variation. The bridge nonlinearity is also eliminated. In applications where regions of positive and negative strain are not available, the two-varying-element bridge is used.

Compared to other position sensors, strain gauges are compact, low-cost, precise, and highly stable, particularly in a full-bridge configuration (Kester 2002; Schitter et al. 2008). However, a major disadvantage is the high measurement noise that arises from the resistive thermal noise and the low sensitivity. The power spectral density of the resistive thermal noise is

$$S(f) = 4kTR \ \ \mathrm{V}^2/\mathrm{Hz}, \tag{5.37}$$

where k is the Boltzmann constant (1.38×10^{-23}), T is the room temperature in Kelvin (300°), and R is the resistance of each element in the bridge. In addition to the thermal noise, the current through the bridge also causes $1/f$.

The strain gauge pictured in Fig. 5.14a has a resistance of 350 Ohms, hence the spectral density is 2.4 nV/$\sqrt{\mathrm{Hz}}$. Since the sensitivity is 0.3633 V/μm, the predicted spectral density is 13 pm/$\sqrt{\mathrm{Hz}}$. This figure agrees with the experimentally measured spectral density plotted in Fig. 5.9. The sensor exhibits a noise density of approximately 15 pm/$\sqrt{\mathrm{Hz}}$ and a $1/f$ noise corner frequency of around 5 Hz. This compares poorly with the noise density of a typical inductive or capacitive sensor which is on the order of 1 pm/$\sqrt{\mathrm{Hz}}$ for a range of 10 μm. Hence, strain gauges are rarely used in systems designed for high resolution. If they are utilized in such systems, the closed-loop bandwidth must be severely restrained.

As an example of strain gauge resolution, we consider a typical two-varying-element strain gauge with an excitation of 5 V and a gauge factor of 1. The full-scale voltage is predicted to be 2.5 mV for a 0.1 % strain. If we assume a $1/f$ noise corner frequency of 5 Hz, $f_l = 0.01$ Hz, and a first-order bandwidth of 1 kHz ($k_e = 1.57$).

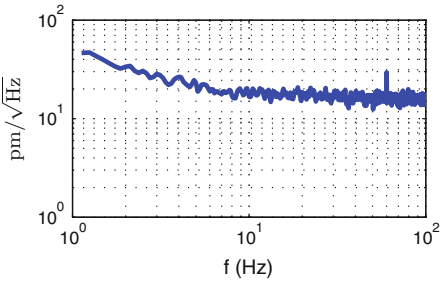


Fig. 5.9 The noise density of the strain sensor and instrumentation. The spectrum can be approximated by a constant spectral density and $1/f$ noise

The resolution predicted by Eq. (5.29) is 580 nV or 230 ppm. In other words, if the full-scale range was 100 μm, the resolution would be 23 nm, which is not competitive.

5.3.2 Piezoresistive Strain Sensors

In 1954, a visiting researcher at Bell Laboratories, Smith, demonstrated that "exceptionally large" resistance changes occur in silicon and germanium when subjected to external strain (Smith 1954). This discovery was the foundation for today's semiconductor piezoresistive sensors that are now ubiquitous in applications such as integrated pressure sensors and accelerometers (Barlian et al. 2009).

Compared to metal foil strain gauges that respond only to changes in geometry, piezoresistive sensors exhibit up to two orders of magnitude greater sensitivity. In addition to their high strain sensitivity, piezoresistive sensors are also easily integrated into standard integrated circuit and MEMS fabrication processes which is highly advantageous for both size and cost. The foremost disadvantages associated with piezoresistive sensors are the low strain range (0.1 %), high temperature sensitivity, poor long-term stability, and slight nonlinearity (1 %) (Barlian et al. 2009). The elimination of these artifacts requires a more complicated conditioning circuit than metal foil strain gauges; however, integrated circuits are now available that partially compensate for nonlinearity, offset, and temperature dependence, for example, the Maxim MAX1450.

As shown in Fig. 5.10, a typical integrated piezoresistive strain sensor consists of a planar n-doped resistor with heavily doped contacts. When the sensor is elongated in the x-axis, the average electron mobility increases in that direction, reducing resistance (Barlian et al. 2009). The effect is reverse during compression, or if the resistor is p-type. Since the piezoresistive effect is due to changes in the crystal lattice, the effect is highly dependent on the crystal orientation. The change in resistance can be expressed as,

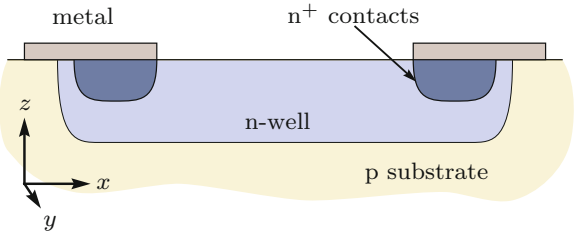


Fig. 5.10 A cross-section of a piezoresistive strain sensor. Deformation of the semiconductor crystal causes a resistance change one-hundred times that of a resistive strain gauge

$$\Delta R = R_G \left[\pi_L \sigma_{xx} + \pi_T \left(\sigma_{yy} + \sigma_{zz}\right)\right], \tag{5.38}$$

where ΔR is the change in resistance; R_G is the nominal resistance; σ_{xx}, σ_{yy}, and σ_{zz} are the tensile stress components in each axis; and π_L and π_T are the longitudinal and transverse piezoresistive coefficients which are determined from the crystal orientation (Barlian et al. 2009).

Due to the temperature dependence and low strain range, piezoresistive sensors are primarily used in microfabricated devices where the difficulties are offset by the high sensitivity and ease of fabrication, for example, meso-scale nanopositioners (DiBiasio and Culpepper 2008) and MEMs devices (Messenger et al. 2009). Discrete piezoresistive sensors are also available for standard macro-scale nanopositioning applications, for example, Micron Instruments SS-095-060-350PU. Discrete piezoresistive strain sensors are significantly smaller than metal foil gauges, for example, the Micron Instruments SS-095-060-350PU is 2.4 mm × 0.4 mm. The sensitivity is typically specified in the same way as a metal foil sensor, by the gauge factor defined in Eq. (5.33). While the gauge factor of a metal foil sensor is between 1 and 2, the gauge factor of the Micron Instruments SS-095-060-350PU is 120.

Due to the temperature dependence of piezoresistive strain sensors, practical application requires a closely collocated half- or full-bridge configuration, similar to a metal foil gauge. The required signal conditioning is also similar to the metal foil gauges. If an accuracy of better than 1 % is required, or if large changes in temperature are expected, the piezoresistive elements must be closely matched and the signal conditioning circuit must be compensated for temperature and nonlinearity. Two fully integrated bridge conditioning circuits include the MAX1450 and MAX1452 from Maxim Integrated Products, USA.

Alike metal foil strain gauges, the noise in piezoresistive sensors is predominantly thermal and $1/f$ noise (Barlian et al. 2009). However, since piezoresistive sensors are semiconductors, the $1/f$ noise can be substantially worse (Barlian et al. 2009). Consider the Micron Instruments SS-095-060-350PU piezoresistive sensor which has a gauge factor of 120 and a resistance of 350 Ω. In a two-varying-element bridge with 2-V excitation, Eq. (5.35) predicts that a full-scale strain of 0.1 % develops 120 mV. The thermal noise due to the resistance is 2.4 nV/$\sqrt{\text{Hz}}$. If the $1/f$ noise corner frequency is assumed to be 10 Hz, the resolution with a first-order bandwidth

of 1000 Hz is 130 nV which implies a 6σ-resolution of 590 nV or 4.9 ppm. Restated, if the full-scale displacement was 100 μm, the resolution would be 0.49 nm.

Although the majority of piezoresistive sensors are integrated directly into MEMS devices, discrete piezoresistive strain sensors are available from: Kulite Semiconductor Products Inc., USA; and Micron Instruments, USA.

5.3.3 Piezoelectric Strain Sensors

In addition to their actuating role, piezoelectric transducers are also widely utilized as high sensitivity strain sensors (Sirohi and Chopra 2000; Fleming and Moheimani 2005; Maess et al. 2008; Fleming et al. 2008; Fleming 2010; Yong et al. 2010, 2013). This is a common use for piezoelectric transducers in fields such as vibration control (Moheimani and Fleming 2006) but not in positioning applications. Beneficially, piezoelectric sensors can provide extremely high strain sensitivity with low measurement noise at high frequencies. However, they are also highly sensitive to temperature, prone to drift, and unable to measure static and low-frequency strains. The key is to utilize piezoelectric strain sensors in applications that benefit from their advantages but are not hindered by their limitations. In nanopositioning applications, piezoelectric strain sensors can be used for damping and vibration control as discussed in Chaps. 7 and 8, and for position measurement when an additional sensor is available, for example, in Ref. Fleming et al. (2008) or Chap. 8.

The basic operation of a piezoelectric strain sensor is illustrated in Fig. 5.11a. In this case the applied force F and resulting strain $\Delta h/h$ is aligned in the same axis as the polarization vector. Recall from Chap. 2 that the polarization vector points in the same direction as the internal dipoles which is opposite in direction to the applied electric field. Thus, compression of the actuator results in a voltage of the same polarity as the voltage applied during polarization. From the stress-voltage form of the piezoelectric constituent equations, the developed electric field E is

$$E = q_{33}\frac{\Delta h}{h}, \tag{5.39}$$

where Δh is the change in thickness, h is the thickness, and q_{33} is the piezoelectric coupling coefficient for the stress-voltage form. The constant q_{33} is related to the piezoelectric strain constant d_{33} by

$$q_{33} = \frac{d_{33}}{\epsilon^T s^D}, \tag{5.40}$$

where ϵ^T is the permittivity under constant stress (in Farad/m), and s^D is the elastic compliance under constant electric displacement (in m^2/N). If the piezoelectric voltage constant g_{33} is known instead of q_{33} or d_{33}, q_{33} can also be derived from $q_{33} = g_{33}/s^D$. By multiplying (5.40) by the thickness h, the measured voltage can be written as:

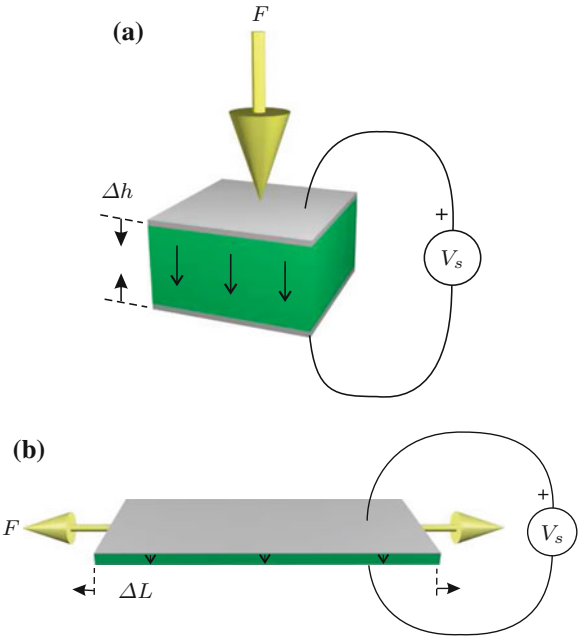


Fig. 5.11 A piezoelectric stack and plate strain sensor. The polarization vector is shown as a *downward arrow*. Axial sensors are typically used to measure dynamic forces while flexional sensors are used to measure changes in strain or curvature

$$V_s = q_{33}\Delta h, \tag{5.41}$$

If there are multiple layers, the voltage is

$$V_s = \frac{q_{33}}{n}\Delta h, \tag{5.42}$$

where n is the number of layers. The developed voltage can also be related to the applied force (Fleming and Leang 2010), as discussed in Sect. 8.2.2.

$$V_s = \frac{nd_{33}}{C}F, \quad \text{or} \quad V_s = \frac{d_{33}h}{n\epsilon^T A}F, \tag{5.43}$$

where C is the transducer capacitance defined by $C{=}n^2\epsilon^T A/h$, and A is the area

The voltage developed by the flexional sensor in Fig. 5.11b is similar to the axial sensor except for the change of piezoelectric constant. In a flexional sensor, the applied force and resulting strain are perpendicular to the polarization vector. Hence, the g_{31} constant is used in place of the g_{33} constant. Assuming that the length L is much larger than the width and thickness, the developed voltage is

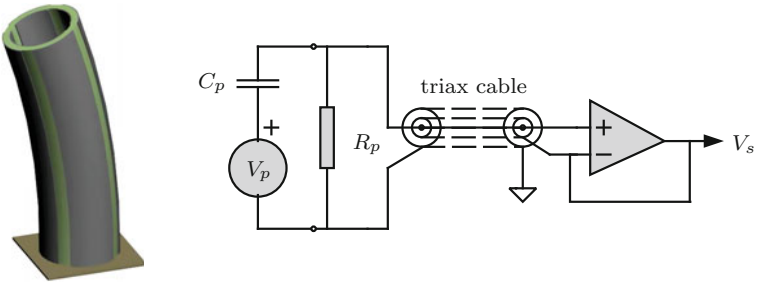


Fig. 5.12 A piezoelectric tube actuator with one electrode utilized as a strain sensor. The electrical equivalent circuit consists of the induced piezoelectric voltage V_p in series with the transducer capacitance. The dielectric leakage and input impedance of the buffer circuit are modeled by the parallel resistance R_p. An effective method for shielding the signal is to use a triaxial cable with the intermediate shield driven at the same potential as the measured voltage. (Tube drawing courtesy K. K. Leang)

$$V_s = \frac{-g_{31}}{L} F, \tag{5.44}$$

which can be rewritten in terms of the stiffness k and strain,

$$V_s = -g_{31} k \frac{\Delta L}{L} \tag{5.45}$$

$$V_s = \frac{-g_{31} A}{s^D L} \frac{\Delta L}{L}, \tag{5.46}$$

where A is the cross-sectional area equal to width $\times$ thickness.

When mounted on a host structure, flexional sensors can be used to detect the underlying stress or strain as well as the curvature or moment (Moheimani and Fleming 2006; Preumont 2006; Sirohi and Chopra 2000). In nanopositioning applications, the electrodes of a piezoelectric tube act as a plate sensor and can be used to detect the strain and hence displacement (Maess et al. 2008; Fleming et al. 2008; Yong et al. 2010). This application is illustrated in Fig. 5.12.

Due to the high mechanical stiffness of piezoelectric sensors, thermal or Boltzmann noise is negligible compared to the electrical noise arising from interface electronics. As piezoelectric sensors have a capacitive source impedance, the noise density $N_{Vs}(\omega)$ of the sensor voltage V_s is due primarily to the current noise i_n generated by the interface electronics. The equivalent electrical circuit of a piezoelectric sensor and high-impedance buffer is shown in Fig. 5.13. Neglecting the leakage resistance R, the noise density of the sensor voltage is

$$N_{Vs}(\omega) = i_n \frac{1}{C\omega}, \tag{5.47}$$

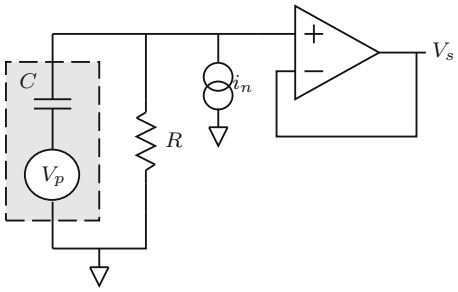


Fig. 5.13 The electrical model of a piezoelectric force sensor. The open-circuit voltage V_p is high-pass filtered by the transducer capacitance C and leakage resistance R. The current source i_n represents the current noise of a high-impedance buffer

where N_{Vs} and i_n are the noise densities of the sensor voltage and current noise, measured in Volts and Amps per $\sqrt{\text{Hz}}$ respectively.

The experimentally measured and predicted noise density of a piezoelectric sensor is plotted in Fig. 5.14. The sensor is a 2-mm Noliac CMAP06 stack mounted on top of 10-mm long actuator, the assembly is mounted in the nanopositioning stage pictured in Fig. 5.15. The sensor has a capacitance of 30 nF and the voltage buffer (OPA606) has a noise density of 2 fA/$\sqrt{\text{Hz}}$. Further details on the behavior of piezoelectric force sensors can be found in Sect. 8.2.2.

In Fig. 5.14b the noise density of the piezoelectric sensor is observed to be more than two orders of magnitude less than the strain and inductive sensors at 100 Hz. The noise density also continues to reduce at higher frequencies. However, at low frequencies the noise of the piezoelectric force sensor eventually surpass the other sensors. As the noise density is equivalent to an integrator excited by white noise, the measured voltage drifts significantly at low frequencies. A time record that illustrates this behavior is plotted in Fig. 5.16. The large drift amplitude is evident. Thus, although the piezoelectric force sensor generates less noise than the strain and inductive sensors at frequencies in the Hz range and above, it is inferior at frequencies below approximately 0.1 Hz.

In addition to noise, piezoelectric force sensors are also limited by dielectric leakage and finite buffer impedance at low-frequencies. The induced voltage V_p shown in Fig. 5.13 is high-pass filtered by the internal transducer capacitance C and the leakage resistance R. The cut-off frequency is

$$f_{hp} = \frac{1}{2\pi RC} \text{ Hz}. \tag{5.48}$$

The buffer circuit used in the results above has an input impedance of 100 MΩ, this results in a low-frequency cut-off of 0.05 Hz. To avoid a phase lead of more than 6 degrees, the piezoelectric force sensor cannot be used to measure frequencies of less than 0.5 Hz.

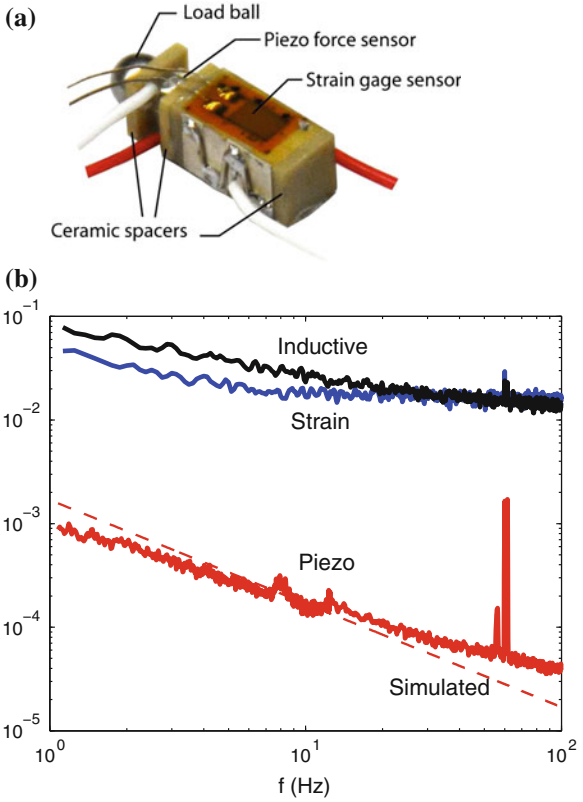


Fig. 5.14 **a** A piezoelectric stack actuator with an integrated force sensor and two resistive strain gages bonded to the *top* and *bottom* surface (the *bottom* gauge is not visible). In **b**, the noise density of the piezoelectric sensor is compared to the resistive strain gauge and a Kaman SMU9000-15N inductive sensor, all signals are scaled to nm/$\sqrt{\text{Hz}}$. The simulated noise of the piezoelectric force sensor is also plotted as a *dashed line*

Piezoelectric actuators and sensors are commercially available from: American Piezo (APC International, Ltd.), USA; CeramTec GmbH, Germany; Noliac A/S, Denmark; Physik Instrumente (PI), Germany; Piezo Systems Inc., USA; and Sensor Technology Ltd., Canada.

5.3.4 Capacitive Sensors

Capacitive sensors are the most commonly used sensors in short-range nanopositioning applications. They are relatively low-cost and can provide excellent linearity, resolution and bandwidth (Baxter 1997). However, due to the electronics required for measuring the capacitance and deriving position, capacitive sensors are

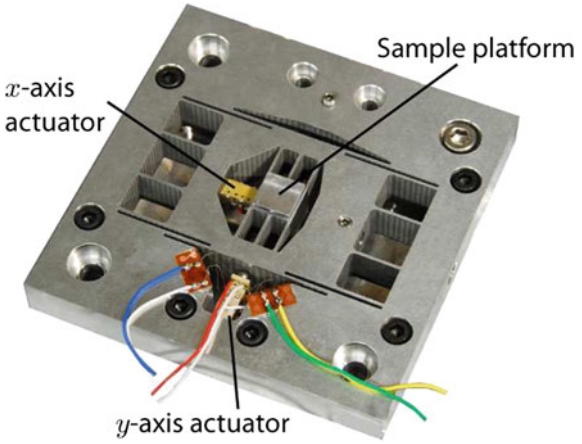


Fig. 5.15 A nanopositioning platform with a two-varying-element strain gauge fitted to the y-*axis* actuator (Fleming and Leang 2010). The nanopositioner is driven by two piezoelectric stack actuators that deflect the sample platform by a maximum of 10 μm in the x and y lateral axes

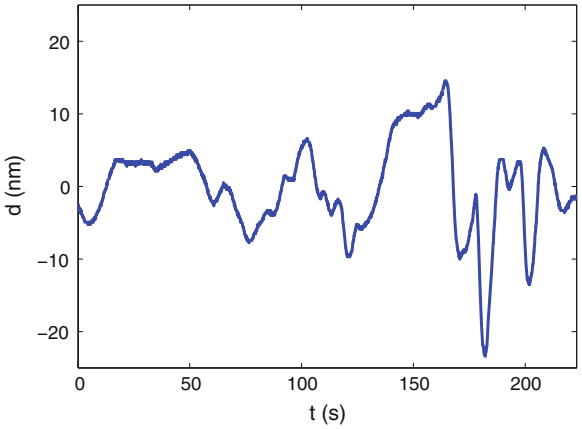


Fig. 5.16 Low-frequency noise of the piezoelectric sensor pictured in Fig. 5.14a, scaled to nanometers. The peak-to-peak noise over 220 s is 38 nm or 26 mV

inherently more complex than sensors such as resistive strain gauges. Larger ranges can be achieved with the use of an encoder-style electrode array (Kim et al. 2006).

All capacitive sensors work on the principle that displacement is proportional to the change in capacitance between two conducting surfaces. If fringe effects are neglected, the capacitance C between two parallel surfaces is

$$C = \frac{\epsilon_0 \epsilon_r A}{h}, \tag{5.49}$$

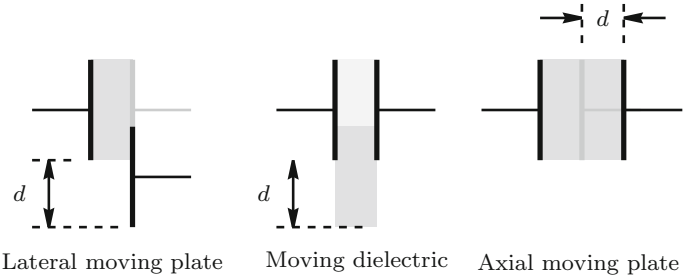


Fig. 5.17 Types of capacitive sensor. The *axial* moving plate produces the highest sensitivity but the smallest practical travel range. *Lateral* moving plate and moving dielectric sensors are most useful in long-range applications

where ϵ_0 is the permittivity of free space, ϵ_r is the relative permittivity of the dielectric (or dielectric constant), A is area between the surfaces, and h is the distance between the surfaces.

Three types of capacitive sensor are illustrated in Fig. 5.17. The lateral moving plate design is used for long range measurements where the plate spacing can be held constant. This is often achieved with two concentric cylinders mounted on the same axis. In this configuration, the change in capacitance is proportional to the change in area and hence position. A similar arrangement can be found in the moving dielectric sensor where the area and distance are constant but the dielectric is variable. This approach is not commonly used because a solid dielectric is required that causes friction and mechanical loading.

The axial moving plate, or parallel plate capacitive sensor is the most common type used in nanopositioning applications. Although the useful range is smaller than other configurations, the sensitivity is proportionally greater. The capacitance of a moving plate sensor is

$$C = \frac{\epsilon_0 \epsilon_r A}{d}, \tag{5.50}$$

hence, the sensitivity is

$$\frac{\mathrm{d}\, C}{\mathrm{d}\, d} = \frac{C_0}{d_0} \text{ F/m}, \tag{5.51}$$

where C_0 and d_0 are the nominal capacitance and distance. Thus, for a sensor with a nominal capacitance of 10 pF and spacing of 100 μm, the sensitivity is 100 fF/μm. The sensitivity of different capacitive sensor types is compared in Hicks et al. (1997).

A practical parallel plate capacitive sensor is illustrated in Fig. 5.18. In addition to the probe electrode, a guard electrode is also used to shield the probe from nearby electric fields and to improve linearity. The guard electrode is driven at the same potential as the probe but is not included in the capacitance measurement. As the fringing effect in the electric field is only present at the outside electrode, the nonlinearity in the capacitance measurement and distance calculation is reduced.

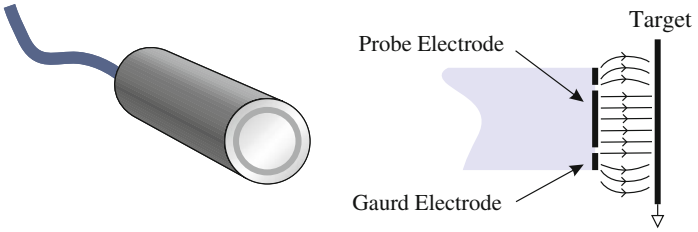


Fig. 5.18 A capacitive sensor probe and electrode configuration. The guard electrode is driven at the same potential as the probe in order to linearize the electric field and reduce fringing effects

A summary of correction terms for different guard electrode geometries can be found in Refs. Hicks et al. (1997) and Baxter (1997).

To measure the capacitance and thus derive the position, a wide variety of circuits are available (Nyce 2004; Baxter 1997). The simplest circuits are timing circuits where the timing capacitor is replaced by the sensor capacitance. Examples include the ubiquitous 555 timer in the one-shot or free-running oscillator modes. The output of a one-shot circuit is a pulse delay proportional to the capacitance. Likewise, the output of the oscillator is a square-wave whose frequency is proportional to capacitance. Although these techniques are not optimal for nanopositioning applications, they are simple, low-cost, and can be directly connected to a microcontroller with no analog-to-digital converters.

A direct measurement of the capacitance can be obtained by applying an AC voltage V to the probe electrode and grounding the target. The resulting current I is determined by Ohms law,

$$I = j\omega VC, \tag{5.52}$$

where ω is the excitation frequency in rad/s. Since the current is proportional to capacitance, this method is useful for the lateral moving plate and moving dielectric configurations where the displacement is also proportional to capacitance. For the axial moving plate configuration, where the displacement is inversely proportional to capacitance, it is more convenient to apply a current and measure the voltage. In this case, the measured voltage in response to an applied current is

$$V = \frac{I}{j\omega C}, \tag{5.53}$$

which is inversely proportional to capacitance and thus proportional to displacement.

Regardless of whether the current or voltage is the measured variable, it is necessary to compute the AC magnitude of the signal. The simplest circuit that achieves this is the single-diode demodulator or envelope detector shown in Fig. 5.19a. Although simple, the linearity and offset voltage of this circuit are dependent on the diode characteristics which are highly influenced by temperature. A better option is the synchronous demodulator with balanced excitation shown in Fig. 5.19b. A synchronous

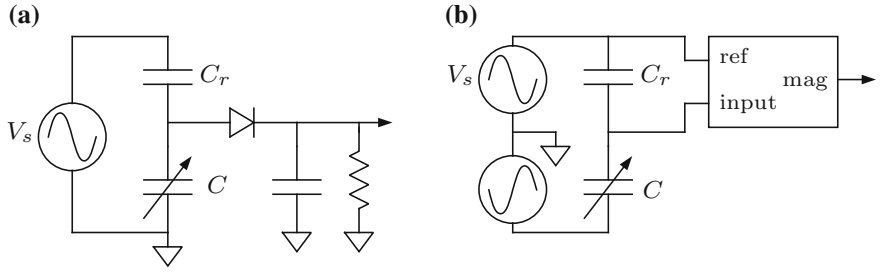


Fig. 5.19 Demodulation circuits for measuring capacitance. The linearity, temperature sensitivity, and noise performance of the synchronous detector is significantly better than the single-diode envelope detector

Table 5.2 A summary of error sources in a parallel plate capacitive sensor studied in Hicks et al. (1997)

Errors due to tilting		
Tilt angle	2 mrad	5 mrad
Nonlinearity	0.08 %	0.6 %
Offset	0.35 %	2.4 %
Scale error	0.8 %	5.4 %
Errors due to bowing		
Bow depth	10 μm	30 μm
Nonlinearity	0.025 %	0.33 %
Offset	5 %	18 %
Scale error	3 %	11 %

The sensor has a gap of 100 μm, a radius of 6 mm, and a nominal capacitance of 10 pF

demodulator can be constructed from a filter and voltage controlled switch (Nyce 2004; Baxter 1997). Integrated circuit demodulators such as the Analog Devices AD630 are also available. Synchronous demodulators provide greatly improved linearity and stability compared to single-diode detectors.

The balanced excitation in Fig. 5.19b eliminates the large DC offset produced by single-ended demodulators, such as Fig. 5.19a. The balanced configuration also eliminates the offset sensitivity to changes in the supply voltage, which greatly improves the stability. Although single-ended excitation can be improved with a full-bridge configuration, this requires a high common-mode rejection ratio, which is difficult to obtain at high frequencies.

In general, capacitive sensors with guard electrodes can provide excellent linearity in ideal conditions (10 ppm or 0.001 %); however, practical limitations can significantly degrade this performance. A detailed analysis of capacitive sensor nonlinearity in Hicks et al. (1997) concluded that the worst sources of nonlinearity are tilting and bowing. Tilting is the angle between the two parallel plates and bowing is the depth of concavity or convexity.

A summary of the error analysis performed in Hicks et al. (1997) is contained in Table 5.2. Considering that the linearity of an capacitive sensor in ideal conditions can

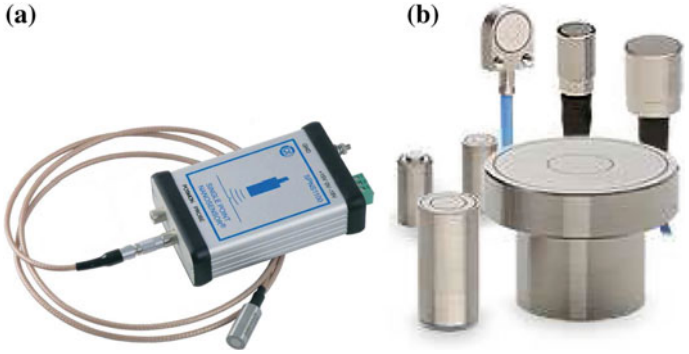


Fig. 5.20 An example of two commercially available capacitive sensors. Photos courtesy of Queensgate Instruments, UK and Micro-Epsilon, Germany

be 0.001 %, the effect of tilting and bowing severely degrades the performance. These errors can be reduced by careful attention to the mounting of capacitance sensors. It is recommended that capacitive sensors be fixed with a spring washer rather than a screw. This can significantly reduce mounting stress on the host structure and sensor. In addition to deformation, excessive mounting forces can slowly relieve over time causing major drifts in offset, linearity, and sensitivity.

The magnitude of error due to tilting and bowing can be reduced by increasing the nominal separation of the two plates, this also increases the range. However, if the area of the sensor is not increased, the capacitance drops, which increases noise.

The noise developed by a capacitive sensor is due primarily to the thermal and shot-noise of the instrumentation electronics. Due to the demodulation process, the noise spectral density is relatively flat and does not contain a significant $1/f$ component. Although the electronic noise remains constant with different sensor configurations, the effective position noise is proportional to the inverse of sensitivity. As the sensitivity is C_0/d_0 (5.51), if the capacitance is doubled by increasing the area, the position noise density is reduced by half. However, if the nominal gap d_0 is doubled to improve the linearity, the capacitance also halves, which reduces the sensitivity and increases the noise density by a factor of four. The position noise density is minimized by using the smallest possible plate separation and the largest area.

A typical commercial capacitive sensor with a range of 100 μm has a noise density of approximately of 20 pm/$\sqrt{\text{Hz}}$ (Fleming et al. 2008). The $1/f$ corner frequency of a capacitive sensor is typically very low, around 10 Hz. With a first-order bandwidth of 1 kHz, the resolution predicted by Eq. (5.29) is 2.4 nm or 24 ppm. This can be reduced to 0.55 nm or 5.5 ppm by restricting the bandwidth to 10 Hz.

Capacitive position sensors are commercially available from: Capacitec, USA; Lion Precision, USA; Micro-Epsilon, Germany; MicroSense, USA; Physik Instrumente (PI), Germany; and Queensgate Instruments, UK. Two commercially available devices are pictured in Fig. 5.20.

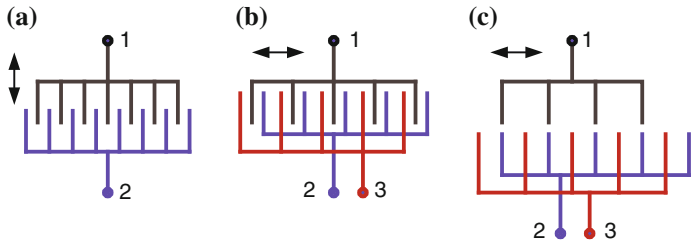


Fig. 5.21 Three examples of MEMs capacitive sensor geometries. **a** Standard comb sensor; **b** Differential comb sensor; **c** Incremental capacitive encoder

5.3.5 MEMs Capacitive and Thermal Sensors

MEMs capacitive sensors operate on a similar principles to their macro-scale counterpart discussed in the previous section. However, due to their small size, a more complicated geometry is required to achieve a practical value of capacitance. The comb type sensor illustrated in Fig. 5.21a is a common variety found in a number of nanopositioning applications, for example Chu and Gianchandani (2003), Zhu et al. (2011). In this configuration, the total capacitance is approximately proportional to the overlap area of each electrode array.

The basic comb sensor can be improved by employing a differential detection method as illustrated in Fig. 5.21b. Here, two sets of excitation electrodes (terminals 2 and 3) are driven 180 degrees out of phase. Thus, at the central position, the potential at terminal 1 is zero. This configuration provides a higher sensitivity than the basic comb sensor and is used extensively in devices such as accelerometers and gyroscopes (Baxter 1997; Kovacs 1998).

To increase the range of motion beyond a single inter-electrode spacing, the configuration in Fig. 5.21c uses withdrawn electrodes to form a capacitive incremental encoder (Kuijpers et al. 2003, 2006a, b). The slider can now move freely in either direction, limited only by the length of the excitation array. As the slider moves horizontally, the induced voltage at terminal 1 alternates between the phase of terminals 2 and 3. A second array is typically used to create a quadrature signal for ascertaining the direction of travel. This approach can provide a large travel range with high resolution but the decoding electronics is more complicated and the performance is sensitive to the separation between the arrays. If the two arrays can be overlain vertically, the capacitance can be increased while the difficulties with array separation are reduced (Lee et al. 2009; Lee and Peters 2009).

Electrothermal sensors are an alternate class of position sensors first utilized in nanopositioning applications by IBM in 2005 (Lantz et al. 2005). An example of a differential electrothermal position sensor is illustrated in Fig. 5.22. Two microheaters are driven by a DC voltage source resulting in a temperature increase. Due to the heat transfer between the microheater and moving heatsink, the temperature of each microheater becomes a function of the overlap area and hence position.

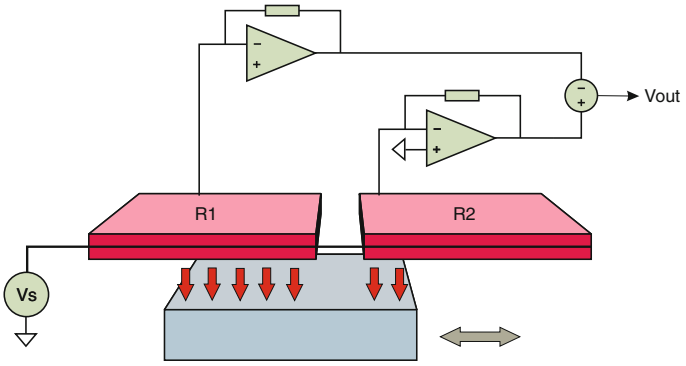


Fig. 5.22 An electrothermal position sensor. The two stationary microheaters are driven by a constant voltage source versus the rate of heat transfer and the resulting temperature is proportional to the overlap between the heater and the heatsink. The position of the heatsink can be estimated by measuring the current difference between the two microheaters which indicates the difference in resistance and temperature

The heatsink position is estimated by measuring the difference in current which is related to the resistance and temperature.

An advantage of electrothermal sensors over capacitive sensors is the compact size which has made them appealing in applications such as data storage (Pantazi et al. 2007; Sebastian et al. 2008; Sebastian and Wiesmann 2008) and nanopositioning (Sebastian and Pantazi 2012; Zhu et al. 2011). The noise performance of electrothermal sensors can be similar or superior to capacitive sensors under certain conditions. However, due to the elevated temperature, electrothermal sensors are known to exhibit a significant amplitude of low-frequency noise (Zhu et al. 2011).

With a range of 100 μm, a thermal position sensing scheme achieved a noise density of approximately 10 pm/$\sqrt{\text{Hz}}$ with a 1/f corner frequency of approximately 3 kHz (Sebastian and Pantazi 2012). This resulted in a resolution of 10 nm over a bandwidth of 4 kHz. As a result of the low frequency noise and drift, an auxiliary position sensor was utilized at frequencies below 24 Hz (Sebastian and Pantazi 2012).

5.3.6 Eddy-Current Sensors

Eddy-current, or inductive proximity sensors, operate on the principle of electromagnetic induction (Fraden 2004; Fericean and Droxler 2007). As illustrated in Fig. 5.23, an eddy-current probe consists of a coil facing an electrically conductive target. When the coil is excited by an AC current, the resulting magnetic field passes through the conductive target and induces a current according to Lenz's law. The current flows at right angles to the applied magnetic field and develops an opposing

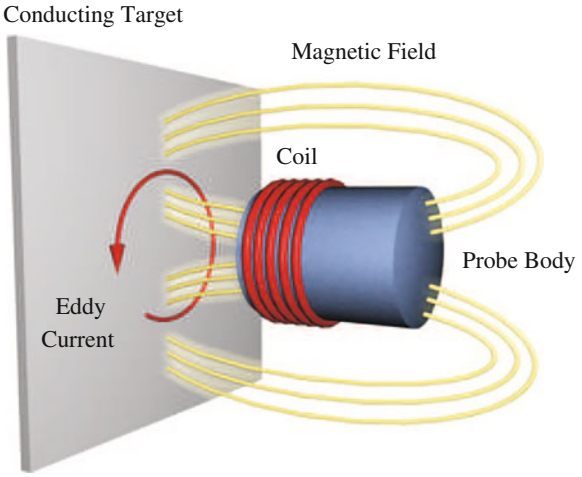


Fig. 5.23 The operating principle of an eddy-current sensor. An alternating current in the coil induces eddy-currents in the target. Increasing the distance between the probe and target reduces the eddy-currents and also the effective resistance of the coil

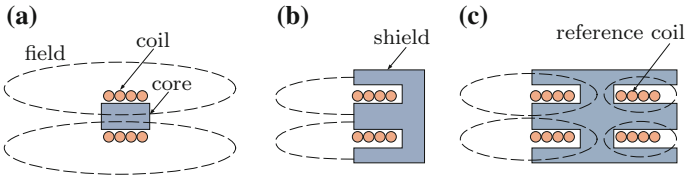


Fig. 5.24 Types of eddy-current sensor. The *unshielded* type has the greatest range but is affected by nearby fields and conductors. A *shield* makes the magnetic field more directional but reduces the range. A reference coil can be used to reduce the sensitivity to temperature

field. The eddy-currents and opposing field become stronger as the probe approaches the target.

The distance between probe and target is detected by measuring the AC resistance of the excitation coil which depends on the magnitude of the opposing field and eddy-current. The required electronics are similar to that of a capacitive sensor and include an oscillator and demodulator to derive the resistance (Roach 1998; Fraden 2004; Nyce 2004).

Three common types of eddy-current sensor are depicted in Fig. 5.24. The unshielded sensor has a large magnetic field that provides the greatest range; however, it also requires the largest target area and is sensitive to nearby conductors. Shielded sensors have a core of permeable material such as Permalloy, which reduces the sensitivity to nearby conductors and requires less target area; however, they also have less range. The balanced type has a second shielded or noninductive coil that is used to null the effect of temperature variation (Li and Ding 2005). The second coil is used in a divider or bridge configuration such as that illustrated in Fig. 5.25.

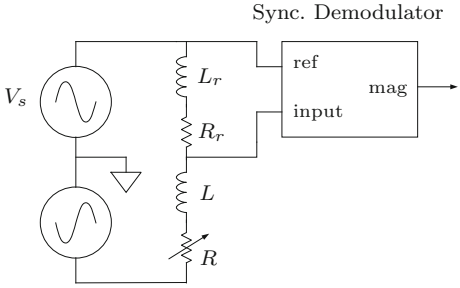


Fig. 5.25 Synchronous demodulation circuit for a balanced eddy-current sensor. L_r and R_r are the inductance and resistance of the reference coil

Another type of position sensor similar to an eddy-current sensor is the inductive proximity sensor, also referred to as a differential reluctance transducer if a reference coil is present. Rather than a conductive target, an inductive proximity sensor requires a ferromagnetic target. Since the reluctance of the magnetic path is proportional to the distance between the probe and target, the displacement can be derived from the coil inductance. Inductive proximity sensors have the same construction and electronics requirement as an eddy-current sensor. Their main drawback compared to eddy-current sensors is the temperature-dependent permeability of the target material and the presence of magnetic hysteresis.

Eddy-current sensors are not as widely used as capacitive sensors in nanopositioning applications due to the temperature sensitivity and range concerns. The temperature sensitivity arises from the need of an electrical coil in the sensor head and the varying resistance of the target. The minimum range of an eddy-current sensor is limited by the minimum physical size of the coil, which imposes a minimum practical range of between 100 and 500 μm. In contrast, capacitive sensors are available with a range of 10 μm, which can provide significantly higher resolution in applications with small travel ranges.

The major advantage of eddy-current and inductive sensors is the insensitivity to dust and pollutants in the air-gap and on the surface of the sensor. This gives them a significant advantage over capacitive sensors in industrial applications.

The noise performance of an eddy-current sensor can be similar to that of a capacitive sensor. For example, the noise density of the Kaman SMU9000-15N which has a range of 500 μm is plotted in Fig. 5.14b. The $1/f$ corner frequency is approximately 20 Hz and the constant density is approximately 20 pm/$\sqrt{\text{Hz}}$. Equation (5.29) predicts a resolution of 5 nm or 10 ppm with a bandwidth of 1 kHz. Due to the physical size of the coils, smaller ranges, and higher resolution is difficult to achieve.

Eddy-current position sensors are commercially available with ranges of approximately 100 μm–80 mm. Manufacturers include: Micro-Epsilon, Germany; Kaman Sensors, USA; MicroStrain, USA; Keyence, USA; Lion Precision, USA; and Ixthus Instrumentation, UK. Two commercially available devices are pictured in Fig. 5.26

Fig. 5.26 Two commercially available eddy-current sensors. Photos courtesy of Lion Precision, USA and Micro-Epsilon, Germany

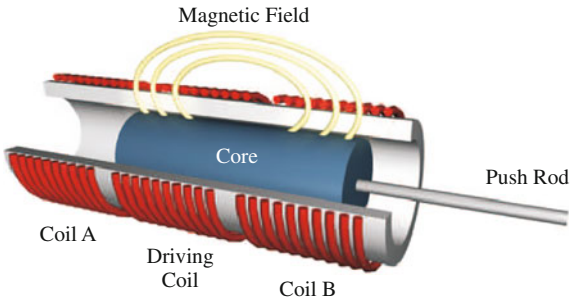


Fig. 5.27 The operating principle of a Linear Variable Displacement Transducer (LVDT). Changes in the core position produce a linear differential change in the coupling between the driving coil and the pick-up coils

5.3.7 Linear Variable Displacement Transformers

Linear Variable Displacement Transformers (LVDTs) are used extensively for displacement measurement with ranges of 1 mm to over 50 cm. They were originally described in a patent by G. B. Hoadley in 1940 (US Patent 2,196,809) and became popular in military and industrial applications due to their ruggedness and high resolution (Nyce 2004).

The operating principle of an LVDT is illustrated in Fig. 5.27. The stationary part of the sensor consists of a single driving coil and two sensing coils wound onto a thermally stable bobbin. The movable component of the transducer is a permeable material such as Nickel-Iron (Permalloy), and is placed inside the bobbin. The core is long enough to fully cover the length of at least two coils. Thus, at either extreme, the central coil always has a complete core at its center.

Since the central coil always has a complete core, all of the magnetic flux is concentrated in the core. As the core moves, the amount of flux passing through each sensor coil is proportional to the length of core contained within. Hence, the displacement of the core is proportional to the difference in voltage induced in the sensor coils. This principle is shown in Fig. 5.28.

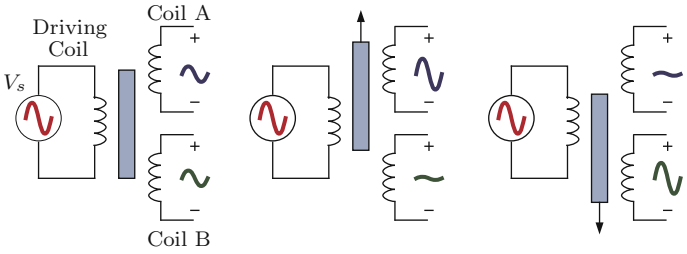


Fig. 5.28 The relationship between the sensor coil voltage and core position in an LVDT. The coil voltage is proportional to the amount or core it contains

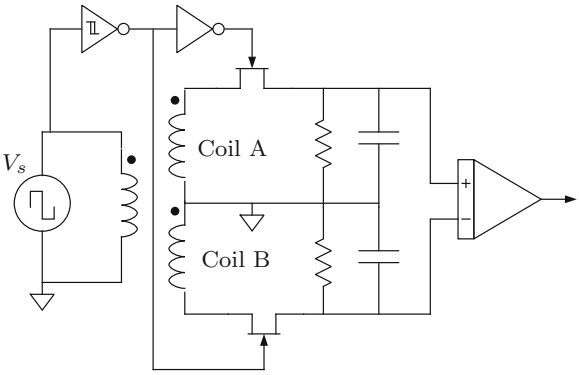


Fig. 5.29 A LVDT conditioning circuit with a synchronous demodulator and differential amplifier (Nyce 2004)

In addition to the components in Fig. 5.27, a bearing is required to guide the motion of the core through the bobbin. An external case is also required that can be constructed from a permeable material to provide magnetic shielding of the coils. It is important that the push-rod be constructed from a nonmagnetic material such as Aluminum or plastic otherwise it contributes erroneously to the coupling between the coils.

The electronics required by an LVDT are similar to that required for a capacitive or inductive sensor. An oscillator excites the driving coil with a frequency of around 1 kHz. Although higher frequencies increase the sensor bandwidth they also induce eddy-currents in the core that are detrimental to performance (Nyce 2004). Alike a capacitive or eddy-current sensor, a demodulator is required to determine the AC magnitude of the voltage induced in each coil. A simple synchronous demodulator circuit for this purpose is shown in Fig. 5.29 (Nyce 2004). The square-wave oscillator is replaced by a sine-wave oscillator if the electronics and LVDT are not physically collocated. Other demodulation circuits include the single-diode demodulator in Fig. 5.19a and the AD630-based demodulator in Fig. 5.19b.

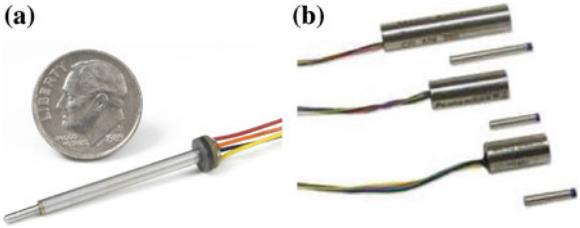


Fig. 5.30 Two commercially available LVDT sensors. Photos courtesy of Singer Instruments, Israel and Macro Sensors, USA

The greatest advantages of LVDTs are the infinitesimal resolution, large range, simplicity, and ruggedness. Very low levels of electrical noise can be achieved due to the low-impedance of the sensing coils. Nonlinearity is also below 1 % without the need for field calibration or mapping functions. The major drawbacks of LVDTs include the limited bandwidth and sensitivity to lateral motion. Due to eddy-currents and the inter winding capacitance, the excitation frequency is limited to a few tens of kHz, which limits the bandwidth to between 100 Hz and 1 kHz. Although classified as a noncontact sensor, bearings are required to guide the core linearly through the bobbin. This can be a significant disadvantage in nanopositioning applications if the sensor adds both friction and mass to the moving platform. However, if the platform is already flexure-guided, additional bearings may not be required. LVDTs are most suited to one-degree-of-freedom applications with relatively large displacement ranges of approximately 1 mm or greater. A range of less than 0.5 mm is difficult to achieve due to the small physical size of the coils. A notable exception is the air core LVDT coils used to detect position in the Asylum Research (USA) atomic force microscopes (Proksch et al. 2007). The air core eliminates eddy- current losses and Barkhausen noise caused by the high permeability materials. An RMS noise of 0.19 nm was reported for a range of 16 μm which equates to a resolution of approximately 1.14 nm and a dynamic range of 71 ppm (Proksch et al. 2007).

The theoretical resolution of LVDT sensors is limited primarily by the Johnson noise of the coils and Barkhausen noise in the magnetic materials (Proksch et al. 2007). However, standard conditioning circuits like the Analog Devices AD598 produce electronic noise on the order of 50 μVp-p with a bandwidth of 1 kHz. This imposes a resolution of approximately 10 ppm when using a driving amplitude of 5 Vp-p. Since the smallest commercially available range is 0.5 mm, the maximum resolution is approximately 5 nm with a 1 kHz bandwidth.

Due to their popularity, LVDTs and the associated conditioning electronics are widely available. Some manufacturers of devices that may be suitable in micro- and nanopositioning applications include: Macro Sensors, USA; Monitran, UK; Singer Instruments, Israel; MicroStrain, USA; Micro-Epsilon, USA; and Honeywell, USA. Two commercially available LVDTs are pictured in Fig. 5.30.

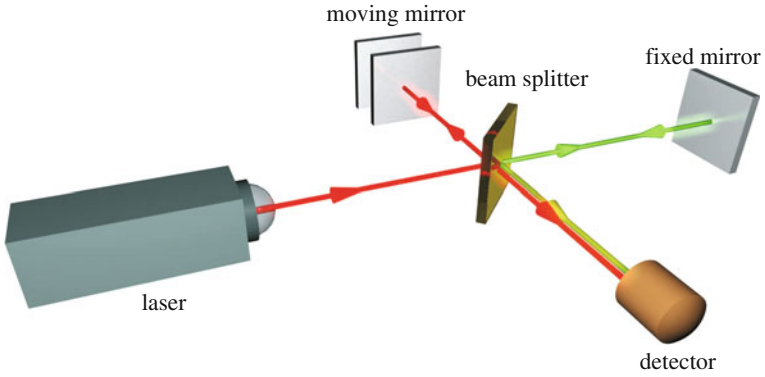


Fig. 5.31 The operation of a Michelson interferometer. The laser light is split into two paths, one that encounters a moving mirror and another that is fixed. The two beams are recombined and interfere at the detector. If the distance between the paths is an integer number of wavelengths, constructive interference occurs

5.3.8 Laser Interferometers

Since 1960, the meter length standard has been defined by optical means. This change arose after Michelson invented the interferometer which improved the accuracy of length measurement from a few parts in 10^7, to a few parts in 10^9 (Hariharan 2007). Thus, in 1960, the meter was redefined in terms of the orange line from a ^{86}Kr discharge lamp.

In 1983, the meter was redefined as the length traveled by light in a vacuum during a time interval of 1/299 792 458 s (Hariharan 2007). This definition was chosen because the speed of light is now fixed and the primary time standard, based on the ^{133}Cs clock, is known to an accuracy of a few parts in 10^{11} (Hariharan 2007). Length measurements are performed by interferometry using lasers with a frequency measured against the time standard. With a known frequency and speed, the laser wavelength can be found to an extremely high accuracy. Stabilized lasers are now available with precisely calibrated wavelengths for metrological purposes. Metrological traceability is described further in Sect. 5.2.7.

The operating principle of a Michelson interferometer is described in Fig. 5.31. A laser beam is split into two paths, one that is reflected by a moving mirror and another reflected by a stationary mirror. The movement of the mirror is measurable by observing the fringe pattern and intensity at the detector. If the distance between the paths is an integer number of wavelengths, constructive interference occurs. The displacement of the moving mirror, in wavelengths, is measured by counting the number of interference events that occur. The phase of the interference, and hence the displacement between interference events, can also be derived from the detector intensity.

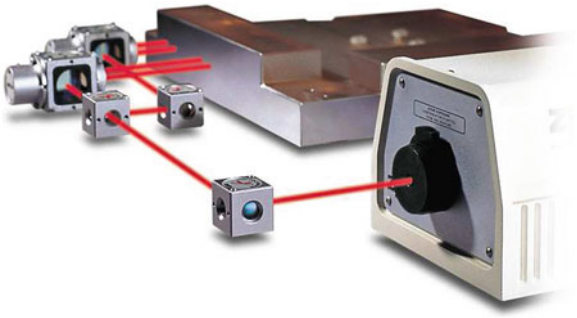


Fig. 5.32 A ZMI™two-*axis* heterodyne interferometer with a single laser source for measuring the angle and displacement of a positioning stage. Courtesy of Zygo, USA

Although simple, the Michelson interferometer is rarely used directly for displacement metrology. Due to the reference path, the Michelson interferometer is sensitive to changes or movement in the reference mirror and the beam splitter. Differences between the optical medium in the reference and measurement path are also problematic. Furthermore, the Michelson interferometer is not ideal for sub-wavelength displacement measurements as the phase sensitivity is a function of the path length. For example, at the peaks of constructive and destructive interference, the phase sensitivity is zero.

Modern displacement interferometers are based on the Heterodyne interferometer by Duke and Gordon from Hewlett-Packard in 1970 (Dukes and Gordon 1970). Although similar in principle to a Michelson interferometer, the heterodyne interferometer, overcomes many of the problems associated with the Michelson design. Most importantly, the phase sensitivity remains constant regardless of the path length.

Since the original work in 1970, a wide variety of improvements have been made to the basic heterodyne interferometer, for example Sommargren (1986). All of these devices work on the heterodyne principle, where the displacement is proportional to the phase (or frequency) difference between two laser beams. In heterodyne interferometers, the displacement signal is shifted up in frequency which avoids $1/f$ noise and provides immunity from low-frequency light source intensity variations.

In the original design, the two frequencies were obtained from a He-Ne laser forced to oscillate at two frequencies separated by 2 MHz. However, later designs utilize acousto-optic frequency shifters to achieve a similar result. An example application of a heterodyne interferometer is pictured in Fig. 5.32. Here, the angle and displacement of a linear positioning stage is measured using two interferometers and a single laser source.

A drawback of conventional interferometers is the large physical size and sensitivity to environmental variations which preclude their use in extreme environments such as within a cryostat or high magnetic field. To allow measurement in such environments, the miniature fiber interferometer, pictured in Fig. 5.33a, was developed (Karrai and Braun 2010). The measuring head contains a single-mode optical

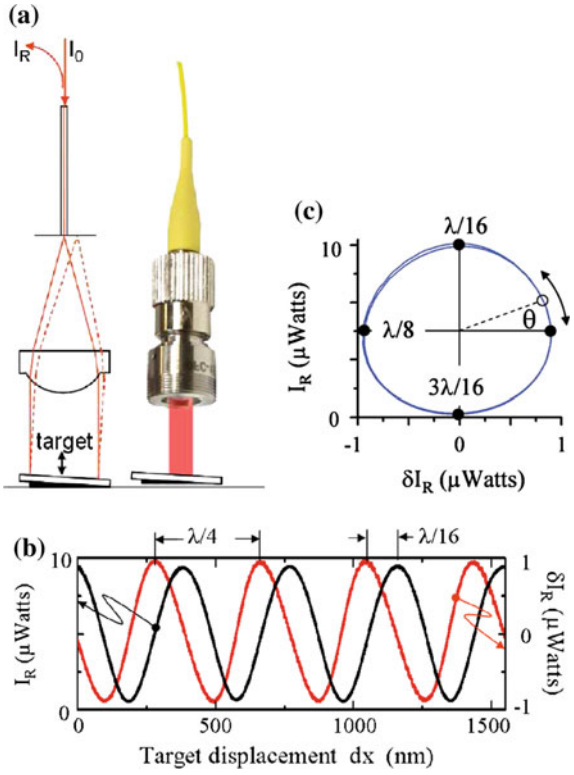


Fig. 5.33 The operating principle of an Attocube FPS miniature fibre interferometer (Karrai and Braun 2010), courtesy of Attocube, Germany. In (**a**) the transmitted light is reflected from the mirror, the fiber surface, the mirror again, and is then focused onto the fiber core. The interferogram plotted in (**b**) shows the direct reflected power (*black*) and the quadrature reflected power (*red*) versus displacement. The quadrature signal is obtained by modulating the laser wavelength and demodulating at the receiver. By plotting the power of the direct and quadrature signals (**c**), the direction of travel and sub-wavelength displacement can be resolved

fiber with a 9 μm core diameter coupled to a collimator lens. Approximately 4 % of the applied light is immediately reflected off the fiber termination and is returned down the fiber, forming the reference beam. The transmitted light passes through the collimator lens and is reflected off the slightly angled target mirror back towards the fiber surface but away from the core. As the fiber surface is a poor reflector, only 4 % of the incident light is reflected from the fiber surface. This reflected light travels back through the lens, is reflected off the mirror and is coupled directly to the fiber core, thus forming a Fabry-Perot interferometer with a cavity length equal to twice the distance between the fiber and mirror.

As the cavity length changes, the two beams interfere so that the reflected power is modulated periodically by the distance as illustrated in Fig. 5.33b. A problem with the basic interferogram is the lack of directional information. To resolve the

direction of travel, the light source wavelength is modulated at a high-frequency and demodulated at the receiver to provide an auxiliary interferogram in quadrature with the original. By considering both the directly reflected power and the demodulated reflected power, the direction of travel and can be deduced from the phase angle shown in Fig. 5.33c.

Since the miniature fiber interferometer is physically separated from the laser and receiver electronics it is both physically small and robust to extreme environments such as high vacuum, cryogenic temperatures, and magnetic fields. Due to the secondary reflection from the fiber surface, the fiber interferometer is also less sensitive to mirror misalignment compared to some other interferometers.

In general, laser interferometers are the most expensive displacement sensors due to the required optical, laser and electronic components. However, unlike other sensors, laser interferometers have an essentially unlimited range even though the resolution can exceed 1 nm. Furthermore, the accuracy, stability, and linearity exceed all other sensors. For these reasons, laser interferometers are widely used in applications such as semiconductor wafer steppers and display manufacturing processes. They are also used in some speciality nanopositioning applications that require metrological precision, for example, the metrological AFM described in Merry et al. (2009).

Aside from the cost, the main drawback of laser interferometers is the susceptibility of the beam to interference. If the beam is broken, the position is lost and the system has to be restarted from a known reference. The position can also be lost if the velocity of the object exceeds the maximum velocity imposed by the electronics. The maximum velocity is typically a few centimeters per second and is not usually a restriction; however, if the object is subject to shock loads, maximum velocity can become an issue.

The noise of laser interferometers is strongly dependent on the instrument type and operating environment. As an example, the Fabry-Perot interferometer discussed in Ref. Karrai and Braun (2010) has a $1/f$ noise corner frequency of approximately 10 Hz and a noise density of approximately 2 pm/$\sqrt{\text{Hz}}$. This results in a resolution of approximately 1.6 nm with a 12 kHz bandwidth. Equation (5.29) predicts a resolution of 0.49 nm with a 1 kHz bandwidth. Although the resolution of interferometers is excellent, small range sensors such as capacitive or piezoresistive sensors can provide higher resolution. However, the comparison is hardly fair considering that interferometers have a range in the meters while small range sensors may be restricted to 10 μm or less.

Some manufacturers of interferometers designed for stage metrology and position control include: Agilent, USA; Attocube, Germany (fiber Interferometer); Keyence, Japan (Fiber Interferometer); Renishaw, UK; Sios, Germany; and Zygo, USA. Instruments from these manufacturers are pictured in Figs. 5.33a and 5.34.

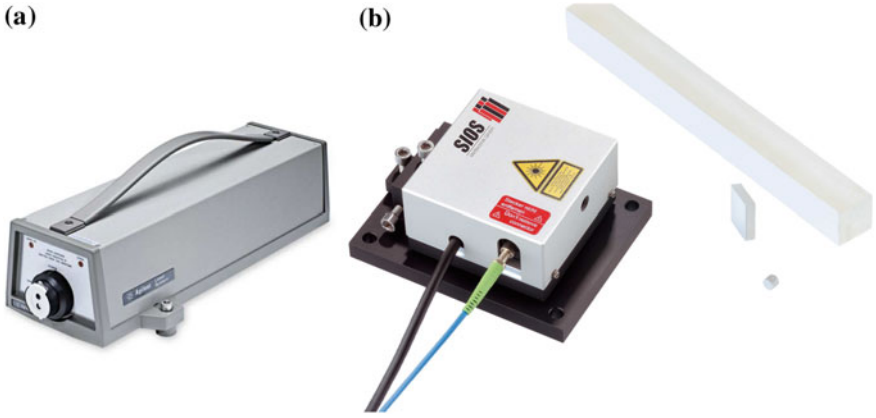


Fig. 5.34 Two commercially available Laser Interferometers. Photos courtesy of Agilent, USA and Sios, Germany

5.3.9 Linear Encoders

A linear encoder consists of two components, the reference scale and the read-head. The read-head is sensitive to an encoded pattern on the reference scale and produces a signal that is proportional to position. Either the scale or the read-head can be free to move, however the scale is typically fixed since the read-head is usually lighter.

The earliest form of linear encoder consisted of a bar with a conductive metal pattern, read by a series of metal brushes (Nyce 2004). Although simple, the constant contact between the brush and scale meant a very limited life and poor reliability.

In the 1950's optical linear encoders became available for machine tools. The reference scales were glass with a photochemically etched pattern. The photolithographic method used to produce the scale resulted in the highest resolution and accuracy at the time.

Although today's optical encoders still produce the highest resolution, other technologies have also become available. Magnetic or inductive linear encoders can not match the absolute accuracy or resolution of an optical scale encoder, however they are cheaper and more tolerant of dust and contamination. The most common type of encoder is possibly the capacitive encoder found in digital calipers. These devices use a series of conductive lines on the slider and scale to produce a variable capacitor.

The operation of a simple reflective optical encoder is illustrated in Fig. 5.35. Light from a laser diode is selectively reflected from the scale onto a photodetector. As the read-head is moved relative to the scale, the peaks in received power correspond the distance between the reflective bars. In between the peaks, the position can be estimated from the received power. Rather than partial reflection, other gratings contain height profiles that modulate the proximity and thus received power (Khiat et al. 2010).

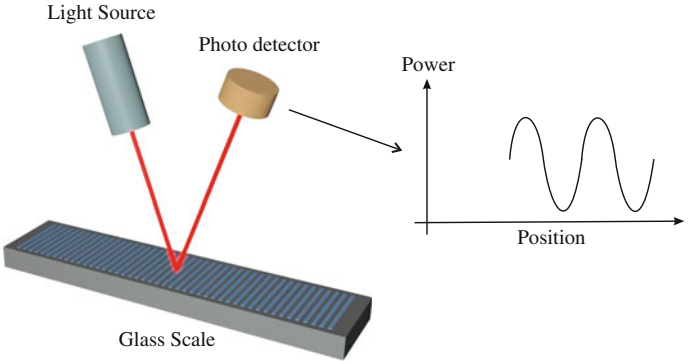


Fig. 5.35 The operation of a simple reflective optical encoder. The *peaks* in the received power correspond to the distance between reflective *bars*

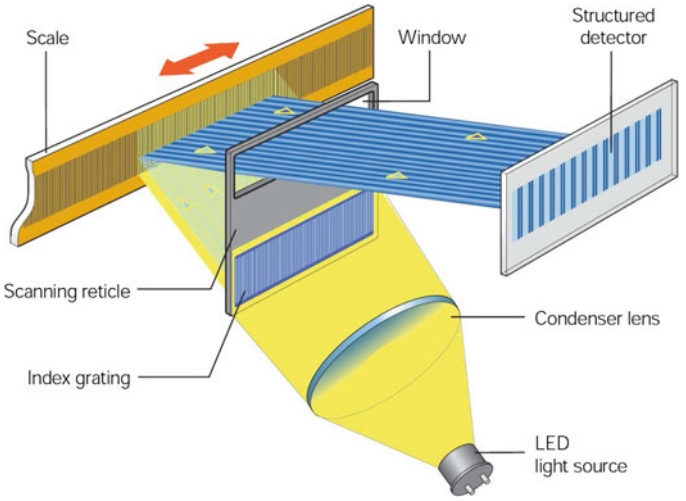


Fig. 5.36 The image scanning technique is used for reference scales with a grating pitch of between 10 and 200 μm. Image courtesy of Heidenhain, Germany

There are two major difficulties with the design illustrated in Fig. 5.35. First, the received power is highly sensitive to any dust or contamination on the scale. Second, it is difficult to determine the direction of motion, particularly at the peaks where the sensitivity approaches zero.

To provide immunity to dust and contamination, commercial optical encoders use a large number of parallel measurements to effectively average out errors. This principle relies on the Moire phenomenon (Sirohi 2009) and is illustrated by the image scanning technique shown in Fig. 5.36. In Fig. 5.36 a parallel beam of light is projected onto a reflective scale through a scanning reticle. The reflected Moire pattern is essentially the binary product of the scanning reticle and the scale and is

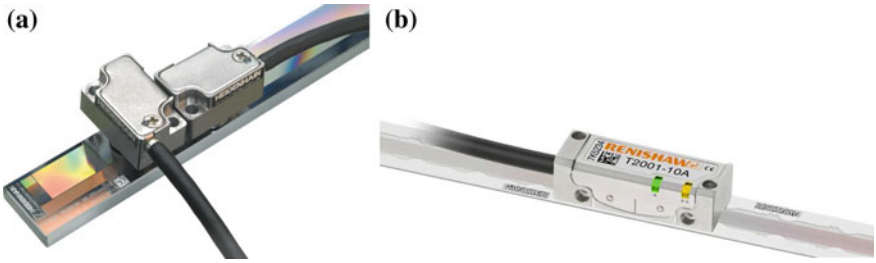


Fig. 5.37 Two commercially available optical linear encoders. Photos courtesy of Heidenhain, Germany and Renishaw, UK

detected by an array of photodetectors. Aside from the immunity to contamination, this technique also provides a quadrature signal that provides directional information.

Optical reference scales are encoded with a geometric pattern that describes either the absolute position or the incremental position. Absolute scales contain additional information that can make them physically larger than incremental scales. Compared to an incremental encoder, an absolute encoder is also typically more sensitive to alignment errors, lower in resolution, slower, and more costly. The benefit of an absolute scale is that the read-head does not need to return to a known reference point after a power failure or read error.

The noise of high resolution optical encoders is described as "jitter" and is typically on the order of 1 nm RMS, or 6 nm peak-to-peak. The overall accuracy is around 5 μm/m (FASTRACK 2014), however accuracies as high as 0.5 μm/m are possible with ranges up to 270 mm (Heidenhain 2014).

The highest resolution optical encoders operate on the principle of interference (Heidenhain 2014; Lee et al. 2007). The technique involves light that is diffracted through a transparent phase grating in the read-head and reflected from a step grating on the scale (Heidenhain 2014). Since this technique operates on the principle of diffraction, extremely small signal periods of down to 128 nm are possible with a resolution on the order of a few nanometers.

Other encoder technologies include techniques where the position information is actually encoded into the medium being scanned. Examples of this approach include hard disk drives (Chen et al. 2006) and MEMS mass storage devices (Sebastian et al. 2008).

Companies that produce linear encoders suitable for nanometer scale metrology include: Heidenhain, Germany; MicroE Systems, USA; and Renishaw, UK. Two instruments from these manufacturers are pictured in Fig. 5.37.

Table 5.3 Summary of position sensor characteristics

Sensor type	Range	DNR (ppm)	Resolution (nm)	Max. BW (kHz)	Accuracy (ppm FSR)
Metal foil	10–500 μm	230	23	1–10	1 %
Piezoresistive	1–500 μm	4.9	0.49	>100	1 %
Capacitive	10–10 mm	24	2.4	100	0.1 %
Electrothermal	10 μm–1 mm	100	10	10	1 %
Eddy current	100 μm–80 mm	10	1	40	0.1 %
LVDT	0.5–500 mm	10	5	1	0.25 %
Interferometer	Meters		0.49	>100	1
Encoder	Meters		6	>100	5

The dynamic range (DNR) and resolution are approximations based on a full-scale range of 100 μm and a first-order bandwidth of 1 kHz

5.4 Comparison and Summary

Due to the extreme breadth of position sensor technologies and the wide range of applications, it is extremely difficult to make direct performance comparisons. In many applications, characteristics such as the physical size and cost play a greater role than performance. Nevertheless, it is informative to compare some aspects of performance.

In Table 5.3 the specifications under consideration are the range, the dynamic range, the 6σ-resolution, the maximum bandwidth, and the typical accuracy. Consider the following notes when interpreting the results in Table 5.3:

- The quoted figures are representative of commercially available devices and do not imply any theoretical limits.
- The dynamic range and 6σ-resolution is an approximation based on a full-scale range of 100 μm and a first-order bandwidth of 1 kHz. The low-frequency limit is assumed to be $f_l = 0.01$ Hz.
- The quoted accuracy is the typical static trueness error defined in Sect. 5.2.6.

Metal foil strain gauges are the simplest and lowest cost sensor considered in this study. Due to their size (a few mm^2) strain gauges are suitable for mounting directly on to actuators or stages with a range from 10 to 500 μm. The parameters in Table 5.3 pertain to the example of a two-varying- element bridge discussed in Sect. 5.3.1. Although strain gauges can be calibrated to achieve higher accuracy, it is reasonable to consider an error of 1 % FSR due to drift and the indirect relationship between the measured strain and actual displacement.

Piezoresistive sensors are smaller than metal foil strain gauges and can be bonded to actuators that are only 1 mm long with a range of up to 1 μm. Although the resolution of piezoresistive sensors is very high, the absolute accuracy is limited by nonlinearity, temperature sensitivity, and inexact matching. An error budget of 1 % FSR is typical. Although strain sensors require contact with the actuator or

flexural components, they do not introduce forces between the reference and moving platforms, thus, in this sense, they are considered to be noncontact.

Capacitive sensors are relatively simple in construction, provide the highest resolution over short ranges, are insensitive to temperature, and can be calibrated to an accuracy of 0.01 % FSR. However, in general purpose applications where the sensor is not calibrated after installation, alignment errors may limit the accuracy to 1 % FSR. The capacitive sensor parameters under consideration are described in Sect. 5.3.4.

Eddy-current sensors can provide excellent resolution for travel ranges greater than 100 μm. They are more sensitive to temperature than capacitive sensors but are less sensitive to dust and pollutants which is important in industrial environments. The quoted noise and resolution is calculated from the example discussed in Sect. 5.3.6.

LVDT sensors are among the most popular in industrial applications requiring a range from a few millimeters to tens of centimeters. They are simple, have a high intrinsic linearity and can be magnetically shielded. However, they also have a low bandwidth and can load the motion with inertia and friction. The maximum resolution is limited by the physical construction of the transducer which is generally suited to ranges of greater than 1 mm. The bandwidth of LVDT sensors is limited by the need to avoid eddy currents in the core. With an excitation frequency of 10 kHz, the maximum bandwidth is approximately 1 kHz.

Compared to other sensor technologies, laser interferometers provide an unprecedented level of accuracy. Stabilized interferometers can achieve an absolute accuracy exceeding 1 ppm, or in other words, better than 1 um/m. Nonlinearity is also on the order of a few nanometers. Due to the low-noise and extreme range, the dynamic range of an interferometer can be as high as a few parts per billion, or upwards of 180 dB. The quoted resolution in Table 5.3 is associated with the Fabry-Perot interferometer discussed in Sect. 5.3.8.

Linear encoders are used in similar applications to interferometers where absolute accuracy is the primary concern. Over large ranges, absolute accuracies of up to 5 ppm or 5 μm/m are possible. Even greater accuracies are possible with linear encoders working on the principle of diffraction. The accuracy of these sensors can exceed 1 ppm over ranges of up to 270 mm, which is equivalent to the best laser interferometers.

5.5 Outlook and Future Requirements

One of the foremost challenges of position sensing is to achieve high resolution and accuracy over a large range. For example, semiconductor wafer stages require a repeatability and resolution in the nanometers while operating over a range in the tens of centimeters (Butler 2011; Mishra et al. 2007). Such applications typically use interferometers or high resolution optical encoders which can provide the required performance but can impose a significant cost. Long range sensors are also becoming necessary in standard nanopositioning applications due to the development of

dual-stage actuators (Michellod et al. 2006; Chassagne et al. 2007; Fleming 2011; Zheng et al. 2011) and stepping mechanisms (Chu and Fan 2006; Merry et al. 2011). Capacitive sensors can be adapted for this purpose by using a periodic array of electrodes (Lee and Peters 2009). Such techniques can also be applied to magnetic or inductive sensing principles. Due to the increasing availability of long range nanopositioning mechanisms, an increased focus on the development of cost-effective long range sensors is required.

A need is also emerging for position sensors capable of measuring position at frequencies up to 100 kHz. Applications include: high-speed surface inspection (Borionetti et al. 2004; Humphris et al. 2006); nanofabrication (Tseng et al. 2008; Vicary and Miles 2008; Tseng 2008; Ferreira and Mavroidis 2006), and imaging of fast biological and physical processes (Fantner et al. 2006; Kobayashi et al. 2007; Schitter et al. 2007; Picco et al. 2007; Ando et al. 2008; Fleming et al. 2010a). Although, many sensor technologies can provide a bandwidth of 100 kHz, this figure is the 3 dB bandwidth where phase and time delay render the signal essentially useless in a feedback loop. High speed position sensors are required with a bandwidth in the MHz that can provide accurate measurements at 100 kHz with negligible phase shift or time delay. Due to the operating principle of modulated sensors such as capacitive and inductive sensors, this level of performance is difficult to achieve due to the impractically high carrier frequency requirement. Applications requiring a very high sensor bandwidth typically use an auxiliary sensor for high bandwidth tasks, for example, a piezoelectric sensor can be used for active resonance damping (Yong et al. 2013; Fleming 2010). Technologies such as piezoresistive sensors (Guliyev et al. 2012) have also shown promise in high-speed applications since a carrier frequency is not required. Magnetoresistive sensors are also suitable for high frequency applications if the changes in field strength can be kept small enough to mitigate hysteresis (Sahoo et al. 2011; Kartik et al. 2012).

Due to the lack of cost-effective sensors that provide both high-resolution and wide bandwidth, recent research has also considered the collaborative use of multiple sensors. For example, in Fleming et al. (2008) a piezoelectric strain sensor and capacitive sensor were combined. The feedback loop utilized the capacitive sensor at low frequencies and the piezoelectric sensor at high frequencies. This approach retains the low-frequency accuracy of the capacitive sensor and the wide bandwidth of the piezo sensor while avoiding the drift from the piezo sensor and wide-band noise from the capacitive sensor. The closed-loop noise was reduced from 5 nm with the capacitive sensor to 0.34 nm with both sensors. Piezoelectric force sensors have also been used for high-frequency damping control while a capacitive, inductive or strain is used for tracking control (Fleming 2010; Fleming and Leang 2010).

Data storage systems are an example application that requires both long range but extreme resolution and increasingly wide bandwidth. In these applications, a media derived position error signal (PES) can provide the requisite range and resolution but not the bandwidth. In Ref. Sebastian et al. (2008) a MEMs storage device successfully combined the accuracy of a media derived position signal with the speed of an electrothermal sensor. Electrothermal sensors have also been combined with capacitive sensors to reduce the inherent 1/f noise (Zhu et al. 2011). Multiple

sensors can be combined by complementary filters (Fleming 2010) or by an optimal technique in the time domain (Fleming et al. 2008) or frequency domain (Sebastian and Pantazi 2012). Given the successful applications to date, it seems likely that the trend of multiple sensors will continue, possibly to the point where multiple sensors are packaged and calibrated as a single unit.

References

Abramovitch DY, Andersson SB, Pao LY, Schitter G (2007) A tutorial on the mechanisms, dynamics, and control of atomic force microscopes. In: Proceedings of American control conference, New York City, NY, pp 3488–3502, July 2007

Ando T, Uchihashi T, Fukuma T (2008) High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes. Prog Surf Sci 83(7–9):337–437

Barlian A, Park W-T, Mallon J, Rastegar A, Pruitt B (2009) Review: semiconductor piezoresistance for microsystems. Proc IEEE 97(3):513–552

Baxter LK (1997) Capacitive sensors: design and applications. IEEE Press, Piscataway

Borionetti G, Bazzalia A, Orizio R (2004) Atomic force microscopy: a powerful tool for surface defect and morphology inspection in semiconductor industry. Eur Phys J Appl Phys 27(1–3): 101–106

Brown RG, Hwang PYC (1997) Introduction to random signals and applied kalman filtering. Wiley, New York

Butler H (2011) Position control in lithographic equipment. IEEE Control Syst 31(5):28–47

Chassagne L, Wakim M, Xu S, Topçu S, Ruaux P, Juncar P, Alayli Y (2007) A 2d nano-positioning system with sub-nanometric repeatability over the millimetre displacement range. Meas Sci Technol 18(11):3267–3272

Chen BM, Lee TH, Peng K, Venkatarmanan V (2006) Hard disk drive servo system. Springer, London

Chu LL, Gianchandani YB (2003) A micromachined 2d positioner with electrothermal actuation and sub-nanometer capacitive sensing. J Micromech Microeng 13(2):279–285

Chu C-L, Fan S-H (2006) A novel long-travel piezoelectric-driven linear nanopositioning stage. Precis Eng 30(1):85–95

Devasia S, Eleftheriou E, Moheimani SOR (2007) A survey of control issues in nanopositioning. IEEE Trans Control Syst Technol 15(5):802–823

DiBiasio CM, Culpepper ML (2008) Design of a meso-scale six-axis nanopositioner with integrated position sensing. In: Proceedings 5th annual international symposium on nanomanufacturing, Singapore

Dong W, Sun LN, Du ZJ (2007) Design of a precision compliant parallel positioner driven by dual piezoelectric actuators. Sens Actuators A 135(1):250–256

Dukes JN, Gordon GB (1970) A two hundred-foot yardstick with graduations every microinch. Hewlett-Packard J 21(2):2–8

van Etten WC (2005) Introduction to noise and random processes. Wiley, West Sussex

Fantner GE, Schitter G, Kindt JH, Ivanov T, Ivanova K, Patel R, Holten-Andersen N, Adams J, Thurner PJ, Rangelow IW, Hansma PK (2006) Components for high speed atomic force microscopy. Ultramicroscopy 106(2–3):881–887

FASTRACK high-accuracy linear encoder scale system. Data sheet l-9517-9356-01-b. Online: www.renishaw.com.

Fericean S, Droxler R (2007) New noncontacting inductive analog proximity and inductive linear displacement sensors for industrial automation. IEEE Sens J 7(11):1538–1545

Ferreira A, Mavroidis C (2006) Virtual reality and haptics for nanorobotics. IEEE Robot Autom Mag 13(3):78–92

Fleming AJ (2012a) Estimating the resolution of nanopositioning systems from frequency domain data. In: Proceedings IEEE international conference on robotics and automation, St. Paul, MN, pp 4786–4791, May 2012
Fleming AJ (2012b) Measuring picometer nanopositioner resolution. In: Proceedings of Actuator 2012, 13th international conference on new actuators, Bremen, June 18–20 2012
Fleming AJ, Moheimani SOR (2005) Control oriented synthesis of high performance piezoelectric shunt impedances for structural vibration control. IEEE Trans Control Syst Technol 13(1):98–112
Fleming AJ, Wills AG, Moheimani SOR (2008) Sensor fusion for improved control of piezoelectric tube scanners. IEEE Trans Control Syst Technol 15(6):1265–6536
Fleming AJ, Kenton BJ, Leang KK (2010) Bridging the gap between conventional and video-speed scanning probe microscopes. Ultramicroscopy 110(9):1205–1214
Fleming AJ, Aphale SS, Moheimani SOR (2010) A new method for robust damping and tracking control of scanning probe microscope positioning stages. IEEE Trans Nanotechnol 9(4):438–448
Fleming AJ, Leang KK (2010) Integrated strain and force feedback for high performance control of piezoelectric actuators. Sens Actuators A 161(1–2):256–265
Fleming AJ (2010) Nanopositioning system with force feedback for high-performance tracking and vibration control. IEEE Trans Mechatron 15(3):433–447
Fleming AJ (2011) Dual-stage vertical feedback for high speed-scanning probe microscopy. IEEE Trans Control Syst Technol 19(1):156–165
Fleming AJ (2012) A method for measuring the resolution of nanopositioning systems. Rev Sci Instrum 83(8):086101
Fraden J (2004) Handboook of modern sensors: physics, designs, and applications. Springer, New York
Guliyev E, Michels T, Volland B, Ivanov T, Hofer M, Rangelow I (2012) High speed quasi-monolithic silicon/piezostack spm scanning stage. Microelectron Eng 98:520–523
Hariharan P (2007) Basics of interferometry, 2nd edn. Academic Press, London
Heidenhain exposed linear encoders. Online: www.heidenhain.com.
Hicks TR, Atherton PD, Xu Y, McConnell M (1997) The nanopositioning book. Queensgate Intstruments Ltd, Berkshire
Humphris A, McConnell M, Catto D (2006) A high-speed atomic force microscope capable of video-rate imaging. Microscopy and analysis: SPM supplement, pp 29–31, Mar 2006
ISO 5725 (1994) Accuracy (trueness and precision) of measurement methods and results
ISO/IEC Guide 98:1993 (1994) Guide to the expression of uncertainty in measurement. Interational Organization for Standardization
JCGM 200:2008 (2008) International vocabulary of metrology basic and general concepts and associated terms (VIM), 3rd edn
Karrai K, Braun P (2010) Miniature long-range laser displacement sensor. In: Proceedings Actuator Conference, Bremen, pp 285–288, June 2010
Kartik V, Sebastian A, Tuma T, Pantazi A, Pozidis H, Sahoo DR (2012) High-bandwidth nanopositioner with magnetoresistance based position sensing. Mechatronics 22(3):295–301
Kester W (2002) Sensor signal conditioning. Analog Devices, Newnes
Khiat A, Lamarque F, Prelle C, Pouille P, Leester-Schädel M, Büttgenbach S (2010) Two-dimension fiber optic sensor for high-resolution and long-range linear measurements. Sens Actuators A Phys 158(1):43–50
Kim M, Moon W, Yoon E, Lee K-R (2006) A new capacitive displacement sensor with high accuracy and long-range. Sens Actuators A Phys 130–131(14):135–141
Kobayashi M, Sumitomo K, Torimitsu K (2007) Real-time imaging of DNA streptavidin complex formation in solution using a high-speed atomic force microscope. Ultramicroscopy 107(2–3):184–190
Kovacs GTA (1998) Micromachined transducers sourcebook. McGraw Hill, Boston
Kuijpers AA, Krijnen GJM, Wiegerink RJ, Lammerink TSJ, Elwenspoek M (2003) 2d-finite-element simulations for long-range capacitive position sensor. J Micromech Microeng 13(4):S183–S189

Kuijpers AA, Krijnen GJM, Wiegerink RJ, Lammerink TSJ, Elwenspoek M (2006) A micromachined capacitive incremental position sensor: part 1 analysis and simulations. J Micromech Microeng 16(6):S116–S124
Kuijpers AA, Krijnen GJM, Wiegerink RJ, Lammerink TSJ, Elwenspoek M (2006) A micromachined capacitive incremental position sensor: part 2 experimental assessment. J Micromech Microeng 16(6):S125–S134
Lantz MA, Binnig GK, Despont M, Drechsler U (2005) A micromechanical thermal displacement sensor with nanometre resolution. Nanotechnology 16(8):1089–1094
Leang KK, Zou Q, Devasia S (2009) Feedforward control of piezoactuators in atomic force microscope systems. Control Syst Mag 29(1):70–82
Lee J-Y, Chen H-Y, Hsu C-C, Wu C-C (2007) Optical heterodyne grating interferometry for displacement measurement with subnanometric resolution. Sens Actuators A Phys 137(1):185–191
Lee J-I, Huang X, Chu P (2009) Nanoprecision MEMS capacitive sensor for linear and rotational positioning. J Microelectromech Syst 18(3):660–670
Lee S-C, Peters RD (2009) Nanoposition sensors with superior linear response to position and unlimited travel ranges. Rev Sci Instrum 80(4):045109
Li Q, Ding F (2005) Novel displacement eddy current sensor with temperature compensation for electrohydraulic valves. Sens Actuators A Phys 122(1):83–87
Lu T-F, Handley D, Yong YK (2004) Position control of a 3 dof compliant micro-motion stage. In: Proceedings control, automation, robotics and vision conference, vol 2, pp 278–1274
Maess J, Fleming AJ, Allgöwer F (2008) Simulation of dynamics-coupling in piezoelectric tube scanners by reduced order finite element models. Rev Sci Instrum 79:015105
Merry R, Uyanik M, van de Molengraft R, Koops R, van Veghel M, Steinbuch M (2009) Identification, control and hysteresis compensation of a 3 DOF metrological AFM. Asian J Control 11(2):130–143
Merry R, Maassen M, van de Molengraft M, van de Wouw N, Steinbuch M (2011) Modeling and waveform optimization of a nano-motion piezo stage. IEEE/ASME Trans Mechatron 16(4):615–626
Messenger R, Aten Q, McLain T, Howell L (2009) Piezoresistive feedback control of a mems thermal actuator. J Microelectromech Syst 18(6):1267–1278
Michellod Y, Mullhaupt P, Gillet D (2006) Strategy for the control of a dual-stage nano-positioning system with a single metrology. In: Proceedings on robotics, automation and mechatronics, pp 1–8, June 2006
Mishra S, Coaplen J, Tomizuka M (2007) Precision positioning of wafer scanners: segmented iterative learning control for nonrepetitive disturbances. IEEE Control Syst 27(4):20–25
Moheimani SOR, Fleming AJ (2006) Piezoelectric transducers for vibration control and damping. Springer, London
Nyce DS (2004) Linear position sensors: theory and application. Wiley, Hoboken
Pantazi A, Sebastian A, Cherubini G, Lantz M, Pozidis H, Rothuizen H, Eleftheriou E (2007) Control of mems-based scanning-probe data-storage devices. IEEE Trans Control Syst Technol 15(5):824–841
Parkin S, Jiang X, Kaiser C, Panchula A, Roche K, Samant M (2003) Magnetically engineered spintronic sensors and memory. Proc IEEE 91(5):661–680
Picco LM, Bozec L, Ulcinas A, Engledew DJ, Antognozzi M, Horton M, Miles MJ (2007) Breaking the speed limit with atomic force microscopy. Nanotechnology 18(4):044030(1–4)
Preumont A (2006) Mechatronics: dynamics of electromechanical and piezoelectric systems. Springer, Heidelberg
Proksch R, Cleveland J Bocek D (2007) Linear variable differential transformers for high precision position measurements. US Patent 7,262,592, 2007
Roach SD (1998) Designing and building an eddy current position sensor. Sensors, Sept 1998. http://www.sensorsmag.com/sensors/electric-magnetic/designing-and-building-eddy-current-position-sensor-772

Sahoo DR, Sebastian A, Häberle W, Pozidis H, Eleftheriou E (2011) Scanning probe microscopy based on magnetoresistive sensing. Nanotechnology 22(14):145501
Salapaka SM, Salapaka MV (2008) Scanning probe microscopy. IEEE Control Syst Mag 28(2): 65–83
Schitter G, Stark RW, Stemmer A (2002) Sensors for closed-loop piezo control: strain gauges versus optical sensors. Meas Sci Technol 13:N47–N48
Schitter G, Åström KJ, DeMartini BE, Thurner PJ, Turner KL, Hansma PK (2007) Design and modeling of a high-speed AFM-scanner. IEEE Trans Control Syst Technol 15(5):906–915
Schitter G, Thurner PJ, Hansma PK (2008) Design and input-shaping control of a novel scanner for high-speed atomic force microscopy. Mechatronics 18(5–6):282–288
Sebastian A, Pantazi A, Pozidis H, Elefthriou E (2008) Nanopositioning for probe-based data storage. IEEE Control Syst Mag 28(4):26–35
Sebastian A, Wiesmann D (2008) Modeling and experimental identification of silicon microheater dynamics: a systems approach. J Microelectromech Syst 17(4):911–920
Sebastian A, Pantazi A (2012) Nanopositioning with multiple sensors: a case study in data storage. IEEE Trans Control Syst Technol 20(2):382–394
Shan Y, Speich J, Leang K (2008) Low-cost IR reflective sensors for submicrolevel position measurement and control. Mechatronics IEEE/ASME Trans 13(6):700–709
Sirohi J, Chopra I (2000) Fundamental understanding of piezoelectric strain sensors. J Intell Mater Syst Struct 11:246–257
Sirohi RS (2009) Optical methods of measurement: wholefield techniques. CRC Press, Boca Raton
Smith CS (1954) Piezoresistance effect in germanium and silicon. Phys Rev 94(1):42–49
Sommargren GE (1986) A new laser measurement system for precision metrology. In: Proceedings of precision engineering conference, Dallas, Nov 1986
Tseng AA (ed) (2008) Nanofabrication: fundamentals and applications. World Scientific, Singapore
Tseng AA, Jou S, Notargiacomo A, Chen TP (2008) Recent developments in tip-based nanofabrication and its roadmap. J Nanosci Nanotechnol 8(5):2167–2186
Vicary JA, Miles MJ (2008) Pushing the boundaries of local oxidation nanolithography: short timescales and high speeds. Ultramicroscopy 108(10):1120–1123
Yong YK, Ahmed B, Moheimani SOR (2010) Atomic force microscopy with a 12-electrode piezoelectric tube scanner. Rev Sci Instrum 81(1–10):033701
Yong YK, Fleming AJ, Moheimani SOR (2013) A novel piezoelectric strain sensor for simultaneous damping and tracking control of a high-speed nanopositioner. IEEE/ASME Trans Mechatron 18(3):1113–1121
Zheng J, Salton A, Fu M (2011) Design and control of a rotary dual-stage actuator positioning system. Mechatronics 21(6):1003–1012
Zhu YK, Moheimani SOR, Yuce MR (2011) Simultaneous capacitive and electrothermal position sensing in a micromachined nanopositioner. Electron Device Lett 32(8):1146–1148
Zhu Y, Bazaei A, Moheimani SOR, Yuce M (2011) Design, modeling and control of a micromachined nanopositioner with integrated electrothermal actuation and sensing. IEEE/ASME J Microelectromech Syst 20(3):711–719

Chapter 6
Shunt Control

As discussed in Chap. 1, the foremost speed limitation in nanopositioning systems arised from the first mechanical resonance mode. This resonance severely limits the speed of nanopositioning systems both in open-loop due to induced vibration, and in closed-loop due to low gain-margin. Attenuation of lightly damped resonance modes can provide extremely large improvements in positioning performance and is hence a foremost priority.

In this chapter, the technique of piezoelectric shunt damping, previously resident in the field of smart structures, is applied to damp mechanical resonance. By connecting an LCR impedance to the terminals of a piezoelectric actuator, mechanical resonance modes can be reduced in magnitude by more than 20 dB. This allows a proportionate increase in both open-loop operating speed, and closed-loop control bandwidth. Beneficially, piezoelectric shunt damping does not require position sensors or, for that matter, any mechanical modifications whatsoever.

6.1 Introduction

As discussed in Chap. 3, nanopositioning systems often exhibit lightly damped low-frequency mechanical resonances. This is particularly true of piezoelectric tube nanopositioners which are often designed with a large length to diameter ratio for high scan ranges. A consequence of designing tubes with large length/diameter ratios is low mechanical resonance frequency. This has been a fundamental problem since the inception of piezoelectric tube scanners and is worsened by the fact that significant payload masses are required in many applications.

In this chapter, a technique for reducing vibration is described that requires only a capacitor, resistor, and inductor connected to the terminals of a piezoelectric actuator (Fleming and Moheimani 2006). It can be used alone, or as part of a feedback controller with improved bandwidth and stability margins. Usually referred to as piezoelectric shunt damping, this technique results in a damped electrical resonance

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_6,

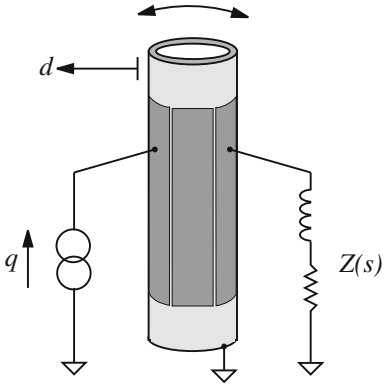


Fig. 6.1 Charge-driven tube scanner with piezoelectric shunt damping circuit

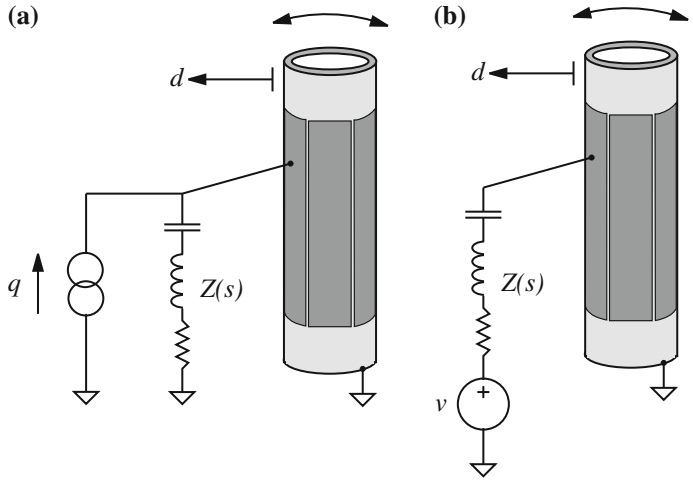


Fig. 6.2 **a** Charge-driven tube scanner. **b** Voltage equivalent circuit

capable of significantly reducing the magnitude of one or more structural modes. Figure 6.1 shows an inductor and resistor connected to the terminals of a charge-driven piezoelectric tube. In this configuration, the inductor and resistor are tuned to damp the first x-axis cantilever mode. Undesired resonance excitation due to scanning and external disturbance is attenuated. Piezoelectric shunt control has also been successfully applied to stack-based positioning systems (Eielsen and Fleming 2010).

Piezoelectric shunt damping requires no feedback sensor and is thus immune to the usual problems of low bandwidth and measurement noise associated with optical and capacitive sensors. Furthermore, as shown in Fig. 6.2, the shunt impedance $Z(s)$ can be applied to the same electrode as the driving charge or voltage source. If the actuator

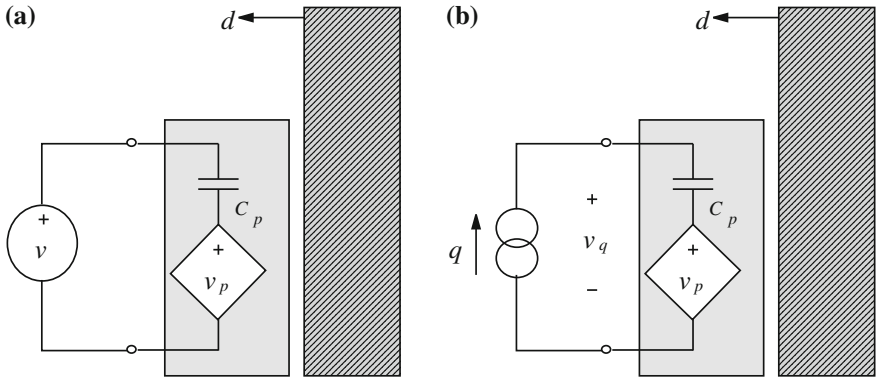


Fig. 6.3 **a** A voltage and **b** charge-driven piezoelectric tube

is a piezoelectric tube, this allows the opposite electrode to be employed for increasing the scan range or as a piezoelectric strain sensor. Piezoelectric shunt damping can be implemented independantly or in conjuction with secondary feedback or feedforward control system.

In the following section, the electromechanical model of a piezoelectric nanopositioner is derived. This is used to analyze the effect of a connected shunt impedance. Implementation issues are then discussed in Sect. 6.3, followed by experimental results, and a chapter summary in Sects. 6.4 and 6.5.

6.2 Shunt Circuit Modeling

The modeling of piezoelectric actuators with attached resonant shunt circuits has traditionally been performed using voltage-driven models. Here, only charge-driven models are utilized. The following subsection introduces the models required to simulate the effect of an attached shunt circuit. Traditional voltage-driven models are initially discussed then related to their charge-driven equivalents as used throughout.

6.2.1 Open-Loop

The open-loop dynamics of a piezoelectric nanopositioner are first considered in the following. Although a piezoelectric tube actuator is used as an example, the modeling process is equally applicable to stack-based nanopositioners.

The electrically equivalent model of a voltage and charge-driven piezoelectric tube is shown in Fig. 6.3. Each electrode acts as a piezoelectric transducer, represented by a strain-dependent voltage source v_p and series capacitor C_p. The polarization vector is assumed to be oriented radially outward, in this case, a positive voltage

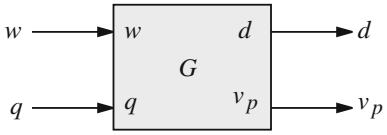


Fig. 6.4 The nanopositioner model describing the deflection d and strain voltage v_p in response to an applied charge q and disturbance w

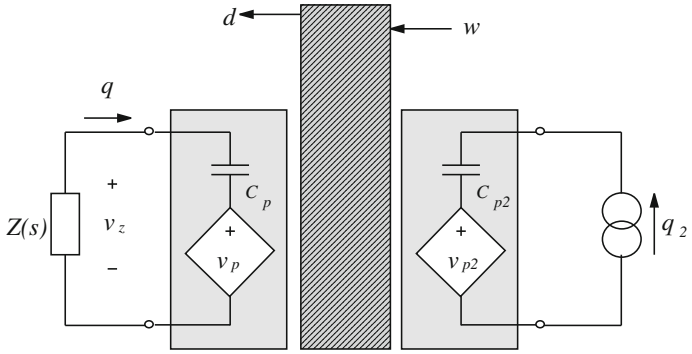


Fig. 6.5 The electrical equivalent of a charge-driven piezoelectric tube with attached shunt circuit

or charge results in a positive deflection. We are interested in the transfer functions from an applied voltage v to the resulting piezoelectric voltage v_p and tip translation d, that is,

$$G_{vv}(s) = \frac{v_p(s)}{v(s)} \quad G_{dv}(s) = \frac{d(s)}{v(s)}\,. \tag{6.1}$$

The transfer functions $G_{vv}(s)$ and $G_{dv}(s)$ can be derived analytically or determined experimentally. Due to the difficulties involved with modeling complicated geometries from first principles, empirical models obtained through system identification are preferable.

In the case of charge actuation, Fig. 6.3b, equivalent transfer functions can be derived. Kirchoff's Voltage Law for the loop is,

$$\frac{-q}{C_p} - v_p + v_q = 0. \tag{6.2}$$

Substituting $v_p = G_{vv} v_q$ and simplifying yields

$$G_{vq}(s) = \frac{v_p(s)}{q(s)} = \frac{1}{C_p} \frac{G_{vv}(s)}{1 - G_{vv}(s)}. \tag{6.3}$$

The displacement transfer function can be derived in a similar fashion,

$$G_{dq}(s) = \frac{d(s)}{q(s)} = \frac{1}{C_p} \frac{G_{dv}(s)}{1 - G_{vv}(s)} \tag{6.4}$$

Off resonance, where $G_{vv}(s) \ll 1$

$$G_{vq}(s) \approx \frac{G_{vv}(s)}{C_p} \quad G_{dq}(s) \approx \frac{G_{dv}(s)}{C_p} \tag{6.5}$$

Thus the relationship between charge and voltage actuation is revealed. Due to the benefits in reducing hysteresis, only charge actuation will be considered in the proceeding sections.

In addition to a charge input, the possibility for a disturbance input w is also desirable. The signal w can be used to study the regulation or rejection of environmental noise. In the following sections, the tube system will be referred to as G, a multi-input multi-output system describing the deflection d and piezoelectric voltage v_p in response to a driving charge q and disturbance w. The inputs and outputs are illustrated in Fig. 6.4. Such a realization is advantageous as the system G will later be identified directly from experimental data using system identification.

6.2.2 Shunt Damping

Although first appearing in Forward (1979), the concept of piezoelectric shunt damping is mainly attributed to Hagood and Von Flotow (1991). A series inductor-resistor network, as shown in Fig. 6.1, was demonstrated to significantly reduce the magnitude of a single structural mode. Together with the inherent piezoelectric capacitance, the network is tuned to the resonance frequency of a single structural mode. Analogous to a tuned mechanical absorber, additional dynamics introduced by the shunt circuit act to increase the effective structural damping Hagood and Von Flotow (1991).

The equivalent electrical model of a shunted piezoelectric tube nanopositioner (as shown in Fig. 6.1) is illustrated in Fig. 6.5. To find the transfer function relating displacement d to the driving charge q_2 Kirchoff's Voltage Law is first applied to the impedance loop then $v_z = -qsZ(s)$ is substituted,

$$\frac{-q(s)}{C_p} - v_p(s) + -q(s)sZ(s) = 0. \tag{6.6}$$

When the opposing tube electrodes are equal in dimension, the charges q and q_2 have an equal but opposite influence on the tube deflection d and v_p. Furthermore

$$v_p = -v_{p2} \tag{6.7}$$

$$\frac{v_p(s)}{q(s)} = \frac{v_{p2}(s)}{q_2(s)} = \frac{-v_p(s)}{q_2(s)} = G_{vq}(s) \tag{6.8}$$

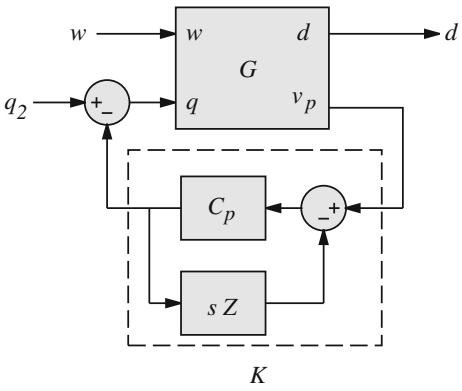


Fig. 6.6 The equivalent feedback diagram where an electrical impedance is connected to the terminals of one tube electrode and the other is driven with charge

The principle of superposition can be applied to find an expression for v_p.

$$v_p(s) = G_{vq}(s)q(s) - G_{vq}(s)q_2(s). \tag{6.9}$$

Rearranging (6.9) in terms of q_2 and substituting into (6.6) yields

$$\frac{v_p(s)}{q_2(s)} = \frac{-G_{vq}(s)}{1 + G_{vq}(s)K(s)} \tag{6.10}$$

where

$$K(s) = \frac{C_p}{1 + C_p s Z(s)}. \tag{6.11}$$

The shunted displacement transfer function can be derived in a similar manner,

$$\frac{d(s)}{q_2(s)} = \frac{G_{dq}(s)}{1 + G_{vq}(s)K(s)} \tag{6.12}$$

Using the principle of superposition, the influence of an external disturbance w can also be included,

$$d(s) = \frac{1}{1 + G_{vq}(s)K(s)} \left(G_{dq}(s)q_2(s) + G_{dw}(s)w(s)\right) \tag{6.13}$$

where G_{dw} is the transfer function measured from an external force w to the displacement d.

From Eqs. (6.12) and (6.13) it is concluded that the presence of an electrical shunt impedance can be viewed equivalently as a strain-voltage feedback control system. A diagrammatic representation of Eq. (6.13) is shown in Fig. 6.6. Further interpretation and analysis can be found in Moheimani et al. (2003).

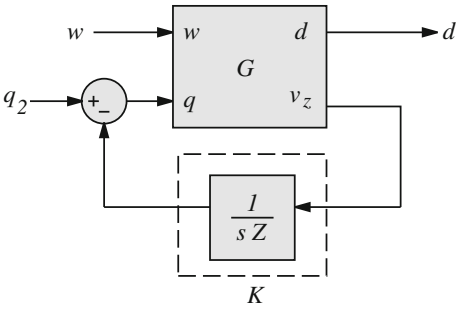


Fig. 6.7 An alternative feedback interpretation considering the terminal voltage v_z rather than the piezoelectric voltage v_p

In some cases (where a second electrode is not available), it may be difficult to obtain a model describing the piezoelectric voltage v_p directly. In such cases, the terminal voltage v_z can also be considered. The equivalent terminal-voltage feedback diagram is shown in Fig. 6.7. v_z is related to v_p by

$$v_z = v_p + \frac{1}{C_p} q, \tag{6.14}$$

that is,

$$\frac{v_z(s)}{q(s)} = G_{vq}(s) + \frac{1}{C_p}. \tag{6.15}$$

Equations (6.6)–(6.13) can be modified accordingly.

6.2.2.1 Hybrid Operation

As mentioned in the introduction, it is advantageous to connect the shunt impedance and driving charge source to the same electrode. In the case of piezoelectric stack actuators, this is the only option. This scenario is depicted in Fig. 6.8. In this subsection, the electrical filtering effect of $Z(s)$ on q_2 is derived. If such a filtering effect can be inverted, the charge source q_2 can be used for scanning, analogous to the case where a shunt impedance is attached to an independent electrode.

Writing Kirchoff's Voltage Law around the loop,

$$-\frac{q}{C_p} - v_p + v_z = 0, \tag{6.16}$$

and substituting the following,

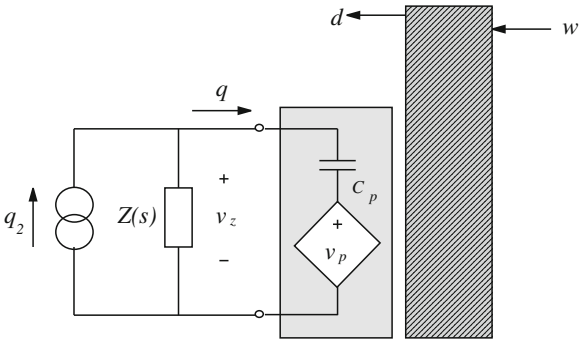


Fig. 6.8 A charge-driven tube electrode with attached parallel shunt circuit

$$q(s) = -\frac{v_z(s)}{sZ(s)} + q_2(s), \tag{6.17}$$

results in the loop equation

$$-\frac{q(s)}{C_p} - v_p(s) - q(s)sZ(s) + q_2(s)sZ(s) = 0. \tag{6.18}$$

Given that $v_p = G_{vq}q$, we can substitute $q = v_p/G_{vq}$ into (6.18). After simplification, the transfer function from q_2 to v_p can be found:

$$\frac{v_p(s)}{q_2(s)} = \frac{sK(s)Z(s)G_{vq}(s)}{1 + G_{vq}(s)K(s)},$$

where K is as given in (6.11). Similarly,

$$\frac{d(s)}{q_2(s)} = \frac{sK(s)Z(s)G_{dq}(s)}{1 + G_{vq}(s)K(s)}. \tag{6.19}$$

Unlike the case in Sect. 6.2.2, the impedance $Z(s)$ distorts the tube transfer function from the driving charge q_2 to the deflection d. Rather than simply adding a strain feedback controller to the mechanical system, the transfer function from q_2 to d now also contains a filter $F(s) = sK(s)Z(s)$. An equivalent feedback diagram is shown in Fig. 6.9.

An obvious technique for eliminating the effect of the filter $F(s)$ is to prefilter the driving charge with $F^{-1}(s)$. Fortunately, this prefiltering and inversion is straightforward to implement in practice. This solution is discussed in Sect. 6.3.

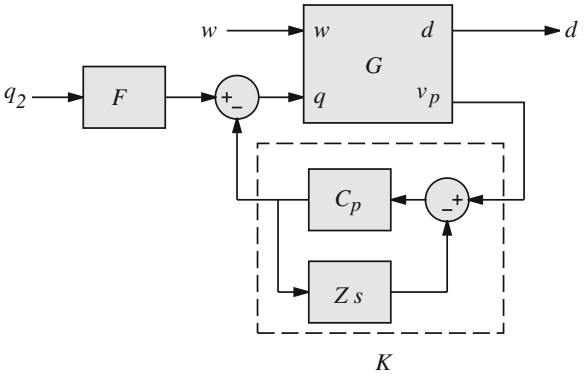


Fig. 6.9 The equivalent feedback diagram where the driving charge and shunt impedance are applied to the same electrode (as shown in Fig. 6.8)

6.2.2.2 Shunt Impedance Design

The Smart Structures and Vibration Control literature contain a multitude of passive, active, linear, and nonlinear piezoelectric shunt impedance designs [reviewed in Fleming (2004) and Moheimani (2003)]. However, only a small subset are suitable damping piezoelectric nanopositioners. The so-called resonant linear shunts meet all of the requisite criteria: They are easy to design, implement, and tune; they offer excellent damping performance (especially for single modes of vibration); they are strictly passive and inject no harmonics; and finally, their presence influences the mechanical dynamics only over a small frequency range. Resonant linear shunts have been shown to emulate the effect of a tuned-mass mechanical absorber Hagood and Von Flotow (1991).

After examination of various impedance designs, the LCR circuit depicted in Fig. 6.2 was found to offer good performance. The presence of a series capacitance is necessitated by the requirement for DC tracking. If the impedance of the network was not infinity at DC, constant tube deflections would require a ramp signal in charge (eventually saturating the amplifier), this is reflected in the scan filter $F(s)$ and its inverse $F(s)^{-1}$.

To damp a single mode of structural vibration, the circuit inductance L, capacitance C, and piezoelectric capacitance C_p are tuned to resonate at the target mechanical frequency ω_1. Although the capacitance value C is essentially arbitrary, values of 1–10 times the piezoelectric capacitance have been found suitable. To equate the frequency of electrical resonance to mechanical resonance, the inductor is tuned as follows:

$$L = \frac{C + C_p}{C C_p \omega_1}. \tag{6.20}$$

The resistance value, dependent on the inherent system damping, is most easily found experimentally. For such systems, resistances in the order of 1 kΩ are typical.

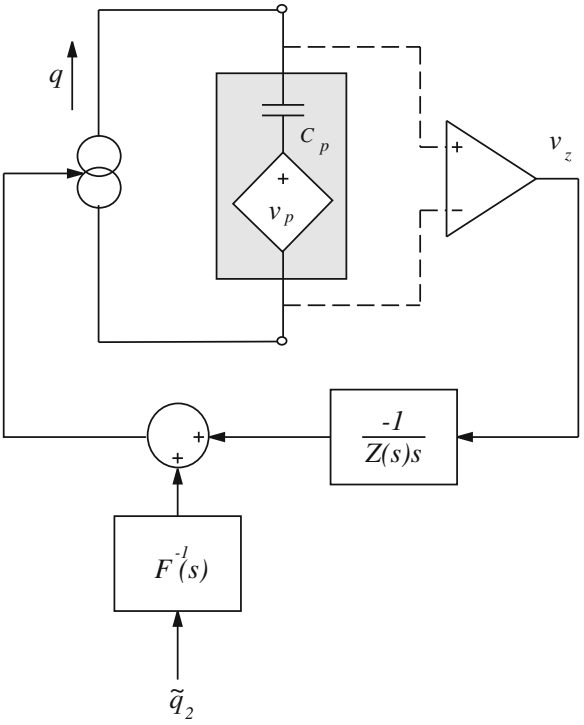


Fig. 6.10 Schematic diagram of a charge-driven tube with integrated shunt circuit

6.3 Implementation

Resonant piezoelectric shunt damping circuits require impractically large values of inductance, typically in the tens of Henrys. For this reason, the shunt damping circuit will be synthesized artificially using the charge amplifier. Consider the schematic shown in Fig. 6.10. Neglecting the filter $F^{-1}(s)$ and input q_2, the charge applied to the piezoelectric tube is equal to

$$q = v_z \frac{-1}{sZ(s)}. \tag{6.21}$$

The impedance (or admittance) experienced by the piezoelectric transducer can be calculated by examining the ratio of current to voltage at its terminals. As the current is equal to $-\dot{q}$, and q is defined by (6.21), the impedance presented to the terminals is simply $Z(s)$ (as defined by the filter in Fig. 6.10). By implementing the filter $\frac{-1}{sZ(s)}$ any arbitrary impedance can be presented to the terminals of the transducer. Simple techniques for designing analog and digital filters that represent $\frac{1}{Z(s)}$ can be found

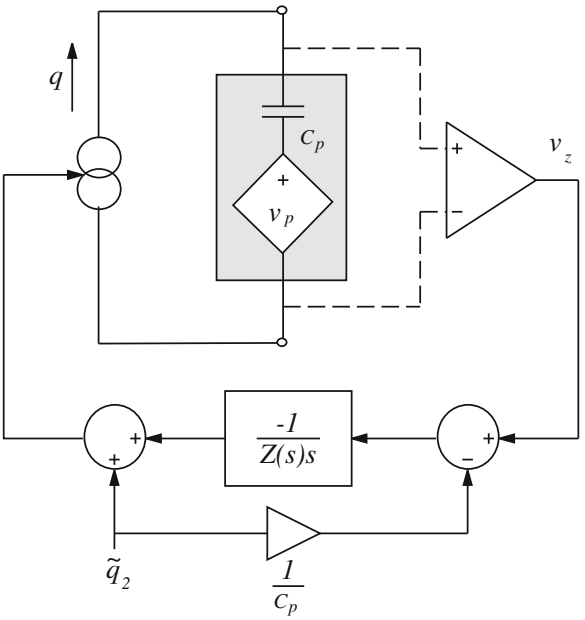


Fig. 6.11 Simplified diagram of a charge amplifier with integrated shunt impedance

in Fleming and Moheimani (2004b). In this work, a dSpace DSP system is used to implement and tune the filter $\frac{-1}{sZ(s)}$.

In addition to the charge required for shunt impedance synthesis, the additive charge q_2 is used for tube scanning. As mentioned in Sect. 6.2.2.1, the additive charge q_2 requires a filter $F^{-1}(s)$ to compensate for the electrical dynamics of the shunt impedance when attached to the same electrode.

A substantial simplification of the system shown in Fig. 6.10 can be made by studying the structure of the filter $F^{-1}(s)$,

$$F^{-1}(s) = \frac{1}{sK(s)Z(s)} = \frac{1 + C_p sZ(s)}{C_p sZ(s)} = \frac{1}{C_p sZ(s)} + 1. \tag{6.22}$$

Considering that the transfer function $\frac{1}{sZ(s)}$ has already been implemented, $F^{-1}(s)$ can be replaced as shown in Fig. 6.11.

6.4 Experimental Results

In this section, a shunt circuit and charge amplifier are employed to drive a piezoelectric tube positioner in one dimension. The tube construction is described in Sect. 3.1.2. Parameters of the shunt impedance and amplifier are listed in Table 6.1.

Table 6.1 Parameters of the charge amplifier and shunt impedance

Charge gain	77.8 nC/V
Voltage Measurement gain	0.1 V/V
L	2.9 H
C	50 nF
R	3.3 kΩ

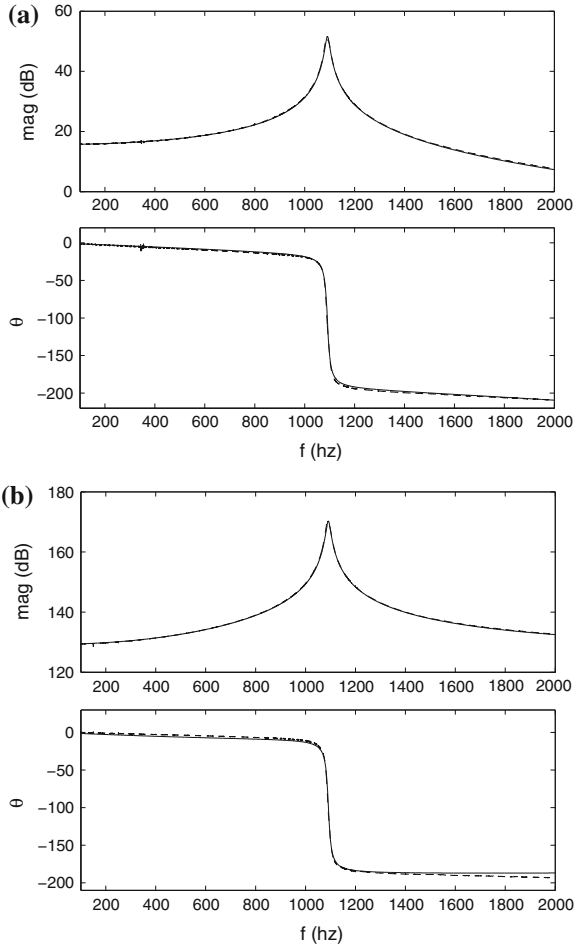


Fig. 6.12 Frequency response of the transfer functions G_{dq} and G_{vq}. Identified model (—), measured (- -). **a** G_{dq}, **b** G_{vq} (m/V)

6.4.1 Tube Dynamics

The first step in designing a shunt circuit is to obtain a model for the piezoelectric tube. This is achieved by measuring, then fitting a model to the transfer functions from charge input to strain-voltage and displacement. The measured frequency responses and model responses are plotted in Figs. 6.12a, b. The system model G

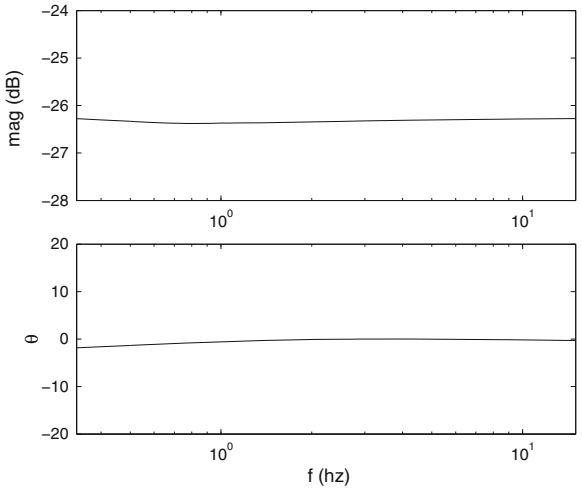


Fig. 6.13 Charge amplifier low-frequency tracking performance. Measured from the charge reference signal (V) to the instrumented load voltage across a 5-nF dummy load

(shown in Fig. 6.4), was obtained by frequency domain subspace system identification (McKelvey et al. 1996). The identification[1] required 12 MIMO data points to return a single input, two output model of order 2. An excellent fit is observed in the frequency domain.

The nominal first resonance frequency and DC charge sensitivity of the tube were measured to be 1,088 Hz and 5.7 m/C (5.7 μm/μC).

6.4.2 Amplifier Performance

Both the low-frequency scanning and high-frequency vibration damping depend on the performance of the charge amplifier and related instrumentation. In the following we examine the two characteristics of foremost importance: low-frequency charge regulation—the ability of the amplifier to reproduce low-frequency inputs without drift, and the bandwidth of charge dominance—the frequency range where hysteresis will be reduced due to dominant charge feedback.

The (low-frequency) transfer function measured from an applied reference signal to the actual charge deposited on a 5 nF dummy load is shown in Fig. 6.13. Excellent low-frequency tracking from 15 mHz to 15 Hz is exhibited by the amplifier and instrumentation. As discussed in Sect. 12.2, the bandwidth of charge dominance was ascertained by zeroing the charge reference and introducing an internal load voltage. The transfer function measured from the internal voltage to the voltage measured

[1] An implementation of the algorithm is freely available by contacting the first author.

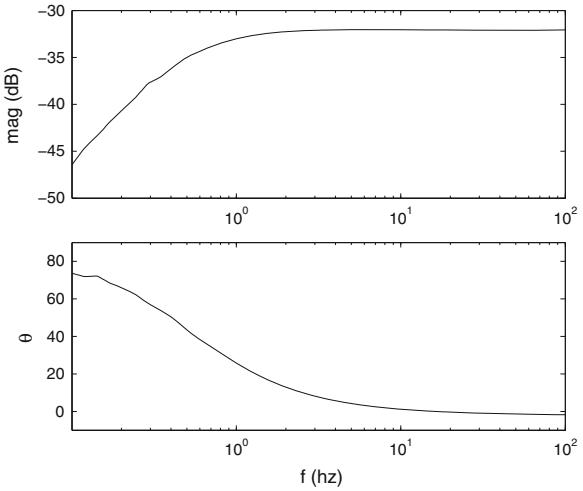


Fig. 6.14 Charge dominance bandwidth. Measured from the internal tube strain voltage v_p to the load voltage

across the load is shown in Fig. 6.14. A charge dominance bandwidth of 0.8 Hz is observed. Frequencies above this bandwidth will experience the linearity benefits of charge actuation.

To illustrate the benefits of charge actuation, the hysteresis exhibited under voltage and charge drive is compared in Fig. 6.15a, b. Hysteresis is reduced by approximately 89 %. Percentage reduction is calculated by measuring the maximum excursion in the minor axis of each plot, then taking the ratio $100 \times \frac{\text{voltage}}{\text{charge}}$. It should be noted that a scan range of ± 3 μm is around 20 % of the full scale deflection, it is often assumed that hysteresis is negligible at such low drives. Similar plots for the same apparatus with a $\pm 8\,\mu$m drive can be found in Fleming and Moheimani (2004a), a greater hysteresis is exhibited, and also heavily reduced through the use of a similar charge drive.

6.4.3 Shunt Damping Performance

6.4.3.1 Scan-Induced Vibration Suppression

Whilst scanning at high frequencies, the greatest cause of tracking error is due to high frequency harmonics exciting the mechanical resonance. The influence of the shunt impedance can be observed to significantly increase the effective damping in Fig. 6.16a. The simulated response shown in Fig. 6.16b shows a good correlation with experimental results. The equivalent decrease in settling time can be observed in Fig. 6.17a.

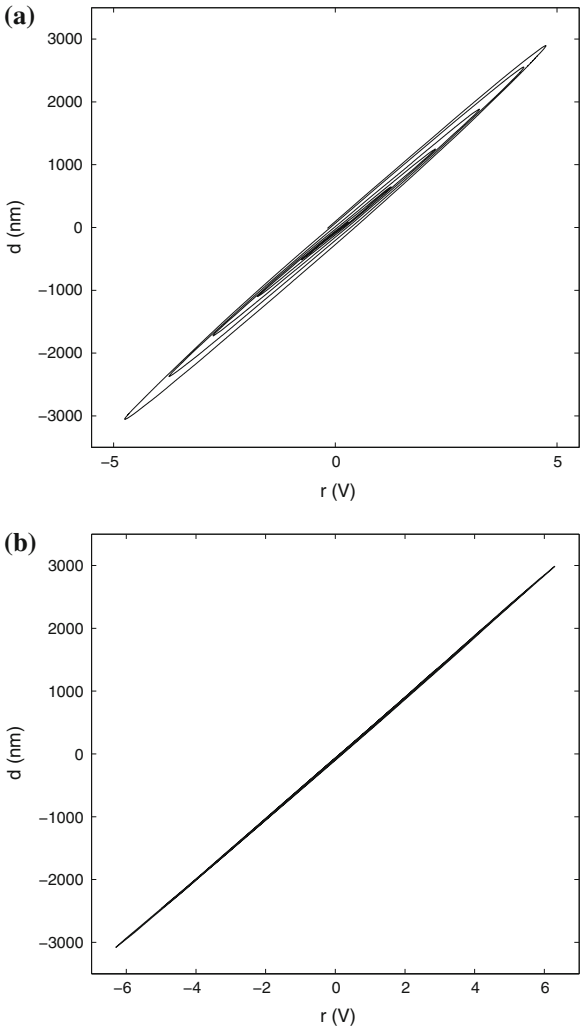


Fig. 6.15 Tube displacement in response for an applied voltage (**a**) and applied charge (**b**). The input signal is a 10-Hz ramped sine wave

To illustrate the improvement in triangular scanning fidelity, an unfiltered 46 Hz Triangle wave was applied to the system. The frequency and lack of filtering was chosen to illustrate the worst-case induced ripple. In practice, the triangle would be filtered or passed through a feedforward controller to reduce vibration. Regardless of the ripple magnitude, the presence of a shunt circuit provides the same decrease in settling time. At high speeds, significant increases in fast-axis resolution can be expected. In the case where feedforward vibration control (Croft et al. 2001) is

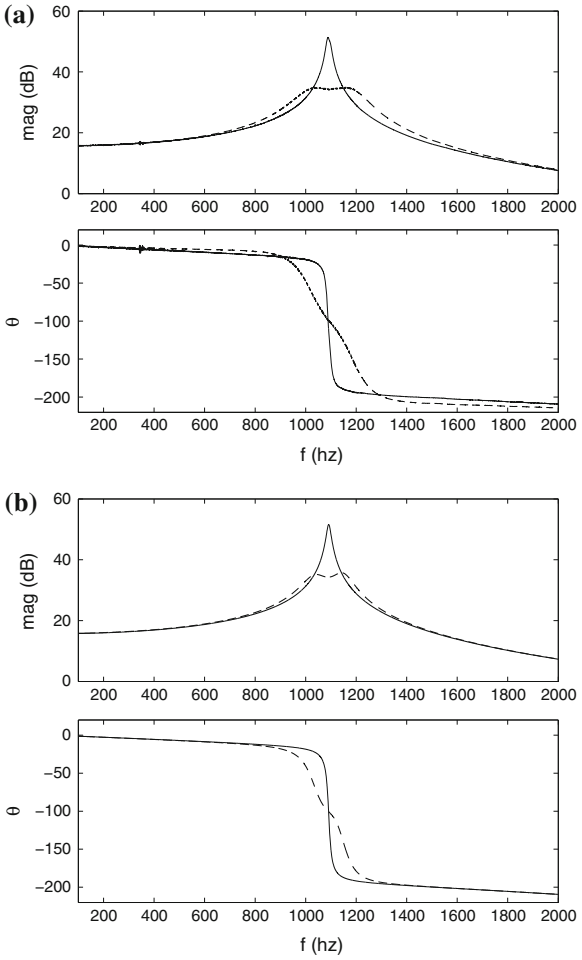


Fig. 6.16 Frequency response of the open-loop (—) and shunt-damped (- -) tube measured from the additive charge input q_2 (C) to the tip displacement d (m). **a** Experimental response, **b** simulated response

applied, the damped mechanical system allows a less agressive feedforward controller and greater immunity to modeling error.

6.4.3.2 Externally Induced Vibration

Another significant source of tracking error is external mechanical noise. Due to the highly resonant nature of the tube, high frequency noise components can excite the mechanical resonance and lead to large erroneous excursions. By applying a voltage to an opposite electrode, we can simulate the effect of a strain disturbance.

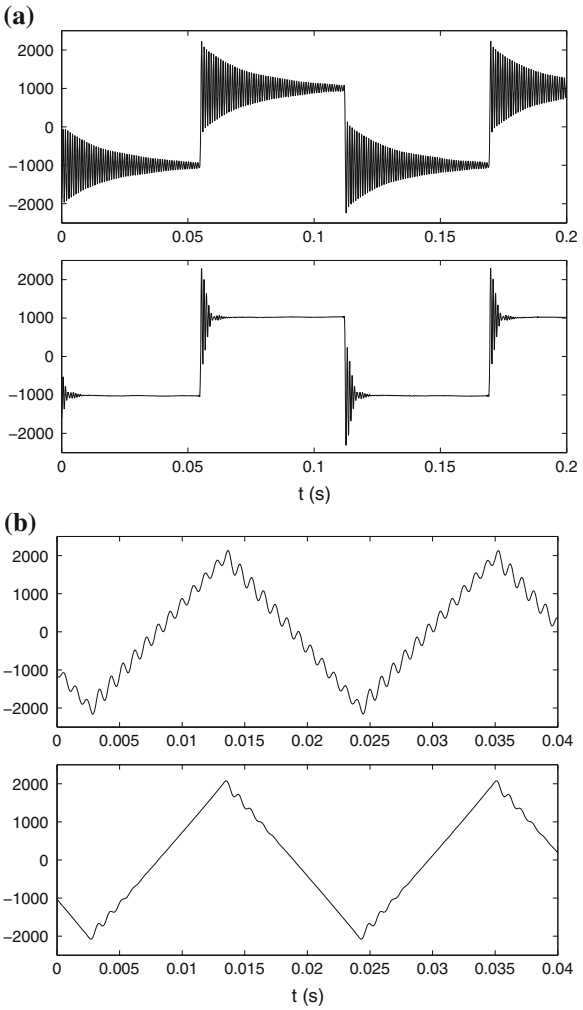


Fig. 6.17 Tube deflection (in nm) resulting from a square wave and triangule wave excitation. (*Top*) Uncontrolled, and (*Bottom*) with LCR shunt impedance, **a** Square wave excitation, **b** 46-Hz Triangle wave excitation

A significant damping of greater than 20 dB can be observed in Fig. 6.18. The effect of such damping can be observed in the time domain by applying a low-frequency scanning signal. With no scan-induced vibration, the external noise is dominant. The reduction of resonant vibration can be seen in Fig. 6.19.

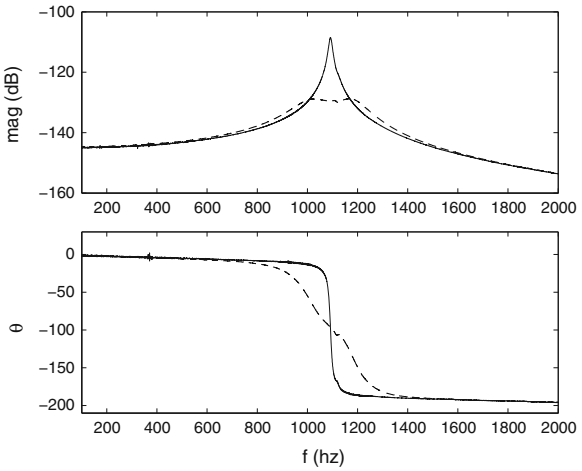


Fig. 6.18 Experimental response. The natural (—) and shunt-damped (- -) tube transfer function from the applied strain disturbance (in V) to the tip displacement d (m)

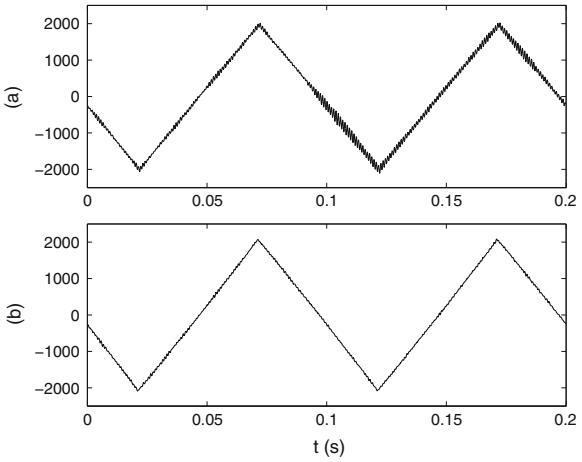


Fig. 6.19 Tube deflection (in nm) resulting from a 1.6-kHz band-limited uniformly distributed random strain disturbance. **a** Uncontrolled, and **b** with LCR shunt impedance

6.4.3.3 Low-Frequency Scanning

The final test of such an apparatus is the ability to track DC charge offsets. In Fig. 6.20 a low-frequency triangle signal was applied to the charge amplifier, at time 130 s a DC offset equivalent to around 1 μm was applied. Aside from the faithful reproduction of a 0.1-Hz triangle wave, the charge amplifier reproduces the offset without drift.

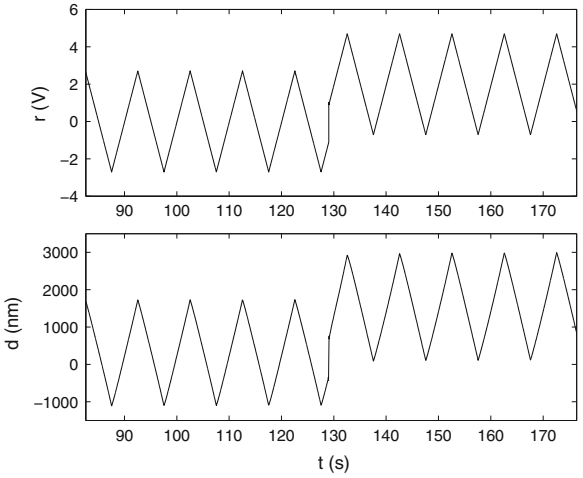


Fig. 6.20 Low-frequency scanning reference and resultant tube displacement with additive DC offset

6.5 Chapter Summary

In this chapter, piezoelectric shunt damping was applied to reduce scan-induced vibration. Piezoelectric shunt damping involves the connection of an electrical impedance to the terminals of a piezoelectric transducer. In experiments considering scan-induced and externally-induced vibration, an LCR network reduces the first resonance mode by 20-dB in magnitude. No feedback sensors are required.

Although charge-driven shunt-damped piezoelectric tubes can be combined with other feedback and feedforward controllers, the simplicity of implementation and performance warrants their use independently.

References

Croft D, Shed G, Devasia S (2001) Creep, hysteresis, and vibration compensation for piezoactuators: atomic force microscopy application. Trans ASME J Dyn Syst Meas Control 123:35–43

Eielsen AA, Fleming AJ (2010) Passive shunt damping of a piezoelectric stack nanopositioner. In: Proceedings of the American Control Conference, Baltimore, MD, June 2010, pp 4963–4968

Fleming AJ (2004) Synthesis and implementation of sensor-less shunt controllers for piezoelectric and electromagnetic vibration control. Ph.D. dissertation, The University of Newcastle, Callaghan 2308, Australia

Fleming AJ, Moheimani SOR (2004a) Hybrid DC accurate charge amplifier for linear piezoelectric positioning'. In: Proceedings of the 3rd IFAC symposium on mechatronic systems, Sydney, Australia

Fleming AJ, Moheimani SOR (2004b) Improved current and charge amplifiers for driving piezoelectric loads, and issues in signal processing design for synthesis of shunt damping circuits. Intell Mater Syst Struct 15(2):77–92

Fleming AJ, Moheimani SOR (2006) Sensorless vibration suppression and scan compensation for piezoelectric tube nanopositioners. IEEE Trans Control Syst Technol 14(1):33–44

Forward RL (1979) Electronic damping of vibrations in optical structures. Appl Opt 18(5):690–697

Hagood NW, Von Flotow A (1991) Damping of structural vibrations with piezoelectric materials and passive electrical networks. J Sound Vibr 146(2):243–268

McKelvey T, Akcay H, Ljung L (1996) Subspace based multivariable system identification from frequency response data. IEEE Trans Autom Control 41(7):960–978

Moheimani SOR (2003) A survey of recent innovations in vibration damping and control using shunted piezoelectric transducers. IEEE Trans Control Syst Technol 11(4):482–494

Moheimani SOR, Fleming AJ, Behrens S (2003) On the feedback structure of wideband piezoelectric shunt damping systems. Smart Mater Struct 12(1):49–56

Chapter 7
Feedback Control

Feedback control is the most commonly used technique for eliminating positioning errors in nanopositioning systems. This chapter provides an overview of feedback control techniques with an experimental comparison of integral control, inversion-based control, and IRC damping control. When the reference trajectory is periodic, repetitive control (RC) can significantly improve the tracking performance of a feedback loop. The RC approach is introduced for nanopositioning.

7.1 Introduction

When operated in open-loop, the static accuracy of a nanopositioning system is limited by piezoelectric hysteresis, creep, cross-coupling from other axes, external disturbances, and temperature drift. To eliminate or reduce these error sources, nanopositioning systems require some form of feedback or feedforward compensation.

As illustrated in Fig. 7.1, a feedback controller works by comparing the commanded position to the actual displacement. By minimizing the positioning error, a feedback controller can compensate for all forms of positioning errors that are within its effective bandwidth. Due to the simplicity and ability to compensate for a wide range of errors, feedback controllers are commonly used in commercial nanopositioning systems.

In applications where fast changes in the reference siganl occur, large positioning errors can also arise from the mechanical resonances of the stage. To avoid excitation of the mechanical resonance in open-loop, the frequency of driving signals is limited to between 1 and 10 % of the resonance frequency (depending on the signal). In applications where the frequency of driving signals should be maximized, for example, in high-speed atomic force microscopy (Ando et al. 2005; Schitter et al. 2007; Humphris et al. 2005; Rost et al. 2005), the nanopositioner is operated in open-loop with driving signals that are shaped to reduce harmonic content. Although

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_7,

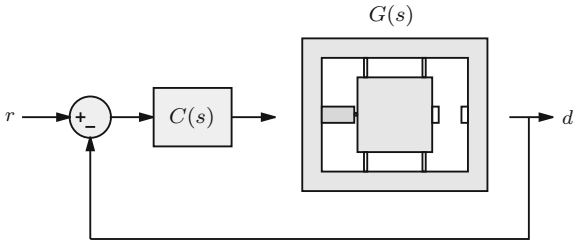


Fig. 7.1 A nanopositioner G in a displacement feedback control loop. The feedback controller $C(s)$ drives the nanopositioner $G(s)$ so that the difference between the reference r and measured position d is minimized

command shaping techniques, reviewed in (Fleming and Wills 2009), can provide a fast response, they do not account for nonlinearity or disturbances.

Since the first resonance mode typically dominates the response, the dynamics of a nanopositioner can be approximated by a second-order low-pass system

$$G(s) = \frac{\omega_n^2}{s^2 + 2\omega_n \zeta s + \omega_n^2}, \tag{7.1}$$

where ω_n and ζ are the natural frequency and damping ratio. Although a second-order system is a highly simplified model, it is sufficient to demonstrate the limitations experienced by some feedback controllers. The magnitude and phase responses of this system are plotted in Fig. 7.4.

The first closed-loop nanopositioning systems were piezoelectric tube scanners with capacitive (Griffith et al. 1990) or optical sensors (Barrett and Quate 1991). Although the early controllers were primarily manually tuned, model-based lead-lag and $\mathcal{H}_\infty$ controllers were also investigated (Tamer and Dahleh 1994).

To improve the gain-margin and closed-loop bandwidth of nanopositioning systems, notch filters or inversion filters can be effective (Leang and Devasia 2007). Such techniques can provide excellent closed-loop bandwidth, up to or greater than the resonance frequency (Abramovitch et al. 2008). However, to achieve high performance, an extremely accurate system model is required. Due to the dependency on model accuracy, a small change in the system dynamics can result in instability. For example, a resonance frequency reduction of 10 % may cause a high-gain inversion-based feedback controller to become unstable. In many applications, the high sensitivity to modeling error is unacceptable as the load mass and resonance frequency of a nanopositioner can vary significantly during service. As a result, high-performance inversion-based controllers are only applied in applications where the resonance frequency is stable, or when the feedback controller can be continually recalibrated (Abramovitch et al. 2008).

Damping control is an alternative method for reducing the bandwidth limitations imposed by mechanical resonance. Damping control uses a feedback loop to artificially increase the damping ratio of a system. With an integral controller, an increase

in ζ allows a proportional increase in the feedback gain and closed-loop bandwidth. Although damping controllers alone cannot increase the closed-loop bandwidth to beyond the resonance frequency, they have the advantage of being insensitive to variations in the resonance frequency. In addition, damping controllers suppress, rather than invert, the mechanical resonance so they can provide better rejection of external disturbances than inversion-based systems.

A number of techniques for damping control have been demonstrated successfully in the literature, these include positive position feedback (PPF) (Fanson and Caughey 1990), polynomial-based control (Aphale et al. 2008), shunt control (Fleming and Moheimani 2006; Fleming et al. 2002), resonant control (Sebastian et al. 2008), and integral resonance control (IRC) (Aphale et al. 2007; Bhikkaji and Moheimani 2008).

In Aphale et al. (2007), IRC was demonstrated as a simple means for damping multiple resonance modes of a cantilever beam. The IRC scheme employs a constant feedthrough term and a simple first-order controller to achieve substantial damping of multiple resonance modes. An adaption of this controller that is suitable for tracking control was reported in (Fleming et al. 2010). The regulator form of IRC is a first-order low-pass filter, which is straightforward to implement. A major benefit of the regulator form is that it can be enclosed in a simple tracking control loop to eliminate drift and effectively reduce nonlinearity at low frequencies.

Optimal controllers with automatic synthesis have also been successfully applied to nanopositioning applications. Examples include robust $\mathcal{H}_\infty$ controllers (Salapaka et al. 2002; Sebastian and Salapaka 2005) and LMI-based controllers (Lee and Salapaka 2009). Robust controllers have also been incorporated with approximate models of hysteresis to improve performance (Chen 1992).

Other control techniques include methods that are targeted at particular trajectories, such as triangular scanning signals (Eielsen et al. 2011). Such periodic reference trajectories often arise in nanopositioning applications (Kenton and Leang 2012). A commonly used technique for controlling systems with periodic inputs or disturbances is RC, as discussed in Sect. 7.10. Another technique that can be used to improve the reference tracking performance of a feedback system is feedforward control (Wu and Zou 2009; Leang and Devasia 2007), which is discussed in Chap. 9.

In the following, an experimental nanopositioner is described for the purpose of examining the performance of three practical controllers. In Sect. 7.3, the performance limitations of basic integral control are discussed. This is followed by a description of inverse control and damping control in Sects. 7.4 and 7.5. In Sect. 7.6, the bandwidth, settling time, and robustness of the three controllers are compared. Each controller is designed to maximize bandwidth while retaining stability margins of at least 6 dB and 60°.

Scanning probe microscopy is an application that requires high-performance control of the sample and probe nanopositioner. The performance implications of each control strategy are demonstrated by applying each technique to an atomic force microscope in Sect. 7.9.

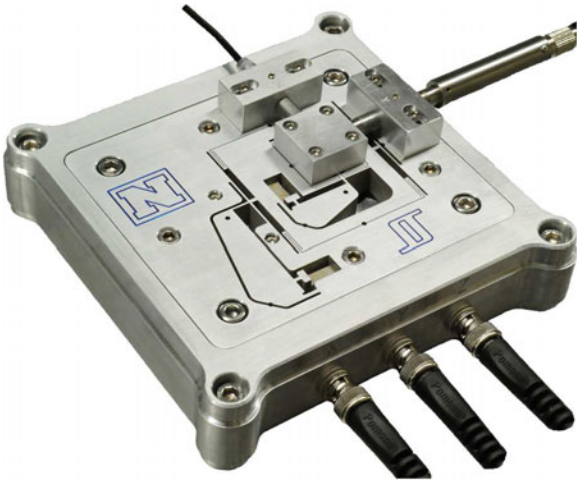


Fig. 7.2 A two-axis serial kinematic nanopositioning platform with a range of 30 μm

7.2 Experimental Setup

To compare the controller characteristics, each technique will be applied to the XY lateral nanopositioning stage pictured in Fig. 7.2. Each axis contains a 12 mm piezoelectric stack actuator (Noliac NAC2003-H12) with a free displacement of 12 μm at 200 V. The flexure design includes a mechanical amplifier to provide a total range of 30 μm. The flexures also mitigate cross-coupling between the axes so that each axis can be controlled independently. The position of the moving platform is measured by a Microsense 6810 capacitive sensor and 6504-01 probe, which has a sensitivity of 2.5 μm/V. The stage is driven by two PiezoDrive PDL200 voltage amplifiers with a gain of 20.

The x-axis, which translates from left to right in Fig. 7.2, has a resonance frequency of 513 Hz. The y-axis contains less mass so the resonance frequency is higher, 727 Hz. Since the x-axis imposes a greater limitation on performance, the comparison will be performed on this axis. However, the design process for the other axis is identical.

The frequency response for a nominal load is plotted in Fig. 7.3a. With the maximum payload, the resonance frequency reduces to 415 Hz as shown in Fig. 7.3b. It can be observed that payload mass significantly modifies the higher frequency dynamics.

For the purpose of control design, a second-order model is procured using the frequency domain least-squares techniques. The model parameters are:

$$G(s) = \frac{2.025 \times 10^7}{s^2 + 48.63s + 1.042 \times 10^7}. \tag{7.2}$$

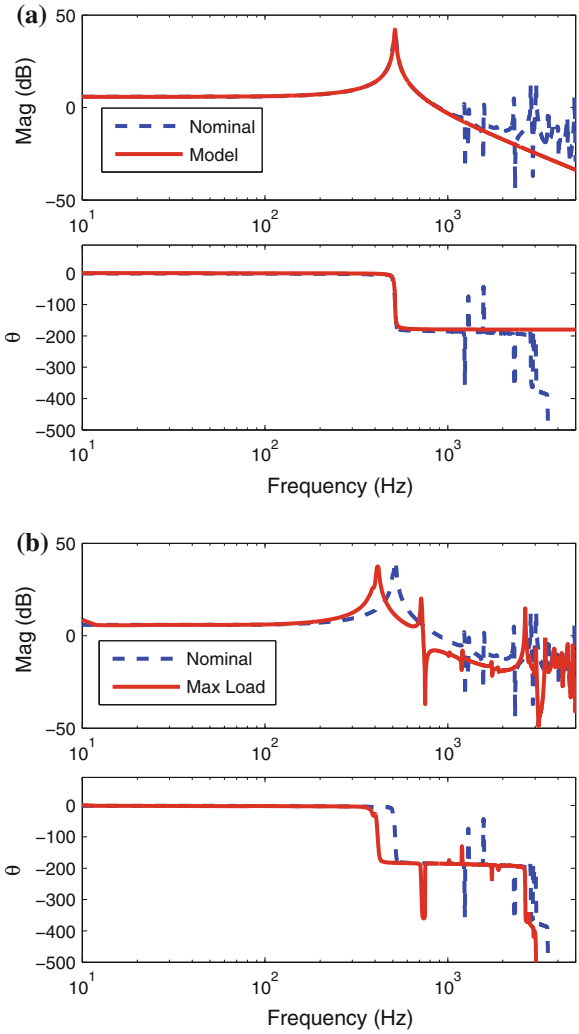


Fig. 7.3 The open-loop frequency response measured from the voltage amplifier input to the sensor output, scaled to μm/V. In **a** the nominal response is compared to the identified model. In **b** the frequency response of the system with maximum load is compared to the nominal response

The frequency response of the model is compared to the experimental data in Fig. 7.3a. The model closely approximates the first resonance mode, which is sufficient for control design.

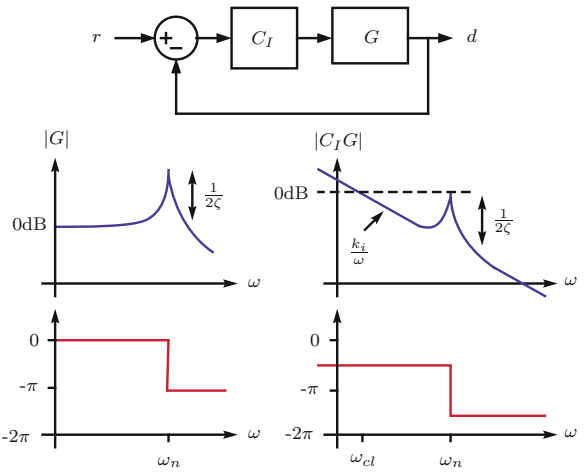


Fig. 7.4 A nanopositioning system G controlled by an integral controller $C_I = k_i/s$. The frequency response of G and the system loop-gain $C_I G$ are plotted on the *left* and *right hand side* respectively

7.3 PI Control

A popular technique for control of commercial nanopositioning systems is sensor-based feedback using integral or proportional-integral control (Li et al. 2006). The transfer function of a PID controller is

$$C_{\text{PID}}(s) = k_p + k_i/s + k_d s, \tag{7.3}$$

However, the derivative term is rarely used due to the increased noise sensitivity and stability problems associated with high frequency resonance modes. PI controllers are simple to tune and effectively reduce piezoelectric nonlinearity at low frequencies. However, the bandwidth of PI tracking controllers is severely limited by the presence of highly resonant modes. The limited closed-loop bandwidth can be explained by examining the loop gain $C_I G$ in Fig. 7.4. Here, the resonant system G is controlled by an integral controller C_I with gain k_i. The factor limiting the maximum feedback gain and closed-loop bandwidth is gain-margin.

Above the natural frequency ω_n, which is approximately equal to the resonance frequency in systems with low damping, the phase lag of the loop-gain exceeds π so the magnitude must be less than 1 (0 dB) for stability in closed-loop. The condition for closed-loop stability is approximately

$$\frac{k_i}{\omega_n} \times \frac{1}{2\zeta} < 1, \text{ or } k_i < 2\omega_n \zeta. \tag{7.4}$$

As the system G is unity gain, the complementary sensitivity function is

$$\frac{d(s)}{r(s)} = \frac{C_I(s)G(s)}{C_{PI}(s)G(s)+1} \approx \frac{k_i}{s+k_i}. \tag{7.5}$$

Thus, the feedback gain k_i is also the approximate 3-dB bandwidth of the complementary sensitivity function and the 0-dB crossing of the loop-gain (in radians per second). From this fact, and the stability condition (7.4), the maximum closed-loop bandwidth is equal to twice the product of damping ratio ζ and natural frequency ω_n, i.e.,

$$\text{max. closed-loop bandwidth} < 2\omega_n\zeta. \tag{7.6}$$

This is a severe limitation as the damping ratio is usually on the order of 0.01, so the maximum closed-loop bandwidth is less than 2 % of the resonance frequency. If a certain amount of gain-margin is required, the bandwidth further reduces to:

$$\text{max. closed-loop bandwidth} < \frac{2\omega_n\zeta}{\text{gain-margin}}, \tag{7.7}$$

where the gain margin is specified as a linear magnitude rather than in dB, for example, 2 rather than 6 dB. The maximum closed-loop bandwidth can also be estimated directly from the frequency response by replacing the factor 2ζ with $1/P$, where P is the linear magnitude of the resonance peak divided by the DC gain, that is

$$\text{max. closed-loop bandwidth} < \frac{\omega_n}{P \times \text{gain-margin}}, \tag{7.8}$$

Due to the second-order resonance, adding a first-order zero to the loop-gain with a proportional term offers little improvement. A derivative term can be beneficial, however this is rarely used as it can destabilize higher frequency modes. A better alternative to derivative action is the notch filter or damping controller discussed in the following sections.

For the nanopositioner under consideration, an integral gain of 15.5 results in a gain-margin of 6 dB and a bandwidth of 13 Hz. The performance is compared to the inversion and damping controllers in Sect. 7.6.

7.4 PI Control with Notch Filters

Techniques aimed at improving the closed-loop bandwidth are typically based on either inversion of resonant dynamics using a notch filter (Abramovitch et al. 2008; Leang and Devasia 2007) or the use of a damping controller (Fleming et al. 2010; Aphale et al. 2008). Inversion techniques are popular as they are simple to implement and can provide a high closed-loop bandwidth if they are finely tuned and the resonance frequency does not vary (Abramovitch et al. 2008). The transfer function of a typical inverse controller is

$$C_{\text{Notch}}(s) = \left(k_p + \frac{k_i}{s}\right) \frac{s^2 + 2\omega_z\zeta_z s + \omega_z^2}{\omega_z^2} \tag{7.9}$$

where ζ_z and ω_z are approximately the damping ratio and first resonance frequency of the nanopositioner. Depending on the implementation method, an additional pole may be required above the bandwidth of interest in order to ensure causality.

The direct inversion controller (7.9) may not be suitable when significant higher-frequency resonances exist. In this case, a notch filter is more appropriate since it attempts to replace the lightly damped resonance with a pair of real poles. Other denominator posibilities include, for example, a pair of complex poles with critical damping. The transfer function is

$$C_{\text{Notch}}(s) = \left(k_p + \frac{k_i}{s}\right) \frac{s^2 + 2\omega_z\zeta_z s + \omega_z^2}{(s + \omega_z)^2} \tag{7.10}$$

If an inverse controller is precisely tuned to the first mechanical resonance, the presence of this mode can be essentially eliminated from the loop-gain. The maximum bandwidth is now limited by the second system resonance rather than the first. Equations (7.7) or (7.8) predict the maximum closed-loop bandwidth based on the resonance frequency and damping ratio of the second significant resonance mode. Additional notch filters can be used to invert higher order resonances, however this requires an extremely accurate system model.

A major consideration with inversion-based control is the possibility for modeling error. In particular, if the resonance frequency drops below the frequency of the notch filter, the phase lag will cause instability. Therefore, a notch filter must be tuned to the lowest resonance frequency that will occur during service. For example, the nanopositioner under consideration has a nominal resonance frequency of 513 Hz and a minimum resonance frequency 410 Hz. Thus, the notch filter is tuned to 410 Hz with an estimated damping of $\zeta_z = 0.01$.

To maintain a gain-margin of 6 dB, the maximum integral gain is $k_i = 44$. The loop-gain during nominal and maximum load conditions is plotted in Fig. 7.5. During nominal conditions, the phase-lag does not exceed 180° until the second resonance mode; however, the first resonance mode remains dominant in the response and can be excited by high-frequency components of the input or disturbances. This behavior is evident in the closed-loop frequency and step responses plotted in Sect. 7.6. Since the notch filter is tuned to the lowest resonance frequency, the system actually performs better with the maximum payload. The loop-gain in Fig. 7.5 shows that the first resonance-mode is almost inverted during this condition.

Due to the sensitivity of inversion-based controllers to variations in the resonance frequency, they are most suited to applications where the resonance frequency is stable, or where the feedback controller can be continually recalibrated (Abramovitch et al. 2008).

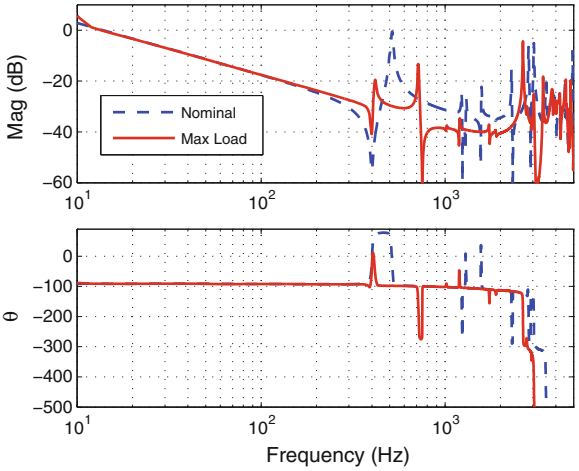


Fig. 7.5 The loop-gain of the nanopositioner and inversion based controller during nominal and maximum load ($C_{\text{notch}}(s)G(s)$)

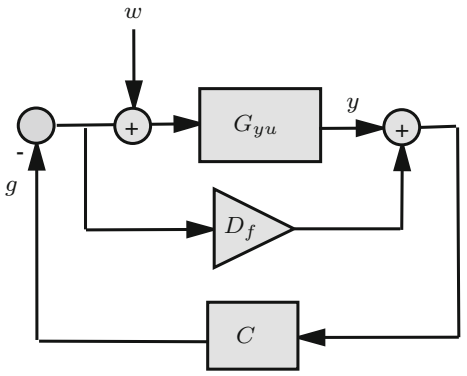


Fig. 7.6 Integral resonance control scheme (Aphale et al. 2007)

7.5 PI Control with IRC Damping

Integral Resonance Control (IRC) was introduced in 2007 as a means for augmenting the structural damping of resonant systems with collocated sensors and actuators (Aphale et al. 2007). A diagram of an IRC loop is shown in Fig. 7.6. It consists of the collocated system G_{yu}, an artificial feedthrough D_f and a controller C. The input disturbance w represents environmental disturbances but can also be used to obtain some qualitative information about the closed-loop response to piezoelectric nonlinearity. That is, if the disturbance rejection at the scan frequency and first few harmonics is large, a significant reduction in hysteresis could be expected.

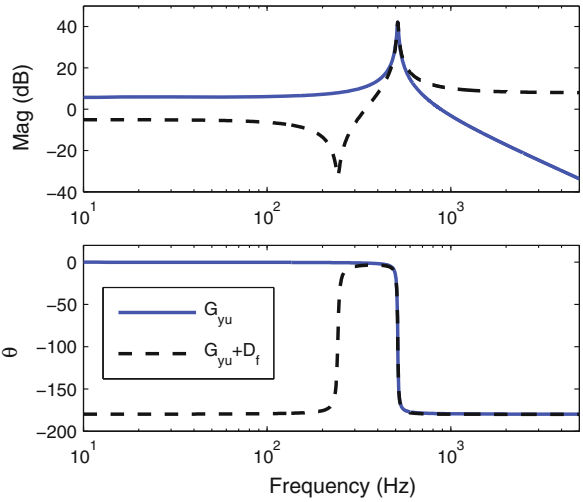


Fig. 7.7 Frequency response of the open-loop system G_{yu} and with artificial feedthrough $G_{yu}+D_f$, where $D_f = -2.5$. The 180° phase change of $G_{yu} + D_f$ is due to the negative feedthrough which also makes the system inverting

The first step in designing an IRC controller is to select, and add, an artificial feedthrough term D_f to the original plant G_{yu}. The new system is referred to as $G_{yu}+D_f$. It has been shown that a sufficiently large and negative feedthrough term will introduce a pair of zeros below the first resonance mode and also guarantee zero-pole interlacing for higher frequency modes (Aphale et al. 2007). Smaller feedthrough terms permit greater maximum damping. Although it is straightforward to manually select a suitable feedthrough term, it can also be computed from Theorem 2 in (Aphale et al. 2007).

For the model G_{yu} described in (7.2), a feedthrough term of $D_f = -2.5$ is sufficient to introduce a pair of zeros below the first resonance mode. The frequency responses of the open-loop system G_{yu} and the modified transfer function $G_{yu} + D_f$ are plotted in Fig. 7.7.

The key behind IRC is the phase response of $G_{yu} + D_f$, which now lies between between $-180°$ and $0°$ as shown in Fig. 7.7. Due to the bounded phase of $G_{yu} + D_f$ a simple negative integral controller

$$C = \frac{-k}{s}, \tag{7.11}$$

can be applied directly to the system. To examine the stability of such a controller, we consider the loop-gain $C \times \left(G_{yu} + D_f\right)$. For stability, the phase of the loop-gain must be within $\pm 180°$ while the gain is greater than zero. The phase of the loop-gain $C \times \left(G_{yu} + D_f\right)$ is equal to the phase of $G_{yu}+D_f - 180°$ for the negative controller gain and a further $90°$ for the single controller pole. The resulting phase response

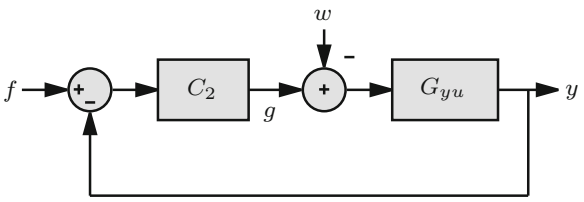


Fig. 7.8 The integral resonance controller of Fig. 7.6 rearranged in regulator form

of the loop-gain lies between $+90°$ and $-90°$. That is, regardless of controller gain, the closed-loop system has a phase margin of $90°$ and an infinite gain-margin with respect to $G_{yu} + D_f$.

A suitable controller gain k can be selected to maximize damping using the root-locus technique (Aphale et al. 2007). For the system under consideration, a gain of $k = 1{,}900$ results in a maximum damping ratio 0.57.

In order to facilitate a tracking control loop, the feedback diagram must be rearranged in a form where the input does not appear as a disturbance. This can be achieved by finding an equivalent regulator that provides the same loop gain, as shown in Fig. 7.8. In Fig. 7.6, the control input g is related to the measured output y by

$$g = C(y - D_f g), \tag{7.12}$$

thus, the equivalent regulator C_2 is

$$C_2 = \frac{C}{1 + CD_f}. \tag{7.13}$$

When $C = \frac{-k}{s}$ the equivalent regulator is

$$C_2 = \frac{-k}{s - kD_f}. \tag{7.14}$$

The closed-loop transfer function of the damping loop is,

$$G_{yf} = \frac{G_{yu}C_2}{1 + G_{yu}C_2}. \tag{7.15}$$

With $D_f = -2.5$ and $k = 1{,}900$, the frequency responses of the open-loop and damped system are plotted in Fig. 7.9.

To achieve integral tracking action, the IRC loop can be enclosed in an outer loop as shown in Fig. 7.10. From the response in Fig. 7.9 or a pole-zero map, it can be observed that the damped system contains the resonance poles, plus an additional first-order pole mid-way between the resonance frequency and the zeros of $G_{yu}+D_f$. To eliminate the additional pole from the loop-gain, an ideal tracking controller is a

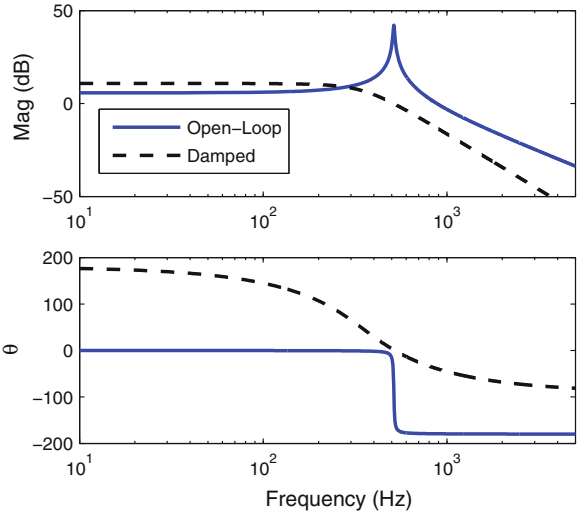


Fig. 7.9 The open- and closed-loop frequency response of the system with integral resonance control

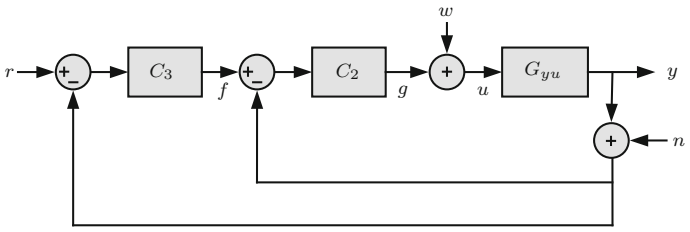


Fig. 7.10 Tracking control system with the damping controller $C_2(s)$ and tracking controller $C_3(s)$. The signal w is the disturbance input and n is the sensor noise

PI controller with a fixed zero at the frequency of the additional pole, that is,

$$C_3 = \frac{-k_i(s+\omega_z)}{s\omega_z}. \tag{7.16}$$

where k_i is chosen in the normal way to provide the desired stability margins or bandwidth. Note that C_3 is inverting to cancel the inverting nature of G_{yf}. For the system under comparison, a gain of $k_i = 245$ results in a phase margin of 60°. The closed-loop response performance is examined in Sect. 7.6.

The transfer function of the closed-loop system is

$$\frac{y}{r} = \frac{C_3 G_{yf}}{1 + C_3 G_{yf}}, \tag{7.17}$$

Table 7.1 Summary of controller parameters

	PI	PI + Notch	PI + IRC
Tracking TF	$\frac{15.5}{s}$	$\frac{44}{s}\frac{2\pi 10^3}{s+2\pi 10^3}$	$\frac{-245}{s}\frac{s+2720}{2720}$
Inverse or damping TF	–	$\frac{s^2+50.27s+6.317\times 10^6}{6.317\times 10^6}$	$\frac{-1900}{s+4750}$

or alternatively,

$$\frac{y}{r} = \frac{C_2 C_3 G_{yu}}{1 + C_2(1 + C_3)G_{yu}}. \tag{7.18}$$

In addition to the closed-loop response, the transfer function from disturbance to the regulated variable y is also of importance,

$$\frac{y}{w} = \frac{G_{yu}}{1 + C_2(1 + C_3)G_{yu}}. \tag{7.19}$$

That is, the disturbance input is regulated by the equivalent controller $C_2(1 + C_3)$.

7.6 Performance Comparison

In Sects. 7.3–7.5, three controllers were designed to maintain a gain and phase margin of at least 6 dB and 60°. The controller parameters are summarized in Table 7.1, and the simulated stability margins are listed in Table 7.2. The integral and inverse controller were limited by gain-margin while the damping controller was limited by phase margin.

The simulated and experimental closed-loop frequency responses are plotted in Figs. 7.11 and 7.12. The frequency where the phase-lag of each control loop exceeds 45° is compared in Table 7.2. In nanopositioning applications, the 45° bandwidth is more informative than the 3 dB bandwidth since it is more closely related to the settling time. Due to the higher permissible servo gain, the PI + IRC controller provides the highest bandwidth by a significant margin.

The simulated and experimental step responses are plotted in Figs. 7.13 and 7.14 and summarized in Table 7.2. The PI+IRC controller provides the shortest step response by approximately a factor of 5, however the response exhibits some overshoot.

Out of the three controllers, the combination of PI control and IRC provides the best closed-loop performance under both nominal and full-load conditions. This is the key benefit of damping control, it is more robust to changes in resonance frequency than inverse control. If the variation in resonance frequency were less, or if the resonance frequency was stable, there would not be a significant difference between the dynamic performance of an inverse controller and damping controller.

Table 7.2 Closed-loop performance summary of integral, inversion based, and damping controller

	PI	PI + Notch	PI + IRC
Gain margin			
Nominal load (dB)	6.1	6.0	14
Full load (dB)	7.0	5.1	10
Phase margin			
Nominal load	inf	89°	69°
Full load (°)	90	89	69
Nominal bandwidth (45°)			
Simulated (Hz)	4.8	13	74
Experimental (Hz)	5.0	13	50
Full-load bandwidth (45°)			
Simulated (Hz)	4.8	13	77
Experimental (Hz)	5.0	13	78
Nominal settling time (99 %)			
Simulated (ms)	160	54	6.2
Experimental (ms)	164	48	9.7
Full-load settling time (99 %)			
Simulated (ms)	170	53	11
Experimental (ms)	165	42	7.6

Since the damping controller requires more design effort than an inverse controller, it is sensible to choose this option when some variation in the resonance frequencies are expected, or if there are multiple low-frequency resonances that are difficult to model.

7.7 Noise and Resolution

The noise sensitivity of each control strategy is the transfer function from the sensor noise n to the actual position y. For the sake of comparison, the noise contribution of the voltage amplifier is assumed to be small compared to the sensor noise. As discussed in Chap. 13, the RMS value or standard deviation of the sensor-induced noise is equal to

$$\sigma = \sqrt{\int_0^\infty S_n(f) \left| \frac{y(2\pi jf)}{n(2\pi jf)} \right|^2 df}, \tag{7.20}$$

where $S_n(f)$ is the power spectral density of the sensor noise. If the sensor noise is Gaussian distributed, the resolution is equal to 6σ. Therefore, if the sensor noise spectral density is constant, the closed-loop resolution is proportional to the area under the noise sensitivity transfer function.

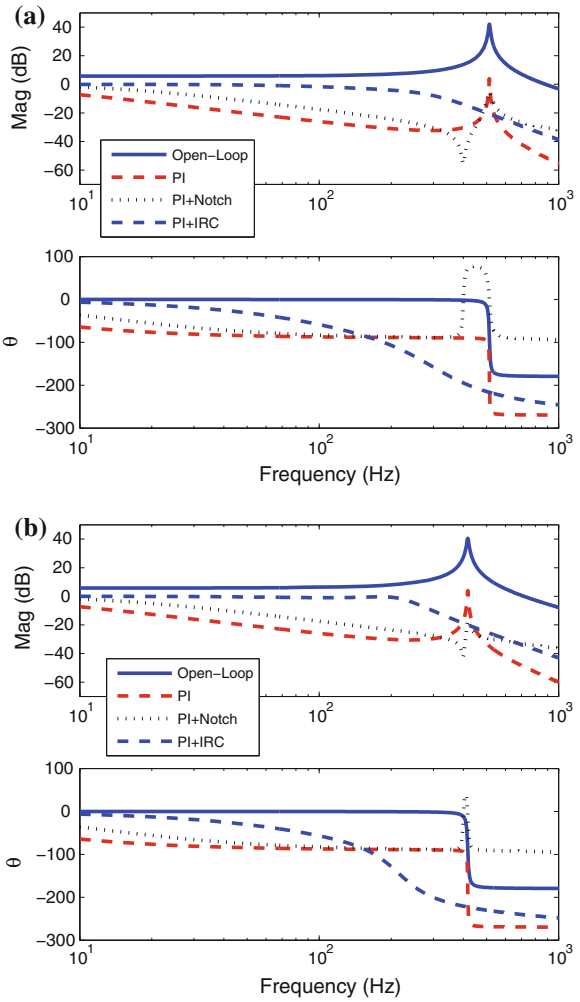


Fig. 7.11 The simulated closed-loop frequency response of each controller under nominal and maximum load conditions. **a** Nominal load, $f_r = 513\,\text{Hz}$. **b** Maximum load, $f_r = 415\,\text{Hz}$

For the PI and inverse controller, the noise sensitivity is the complementary sensitivity function with opposite sign, that is

$$\frac{y}{n} = \frac{-C_3 G_{yu}}{1 + C_3 G_{yu}}. \tag{7.21}$$

However, with a damping controller as shown in Fig. 7.10, the noise sensitivity is not identical to the complementary sensitivity (7.17). Rather, it is

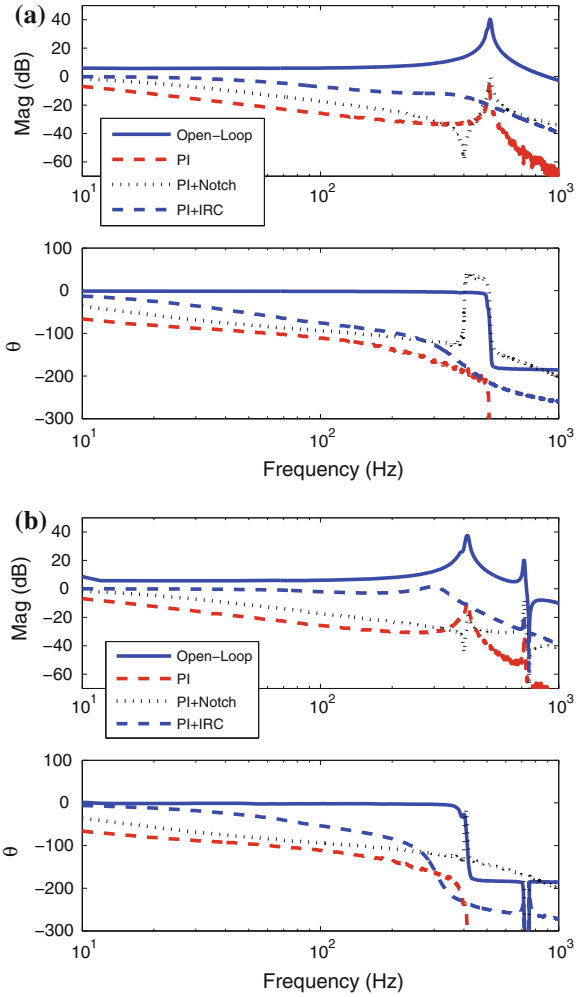


Fig. 7.12 The experimental closed-loop frequency response of each controller under nominal and maximum load conditions. **a** Nominal load, $f_r = 513\,\text{Hz}$. **b** Maximum load, $f_r = 415\,\text{Hz}$

$$\frac{y}{n} = \frac{-C_2(1+C_3)G_{yu}}{1+C_2(1+C_3)G_{yu}}. \tag{7.22}$$

It can be observed from Eq. (7.21) that the noise sensitivity for a standard control loop can be reduced by reducing the closed-loop bandwidth or controller gain. However with a damping controller, the noise sensitivity bandwidth is dominated by the damping control loop, not the tracking loop. This is a drawback since the noise sensitivity bandwidth cannot be reduced by varying the tracking controller gain. However, since the noise sensitivity of the IRC system is not strongly affected by the tracking

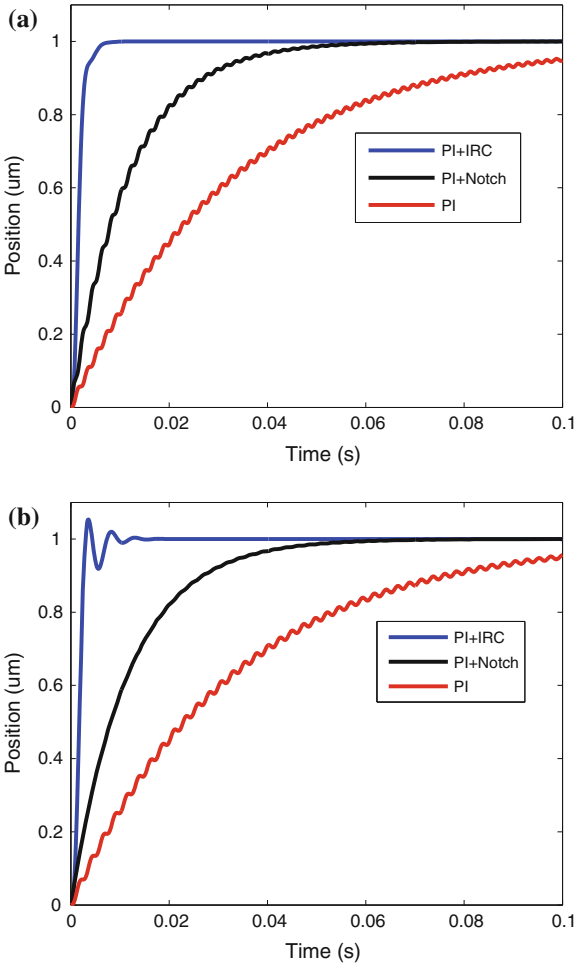


Fig. 7.13 The simulated closed-loop step response of each controller under nominal and maximum load conditions. **a** Nominal load, $f_r = 513\,\text{Hz}$. **b** Maximum load, $f_r = 415\,\text{Hz}$

controller gain C_3, the tracking controller can be tuned to the highest practical gain since there is little noise penalty in doing so.

The noise sensitivity of each control strategy is plotted in Fig. 7.15. Due to the wide bandwidth of the damping controller, the noise sensitivity bandwidth is significantly greater than the PI and inverse controllers.

A straightforward technique for estimating the positioning resolution is to measure the sensor noise and filter it by the noise sensitivity function. Following the guidelines in Sect. 13.9.3, the sensor noise was amplified using an SR560 amplifier with a gain of 10,000 and a bandwidth of 0.03–10 kHz. A 100 s of data was recorded at a sampling rate of 30 kHz. A 3 s record of the closed-loop position noise for each controller

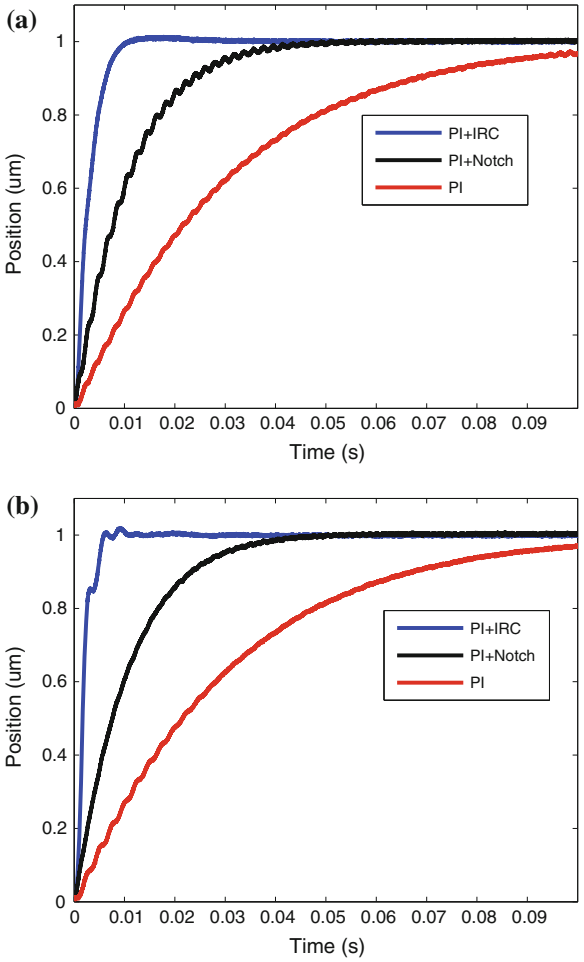


Fig. 7.14 The experimental closed-loop step response of each controller under nominal and maximum load conditions. **a** Nominal load, $f_r = 513\,\text{Hz}$. **b** Maximum load, $f_r = 415\,\text{Hz}$

is plotted in Fig. 7.16. While the PI and inverse controller contain low-frequency noise plus randomly excited resonance, the IRC controller results in a more uniform spectrum but with a wider noise bandwidth. Considering that the IRC controller increases the closed-loop bandwidth from 5 to 78 Hz (compared to PI control), the decrease in resolution from 0.27 to 0.43 nm is small.

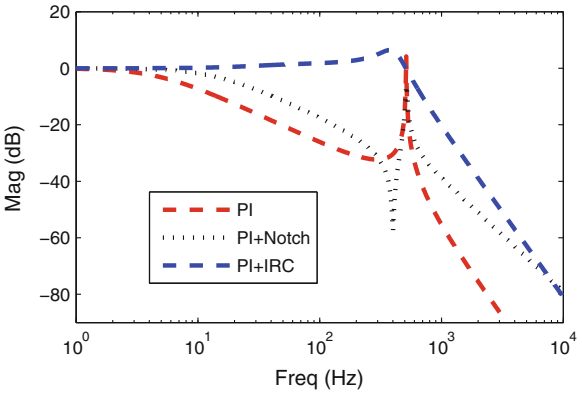


Fig. 7.15 The noise sensitivity of each control strategy

7.8 Analog Implementation

Due to the simplicity of the IRC damping and tracking controller, it is straightforward to implement in both analog and digital form. The IRC damping and tracking controller shown in Fig. 7.10 can be implemented directly with the analog circuit shown in Fig. 7.17. Although the controller requires only two opamps, the four-opamp circuit shown in Fig. 7.17 is easier to understand, trouble-shoot, and tune (if necessary).

The operation of the circuit is self-explanatory. The first stage is a unity-gain differential amplifier that implements the subtraction function $r - y$. The second stage implements the PI tracking controller. The corresponding circuit transfer function of the PI controller is

$$C_3(s) = -\frac{s + \frac{1}{r_{3b}c_3}}{r_{3a}c_3 \frac{1}{r_{3b}c_3} s}, \tag{7.23}$$

which results in the equality $r_{3a}c_3 = 1/k_i$ and $r_{3b}c_3 = 1/\omega_z$

The third stage is a unity-gain differential amplifier with two noninverting inputs for f and u_f. The final stage implements the IRC controller C_2, where

$$C_2(s) = \frac{-k}{s - kD_f}. \tag{7.24}$$

The circuit transfer function is

$$C_2(s) = \frac{-\frac{1}{r_{2a}c_2}}{s + \frac{1}{r_{2b}c_2}}. \tag{7.25}$$

As k is positive and D_f is negative, the equalities are

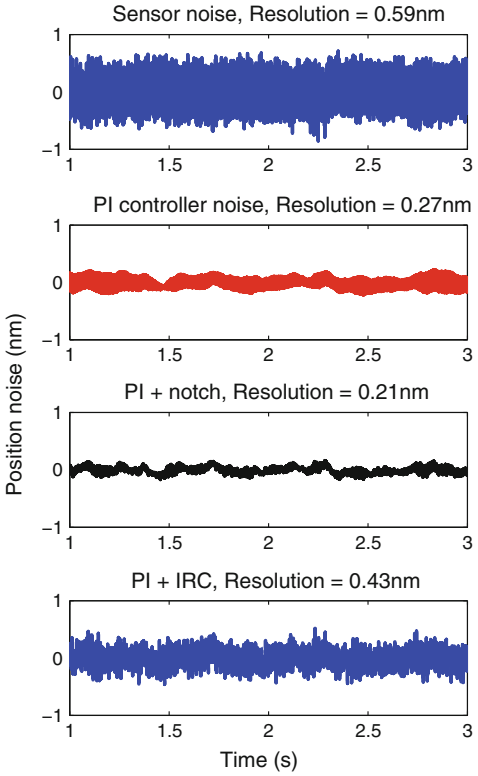


Fig. 7.16 The closed-loop noise of each control strategy and the corresponding 6σ-resolution

$$r_{2a}c_2 = \frac{1}{k}, \text{ and } r_{2b}c_2 = \frac{1}{kD_f}. \tag{7.26}$$

In both of the integrating stages, a 100 nF film capacitor (e.g., Polypropylene) is recommended as these capacitors are highly linear and temperature stable. The capacitance value should not be less than 100 nF to avoid large resistances that contribute thermal noise and amplify current noise. The opamps should have a gain-bandwidth product of around 10 Mhz or greater to avoid controller phase lag. The opamps should also be suited to a source impedance in the kΩ range with the lowest possible noise corner frequency. The Texas Instruments OPA4227 is a suitable device, which is readily available at low cost.

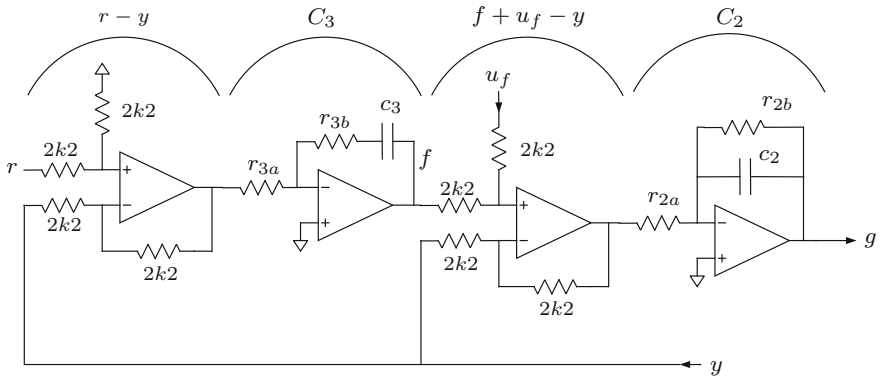


Fig. 7.17 Analog implementation of an IRC damping and PI tracking controller

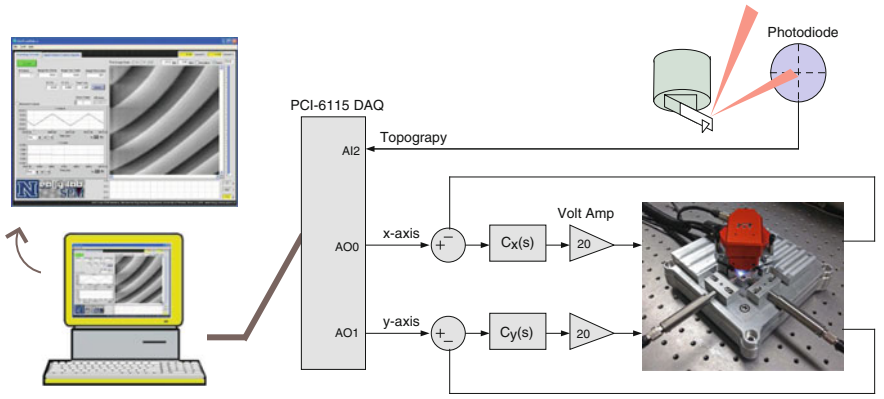


Fig. 7.18 The AFM system in constant-height contact-mode. A lab view application creates the scanning waveforms and records the image while an xPC Target performs the real-time control

7.9 Application to AFM Imaging

To illustrate the impact of positioning bandwidth on application performance, the nanopositioner will be employed for lateral scanning in an atomic force microscope (Abramovitch et al. 2007; Salapaka and Salapaka 2008; Ando et al. 2008; Schitter 2009; Clayton et al. 2009; Fleming et al. 2010).

The experimental setup is shown in Fig. 7.18. A National Instruments PCI-6115 data acquisition card and LabView application[1] are used to generate the raster signals and acquire the image (Fleming et al. 2010). The AFM head is a NanoSurf EasyScan microscope which is only used for holding the cantilever and measuring the deflection. The microcantilver is a Budget Sensors ContAl cantilever with a stiffness of

[1] The easyLab SPM Interface is available by contacting K. K. Leang at kam@unr.edu.

0.2 N/m and the sample under consideration is a silicon calibration grating with a period of 6 μm and a height of 20 nm.

The scan waveforms are standard triangular raster signals. To acquire an image, the y-axis is driven with a slow ramp while the x-axis reference is a 10 Hz triangular waveform. With a scan rate of 10 Hz, a 200 × 200 pixel image is acquired in 20 s. Due to the slow scan rate of the y-axis, the tracking error can be neglected. However, significant positioning errors can arise from the x-axis response. The positioning error for each controller and the resulting image is plotted in Fig. 7.19. The higher bandwidth of the IRC control system is observed to significantly reduce scan-induced imaging artifacts.

7.10 Repetitive Control

7.10.1 Introduction

Many applications in nanopositioning require the stage to track periodic reference trajectories with precision. For example, in AFM a nanopositioner is used to raster back and forth a probe tip relative to a sample surface to obtain high-resolution topographical images, directly measure various properties of a specimen, and even investigate nano-scale dynamic interactions in real time (Radmacher 1997; Salapaka and Salapaka 2008; Ando et al. 2008). The periodicity of the desired trajectory lends itself nicely for applying RC for precision positioning, even at relatively high speed. Recently, the RC approach has been applied to piezo-based positioners and SPMs (Aridogan et al. 2009; Merry et al. 2011; Shan and Leang 2012a, b, 2013).

Repetitive control (RC) is a direct application of the internal model principle (Francis and Wonham 1976), where a signal generator—the transfer function of the reference trajectory—is incorporated into a feedback loop to provide high gain at the fundamental frequency of the reference trajectory and its harmonics (Inoue et al. 1981; Hara et al. 1988). Repetitive controllers have been used to address run-out issues in disk drive systems (Chew and Tomizuka 1990; Steinbuch et al. 2007) and to improve the performance of machine tools (Li and Li 1996; Chen and Hsieh 2007). Compared to traditional proportional-integral or proportional-integral-derivative (PID) feedback controllers, where careful tuning is required and the residual tracking error persists from one operating cycle to the next, RC has the ability to reduce the error as the number of operating cycles increases. For applications in which the desired trajectory is periodic and the signal period is known a priori, a repetitive controller offers many advantages. First, it can be plugged into an existing feedback control loop to enhance performance for scanning applications. Second, compared to iterative learning control (ILC) (Arimoto et al. 1984; Moore et al. 1992), a control method that has been used extensively for piezo-based positioning systems (Leang and Devasia 2006; Wu and Zou 2007), RC does not require the initial condition to be reset at the start of each iteration trial (Hara et al. 1988). Therefore, the

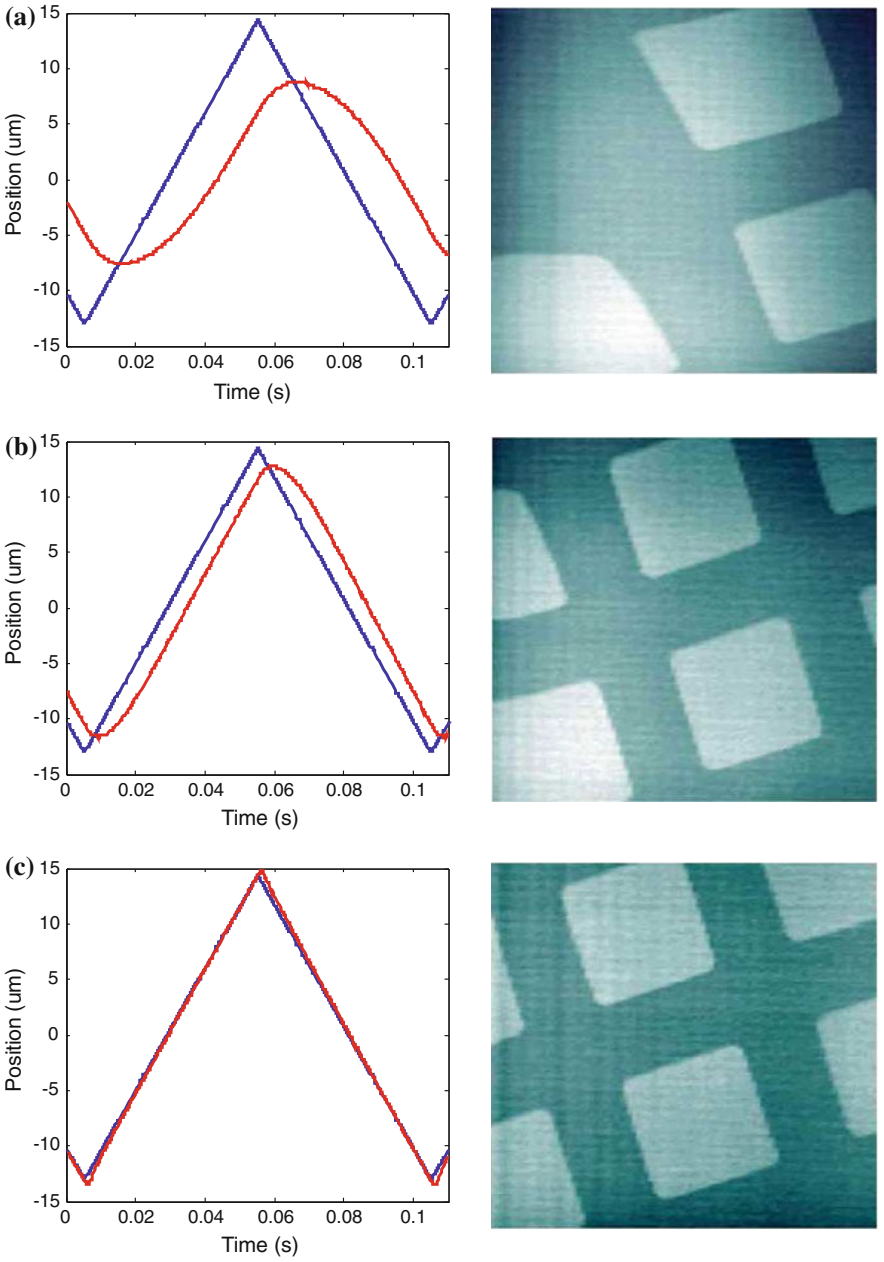


Fig. 7.19 The x-axis scanning performance and resulting image for each of the three controller strategies. The scanning trajectory is a full-range (27 μm) 10 Hz *triangle wave*. **a** PI control, **b** PI + notch, **c** PI + IRC damping

implementation is simplified. Third, compared to model-based feedforward approaches (Clayton et al. 2009; Croft et al. 2001), RC does not require extensive modeling of the system. Due to variations in the system dynamics, for example due to aging (Lowrie et al. 1999) or temperature variations (Lee and Saravanos 1998), open-loop feedforward approaches often lack robustness. On the other hand, the feedback mechanism built into RC provides robustness to parameter variation. Finally, RC can be easily implemented digitally, and thus high-speed data acquisition and control hardware such as field-programmable gate array systems (Fantner et al. 2005) can take advantage of the RC structure for precision control. It has also recently been demonstrated that RC can also be implemented using a single FIR filter which dramatically simplifies the implementation (Teo and Fleming 2014).

7.10.2 Repetitive Control Concept and Stability Considerations

To illustrate the concept of RC, consider an example of the discrete-time RC closed-loop system shown in Fig. 7.20a. The dynamics of the nanopositioner, assumed to be linear, is represented by $G_p(z)$, where $z = e^{j\omega T_s}$, $\omega \in (0, \pi/T_s)$. In the block diagram, $G_c(z)$ is a feedback controller, such as an existing PID controller; $Q(z)$ is a low-pass filter for robustness; k_{rc} is the RC gain; and $P_1(z) = z^{m_1}$ and $P_2(z) = z^{m_2}$, where m_1, m_2 are non-negative integers, are positive phase lead compensators to enhance the performance of the RC feedback system. It is emphasized that the phase lead compensators z^{m_1} and z^{m_2} provide a linear phase lead of (in units of radians)

$$\theta_{1,2}(\omega) = m_{1,2} T_s \omega, \text{ for } \omega \in (0, \pi/T_s). \tag{7.27}$$

The key component of the repetitive controller is the signal generator. To create a signal generator with period T_p, the inner loop contains the pure delay z^{-N}, where the positive integer $N = T_p/T_s$ is the number of points per period T_p; and T_s is the sampling time. An analysis of the performance of the closed-loop system is presented below, where the following assumptions are considered: (1) the reference trajectory $R(z)$ is periodic and has period T_p and (2) the closed-loop system without the RC loop is asymptotically stable, i.e., $1 + G_c(z)G_p(z) = 0$ has no roots outside of the unit circle in the z-plane.

Assumptions 1 and 2 are easily met for many applications in nanopositioning, including SPMs. For example in AFM imaging, the lateral movements of the piezoactuator are periodic, such as a triangle scanning signal. Also, most SPMs are equipped with feedback controllers $G_c(z)$ to control the lateral positioning, which can be tuned to be stable.

The transfer function of the signal generator [or RC block, Fig. 7.20a] that relates $E(z)$ to $A(z)$ is given by

$$\frac{A(z)}{E(z)} = \frac{Q(z)P_1(z)z^{-N}}{1 - Q(z)P_1(z)z^{-N}} = \frac{Q(z)z^{(-N+m_1)}}{1 - Q(z)z^{(-N+m_1)}}. \tag{7.28}$$

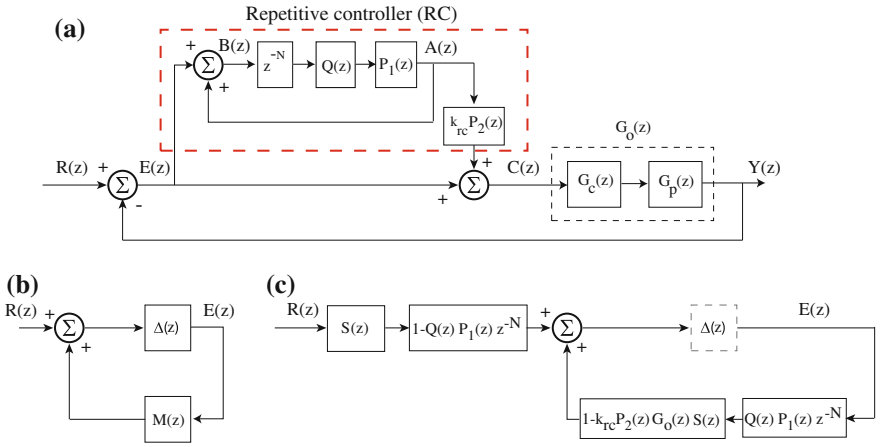


Fig. 7.20 The repetitive control (RC) feedback system. **a** The block diagram of the proposed RC system. **b** Positive feedback system for stability analysis. **c** Positive feedback system representing the block diagram in part **a** for stability analysis

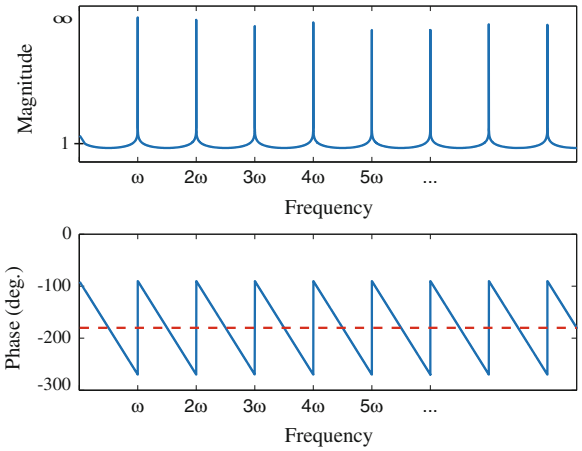


Fig. 7.21 Magnitude and phase versus frequency for signal generator $z^{-N}/(1 - z^{-N})$, where $z = e^{j\omega T_s}$

In the absence of both the low-pass filter $Q(z)$ and positive phase lead $P_1(z) = z^{m_1}$, the poles of the signal generator are $1 - z^{-N} = 0$; therefore, the frequency response of the signal generator shown in Fig. 7.21 reveals infinite gain at the fundamental frequency and its harmonics $\omega = 2n\pi/T_p$, where $n = 1, 2, 3, \ldots$. The infinite gain at the harmonics is what gives the RC its ability to track a periodic reference trajectory. As a result, RC is a useful control method for applications such as SPM in which the scanning motion is repetitive. Unfortunately, the RC also contributes phase lag which causes instability. Therefore, the stability, robustness, and tracking performance of

the RC closed-loop system must be carefully considered. In the following, these issues will be addressed, and the conditions for how to choose the RC gain k_{rc} are presented, along with a discussion of the effects of the phase lead compensators $P_1(z)$ and $P_2(z)$ on the performance of the closed-loop system.

To analyze the stability of the closed-loop RC system shown in Fig. 7.20a, consider the transfer function relating the reference trajectory $R(z)$ and the tracking error $E(z)$,

$$\frac{E(z)}{R(z)} = \frac{1 - H(z)}{1 - H(z) + \big[(k_{\text{rc}} P_2(z) - 1)H(z) + 1\big]G_o(z)}, \tag{7.29}$$

where $H(z) = Q(z)z^{(-N+m_1)}$ and $G_o(z) = G_c(z)G_p(z)$. Multiplying the numerator and denominator of (7.29) by the sensitivity function $S(z) = 1/(1 + G_o(z))$ of the feedback system without the repetitive controller, the following transfer function is obtained:

$$S_{\text{rc}}(z) = \frac{E(z)}{R(z)} = \frac{\big[1 - H(z)\big]S(z)}{1 - H(z)\big[1 - k_{\text{rc}} P_2(z)G_o(z)S(z)\big]}. \tag{7.30}$$

The $S_{\text{rc}}(z)$ shown above is referred to as the sensitivity function of the closed-loop RC system.

The stability conditions for the RC system can be determined by simplifying the block diagram in Fig. 7.20a to the equivalent interconnected system shown in Fig. 7.20b, which results in Fig. 7.20c. Then the RC sensitivity transfer function (7.30) can be associated with the $M(z)$ and $\Delta(z)$ terms in Fig. 7.20c for stability analysis.

Because the closed-loop system without the RC loop is assumed to be asymptotically stable, then the sensitivity function without RC, $S(z)$, has no poles outside the unit circle in the z-plane, so it is stable. Likewise, $1 - H(z)$ is required to be bounded input - bounded output stable. Replacing $z = e^{j\omega T_s}$, the positive feedback closed-loop system in Fig. 7.20c is internally stable according to The Small Gain Theorem (Zhou and Doyle 1998) when

$$\begin{aligned} &\big|H(z)\big[1 - k_{\text{rc}} P_2(z)G_o(z)S(z)\big]\big| = \\ &\quad \Big|H(e^{j\omega T_s})\big[1 - k_{\text{rc}} e^{j\theta_2(\omega)} G_o(e^{j\omega T_s})S(e^{j\omega T_s})\big]\Big| < 1, \end{aligned} \tag{7.31}$$

for all $\omega \in (0, \frac{\pi}{T_s})$, where the phase lead $\theta_2(\omega)$ is defined by Eq. (7.27). By satisfying condition (7.31), the closed-loop RC system shown in Fig. 7.20a is asymptotically stable.

In general, both the RC gain k_{rc} and the phase lead $\theta_2(\omega)$ affect the stability and robustness of RC as well as the rate of convergence of the tracking error. In the following, condition (7.31) is used to determine explicitly the range of acceptable k_{rc} for a given $Q(z)$ and $G_o(z)$. The effects of the phase lead $\theta_2(\omega)$ on robustness and the phase lead $\theta_1(\omega)$ on the tracking performance will be discussed next.

Let $T(z)$ represent the complementary sensitivity function of the closed-loop feedback system without RC, that is, $T(z) = G_o(z)S(z)$. Suppose the magnitude of the low-pass filter $|Q(z)|$ approaches unity at low frequencies and zero at high frequencies, hence $|Q(e^{j\omega T_s})| \leq 1$, for $\omega \in (0, \pi/T_s)$. Therefore, condition (7.31) becomes

$$\left|1 - k_{\text{rc}} e^{j\theta_2(\omega)} T(e^{j\omega T_s})\right| < 1 \leq \frac{1}{|Q(e^{j\omega T_s})|}. \tag{7.32}$$

Replacing the complementary sensitive function with $T(e^{j\omega T_s}) = A(\omega)e^{j\theta_T(\omega)}$, where $A(\omega) > 0$ and $\theta_T(\omega)$ are the magnitude and phase of $T(e^{j\omega T_s})$, respectively, Eq. (7.32) becomes

$$\left|1 - k_{\text{rc}} A(\omega) e^{j[\theta_T(\omega)+\theta_2(\omega)]}\right| < 1. \tag{7.33}$$

Finally, solving Eq. (7.33) leads to the following two conditions for the RC gain k_{rc} and linear phase lead $\theta_2(\omega)$ to ensure stability:

$$0 < k_{\text{rc}} < \frac{2\cos[\theta_T(\omega) + \theta_2(\omega)]}{A(\omega)} \quad \text{and} \tag{7.34}$$

$$-\pi/2 < [\theta_T(\omega) + \theta_2(\omega)] < \pi/2. \tag{7.35}$$

By Eq. (7.35), the lead compensator $P_2(z) = z^{m_2}$ accounts for the phase lag of the closed-loop feedback system without RC. In fact, $P_2(z)$ enhances the stability margin of the closed-loop RC system by increasing the frequency at which the phase angle crosses the $\pm 90°$ boundary. This frequency will be referred to as the crossover frequency.

7.10.3 Dual-Stage Repetitive Control

The challenges with designing and implementing RC include stability, robustness, and achieving good steady-state tracking performance. One solution to the stability and robustness problem is to incorporate a low-pass filter into the RC loop (Tomizuka et al. 1998) or employ a simple frequency aliasing filter (Ratcliffe 2005). It is pointed out that a tradeoff is made between robustness and high-frequency tracking when such filters are used. The steady-state tracking performance of RC can be improved as shown above, for example by cascading a phase-lead compensator to account for the phase lag of the low-pass filter to increase the controller gain at the harmonics of the reference trajectory (Broberg and Molyet 1994; Aridogan et al. 2009). High-order RC has been studied in (Steinbuch et al. 2007) to improve performance and robustness in the presence of noise and variations in the signal period.

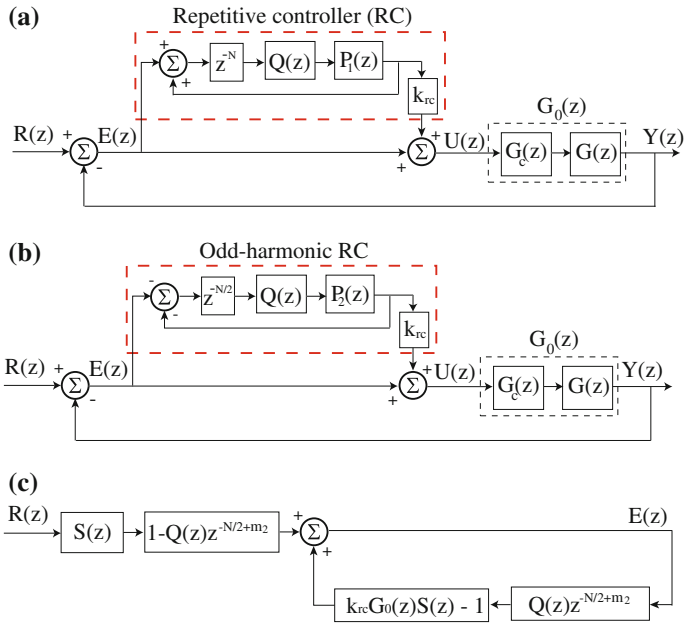


Fig. 7.22 **a** A conventional RC system, where the $R(z)$ represents a periodic reference trajectory, $Y(z)$ is the system output, $G_c(z)$ is the controller and $G(z)$ is the plant dynamics. **b** An odd-harmonic RC with a linear phase-lead compensator $P_2(z) = z^{m_2}$ and a RC gain k_{rc} to enhance performance. **c** An equivalent block diagram of (**b**) for stability analysis

Dual-stage repetitive control (dual-RC) can be used to further improve tracking performance. The dual-RC design is motivated by the need to further reduce the magnitude of the sensitivity function of the closed-loop system to help lower the tracking error. This is achieved by cascading a conventional RC with an odd-harmonic RC (Zhou et al. 2007; Shan and Leang 2012a), effectively '*squaring*' the controller. This structure not only lowers the tracking error compared to conventional RC, but also offers good robustness for tracking odd-harmonic trajectories. It is noted that a similar dual-RC structure has been studied in (Kim and Tsao 2004), where two identical RCs are cascaded together (series connection); and a parallel configuration is presented in (Zhou et al. 2007). In contrast, the proposed dual-RC cascades an enhanced conventional RC with an odd-harmonic RC, and the series configuration is specifically tailored for tracking periodic scanning trajectories such as triangle signals with odd harmonics. Such reference signals are commonly used in piezo-based nanopositioners for raster-type and scanning applications, like AFM imaging.

The tracking performance of the conventional RC system shown in Fig. 7.22a is governed by the sensitivity function

$$S_{\text{rc}}(z) \triangleq \frac{E(z)}{R(z)} = \frac{[1 - H_1(z)]S(z)}{1 - H_1(z)[1 - k_{\text{rc}}G_0(z)S(z)]}, \tag{7.36}$$

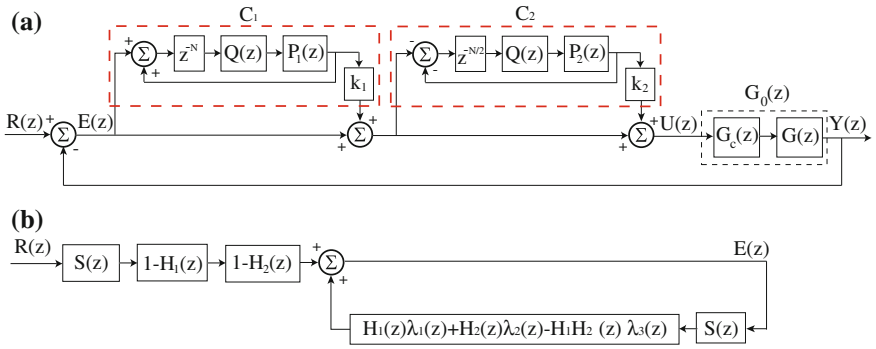


Fig. 7.23 **a** A dual-stage RC design consisting of a conventional RC (C_1) cascaded with an odd-harmonic RC (C_2) and **b** the equivalent block diagram of (**a**) for stability analysis

where $H_1(z) = Q(z)z^{-N+m_1}$ and $S(z) = 1/[1 + G_0(z)]$ is the sensitivity function of the feedback system without the repetitive controller. One approach to improve the tracking performance of the conventional RC is to reduce the magnitude of S_{rc} by cascading together two signal generators, effectively producing a squaring effect (Kim and Tsao 2004). However, the reference trajectories used in the scanning operation in SPMs are generally odd-harmonic signals (e.g., triangle trajectories), it is preferred that an odd-harmonic RC (Zhou et al. 2007) as depicted in Fig. 7.22b be cascaded with a conventional RC as shown in Fig. 7.23a, instead of cascading two conventional RCs. By doing this, the resultant sensitivity function is

$$\widetilde{S}_{\text{rc}}(z) = \frac{[1 - H_1(z)][1 - H_2(z)]}{W(z) + [1 - H_1(z)(1 - k_1)][1 - H_2(z)(1 - k_2)]G_0(z)}, \tag{7.37}$$

where $W(z) = [1-H_1(z)][1-H_2(z)]$ and $H_2(z) = -z^{-\frac{N}{2}+m_2}Q(z)$. The advantage of the enhanced dual-RC design over cascading two conventional RCs together is added performance and robustness. Cascading two conventional RCs together results in excessive gain at the even harmonics, which can degrade the system's performance for tracking odd-harmonic reference trajectories (Costa-Castello et al. 2004). The performance of the enhanced dual-RC is illustrated by comparing the magnitude response of the sensitivity function $\widetilde{S}_{\text{rc}}(z)$ of the enhanced dual-RC in Eq. (7.37) to the magnitude response of the sensitivity function $S_{\text{rc}}(z)$ of the conventional RC in Eq. (7.36) and the sensitivity function $\overline{S}_{\text{rc}}(z)$ of the odd-harmonic RC in Fig. 7.22b, given by

$$\overline{S}_{\text{rc}}(z) = \frac{[1 - H_2(z)]S(z)}{1 - H_2(z)[1 - k_{\text{rc}}G_0(z)S(z)]}. \tag{7.38}$$

The comparison of the three RC configurations is shown in Fig. 7.24, where the frequency response functions are generated in Matlab using the '`margin`' command

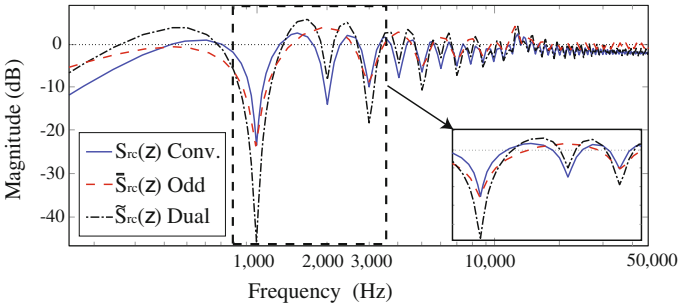


Fig. 7.24 Comparison of magnitude versus frequency plots for the sensitivity functions for different RC configurations, where $S_{\text{rc}}(z)$ denotes the conventional RC (*solid line*), $\bar{S}_{\text{rc}}(z)$ is for the odd-harmonic RC (*dash line*), and $\tilde{S}_{\text{rc}}(z)$ represents the dual RC (*dash-dot line*)

using $N = 100, m_1 = m_2 = 0, Q(z) = 1$, and $T_s = 10\,\mu s$ as an illustrative example. The results reveal that the odd-harmonic RC has little affect on the even-harmonics like the conventional RC (gain at first even harmonic: -13.7 dB for conventional RC, 4.49 dB for odd-harmonic RC, and -8.69 dB for dual-RC). Instead, the magnitude of the sensitivity function for the dual-RC is significantly lower than the conventional RC at the odd-harmonics (-24.4 dB for conventional RC vs. -47.1 dB for dual-RC at the first odd harmonic). This implies that (1) the odd-harmonic RC has the same tracking performance as the conventional RC for tracking odd-harmonic trajectories but it provides the system with more robustness by reducing the gain at the even harmonics, which effectively minimizes the amplification of signals in that frequency range, such as noise and (2) the dual-RC provides higher gain than the conventional RC at the odd-harmonics; therefore, the dual-RC will improve the tracking of trajectories with odd-harmonics.

The stability conditions for the dual-RC is presented as follows. First, the stability conditions for the odd-harmonic RC is presented, then the conditions for the dual-RC is presented. Readers are referred to Shan and Leang (Shan and Leang 2012a) for details of the stability analysis and proof.

Let T_s be the sampling time. Consider the odd-harmonic RC shown in Fig. 7.22b and the following assumptions: (1) the reference trajectory $R(z)$ is periodic in time with period T_p and (2) the closed-loop system without the RC is asymptotically stable, i.e., $1 + G_0(z) = 0$ has no roots outside of the unit circle in the z-plane. For the odd-harmonic RC, if Assumptions 1 and 2 hold and if $|Q(e^{j\omega T_s})| \leq 1$ and

$$0 < k_{\text{rc}} < \frac{2\cos[\theta_T(\omega)]}{A(\omega)} \text{ and } -\pi/2 < \theta_T(\omega) < \pi/2, \tag{7.39}$$

for $\omega \in (0, \pi/T_s)$, then the odd-harmonic RC feedback system shown in Fig. 7.22b is asymptotically stable. This result states that within an acceptable operating frequency range, there exists a sufficiently small RC gain k_{rc} such that the closed-loop

odd-harmonic RC system is stable. Next, the stability conditions for the dual-RC, created by cascading an odd-harmonic RC with the conventional RC, is presented.

Consider the enhanced dual-RC system shown in Fig. 7.23a. If Assumptions 1 and 2 hold and if $|Q(e^{j\omega T_s})| \leq 1$ and

$$\frac{3\cos[\theta_T(\omega)] - \Delta}{3A(\omega)} < k_1, k_2 < 1 + \sqrt{1 + \frac{3\cos[\theta_T(\omega)] + \Delta}{3A(\omega)}},$$
$$-\pi/9 \leq \theta_T(\omega) \leq \pi/9, \tag{7.40}$$

with $\Delta = \sqrt{9\cos^2[\theta_T(\omega)] - 8}$ for $\omega \in (0, \pi/T_s)$, then the closed-loop system in Fig. 7.23a is asymptotically stable (see Shan and Leang (2012a) for details of the stability analysis and proof). Therefore by satisfying the above conditions, that is by picking appropriate values for the RC gains, k_1 and k_2, within a particular operating frequency range, the dual-RC is guaranteed stable.

7.10.4 Handling Hysteresis

In the above analysis, the effects of hysteresis were not considered explicitly in the RC design. Hysteresis can drastically affect the performance of a closed-loop controller, particularly if the controller is designed around a linear dynamics model (Main and Garcia 1997). To keep the analysis simple, an approach to minimize the affect of hysteresis for RC is optimizing the resident feedback controller $G_c(z)$ in such a way that the closed-loop performance accounts for the hysteresis behavior over the bandwidth of interest. Additionally, it has been shown that high-gain feedback control is effective for significantly reducing hysteresis behavior (Leang and Devasia 2007). Another approach is depicted in Fig. 7.25a, where an internal feedback loop is used to linearize the plant dynamics (Choi et al. 2002). Recently in (Shan and Leang 2012b) the design of RC which factors in the hysteresis effect was studied. If the hysteresis nonlinearity exceeds a particular bound, the hysteresis can be accounted for using model-based feedforward compensation as illustrated in Fig. 7.25b (Ahn 2003; Shan and Leang 2012a, b) (see Chap. 11). Therefore, compensating for the hysteresis effect permits the application of the analysis presented above.

7.10.5 Design and Implementation

Two repetitive controllers were designed, implemented, and their responses were compared to PID control. The first was a standard RC with a low-pass filter $Q(z)$ in the RC loop. The standard RC did not include phase lead compensators. The second RC contained the two phase lead compensators z^{m_1} and z^{m_2} to improve the

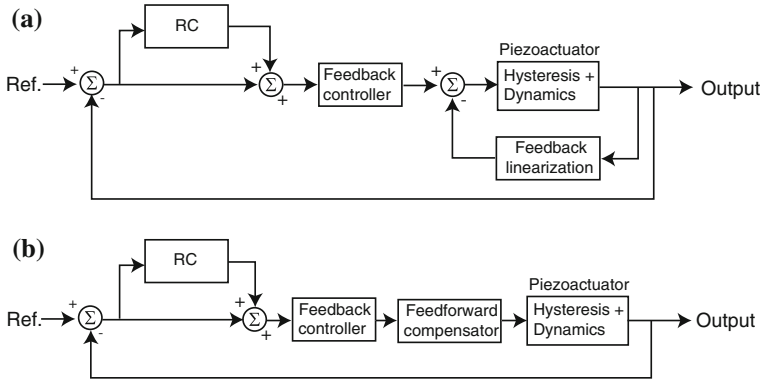


Fig. 7.25 Techniques to account for hysteresis in RC design. **a** feedback-linearization approach and **b** feedforward hysteresis compensation (Shan and Leang 2012a, b)

tracking performance and stability, respectively. The details of the design process are described below.

The experimental AFM system (Molecular Imaging PicoPlus model) and block diagram of the control system are shown in Fig. 7.26a, b, respectively. The AFM uses a piezoelectric tube-shaped actuator for positioning the cantilever and probe tip. The AFM was customized to permit the application of control signals to control the movement of the piezoactuator in the three coordinate axes (x, y, and z). Inductive sensors were used to measure the displacements of the piezoactuator and the signals were accessible through a custom signal access module. The gain of the inductive sensors were 96–97 μm/V in the x-axis and y-axis, respectively. A PC computer and data acquisition system running custom C code were used to implement the RC control system. The sampling frequency of the data acquisition and control hardware was 10 kHz.

The RC was applied to track a periodic reference trajectory in the x-axis as an illustrative example. This axis was the fast-scanning axis because the probe tip was moved back and forth at least 100 times faster than the up and down motion in the y-direction during imaging. For example, a 100×100 pixel image requires the AFM tip to scan back and forth across the sample surface 100 times and slowly move from top to bottom. It is noted that the effects due to cross-coupling in piezo-tube actuators were not considered in this work. Interested readers are referred to the work of Tien et al. (2005), for additional details to further improve tracking performance.

A linear dynamics model for the piezoactuator was obtained for designing the RC system. The model was found by curve fitting the measured frequency response function. The frequency response along the x axis was measured using a dynamic signal analyzer (DSA, Hewlett Packard, Model 35670A). The response was measured over small ranges to minimize the effects of hysteresis and above 1 Hz to avoid the effects of creep (Croft et al. 2001). The resulting frequency response curves are shown in Fig. 7.27. A linear 12th-order transfer function model $G(s)$ (dash-dot line

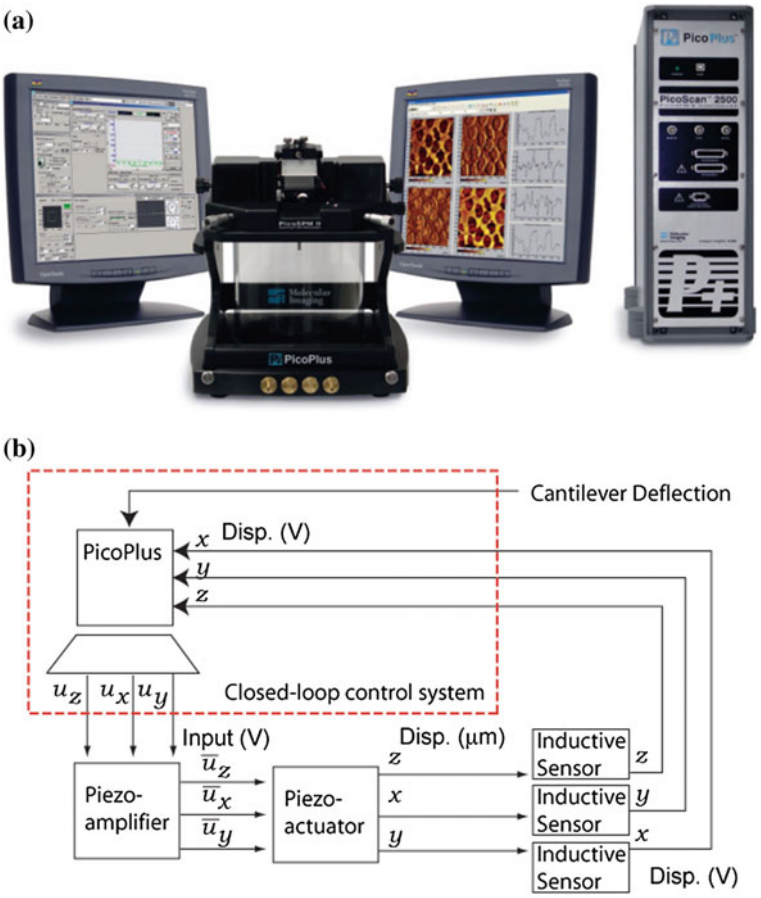


Fig. 7.26 **a** The AFM system and **b** a block diagram of the AFM and control system. An external computer running custom C code was used to implement the control algorithm

in Fig. 7.27) was curve fitted to the measured frequency response function. The continuous-time model was then converted to the discrete-time model $G_p(z)$ using the Matlab function `c2d` with a sampling frequency of 10 kHz (shown by the dashed line in Fig. 7.27).

Prior to integrating the RC, a PID controller was designed for the piezoactuator to control the motion along the x axis. The PID controller is given by

$$G_c(z) = K_p + K_i \left(\frac{z}{z-1} \right) + K_d \left(\frac{z-1}{z} \right), \tag{7.41}$$

where the Ziegler-Nichols method (Franklin et al. 2006) was used to tune the parameters of the controller to $K_p = 1$, $K_i = 1450$, and $K_d = 0.0002$. The PID controller

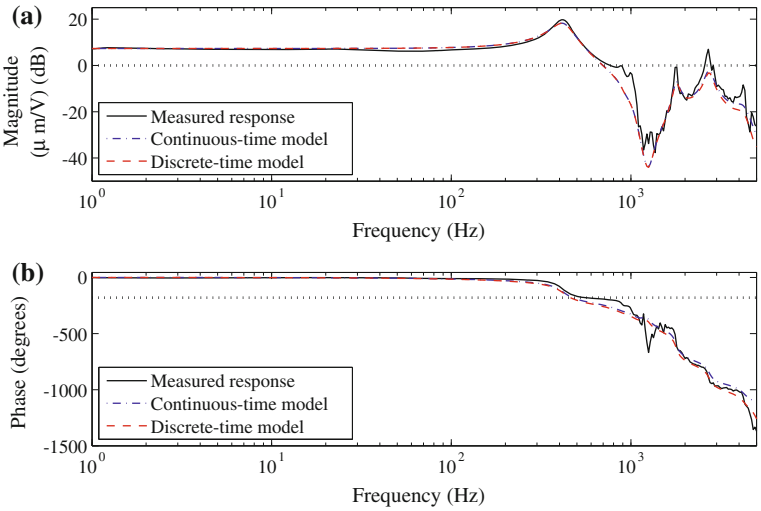


Fig. 7.27 The frequency response of piezoactuator along the x axis. The *solid line* is the measured response; the *dash-dot* line represents the linear continuous-time model $G(s)$; and the *dash line* is the linear discrete-time model $G_p(z)$ using Matlab function `c2d` with zero-order hold and sampling frequency of 10 kHz

was implemented at a sampling frequency of 10 kHz. The performance of the PID controller to a step reference is shown in Fig. 7.28a. It can be observed that without PID control, the open-loop response shows significant overshoot. Also, after 30 ms creep effect becomes noticeable. Creep is a slow behavior and after several minutes the tracking error can be in excess of 20 % (Leang and Devasia 2006). On the other hand, the PID controller minimized the overshoot and creep effect.

The response of the PID controller for tracking a triangular trajectory at 1, 5, and 25 Hz are shown in Fig. 7.28b. Triangle reference signals are commonly used in AFM imaging. The maximum tracking error for the three cases are shown in Fig. 7.28c. The error at 1 Hz (low speed) was relatively small, approximately 1.48 % of the 10-μm range (±5 μm). However, at 25 Hz (high speed) scanning the error was unacceptably large at 10.70 %. Due to vibrational dynamics and hysteresis effects, open-loop AFM imaging is limited to less than 2–3 Hz. The objective was to reduce the tracking error by adding a repetitive controller to the PID loop.

The next steps are to design the low-pass filter and phase lead z^{m_2} for stability and robustness, followed by designing the phase lead z^{m_1} to minimize the steady-state tracking error. The steps are outlined as follows:

First, the RC was designed for stability and robustness. This involves designing a low-pass filter $Q(z)$ and adding phase lead via m_2 to satisfy the conditions given by Eqs. (7.34) and (7.35). The following low-pass filter was used in the RC loop,

$$Q(z) = \frac{a}{z+b}, \tag{7.42}$$

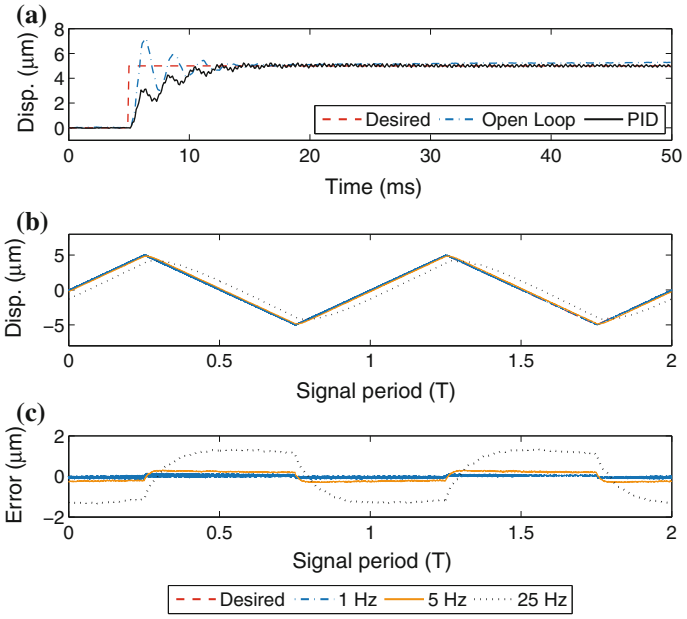


Fig. 7.28 The measured responses of the PID controller to **a** a step reference and **b** triangle references at 1, 5, and 25 Hz. **c** The tracking error for the triangle reference signals associated with plot (**b**)

where $|a| + |b| = 1$. The cutoff frequency ω_Q of the low-pass filter was chosen below the $\pm 90°$ crossover frequency to satisfy Eq. (7.35). The low-pass filter cutoff frequency is limited by the crossover frequency. Also, the cutoff frequency limits the achievable scan rate to about one-tenth of the cutoff frequency, i.e., $\omega_Q/10$.

The phase response $\theta_T(\omega)$, of the closed-loop feedback system without RC, and different phase lead $\theta_2(\omega)$ are shown in Fig. 7.29. Without phase lead ($m_2 = 0$), the $\pm 90°$ crossover frequency was approximately 486 Hz. This value sets the maximum cutoff frequency for the low-pass filter and the maximum scan rate.

Next, simulations were done to show the tracking performance of RC. The chosen cutoff frequency for $Q(z)$ was 250 Hz and zero phase lead ($m_2 = 0$) was used. Therefore, the maximum scan rate is 25 Hz. It is noted that for higher rate scanning, the cutoff can be increased, but only up to 486 Hz when $m_2 = 0$ (see Fig. 7.29). The 250 Hz cutoff frequency was chosen because it provided a safety margin of approximately two. Then, the RC gain was determined by satisfying Eq. (7.34), for instance picking $k_{rc} = 0.40$. The simulated tracking response for $\pm 25\,\mu$m scan range at 25 Hz is shown in Fig. 7.30. The first two plots, Fig. 7.30a1 and b1, show the tracking performance and error, respectively, for a stable RC system without any phase lead compensators, i.e., $m_1 = m_2 = 0$. In this case, increasing k_{rc} and/or the low-pass filter's cutoff frequency caused instability. Reducing the RC gain, however,

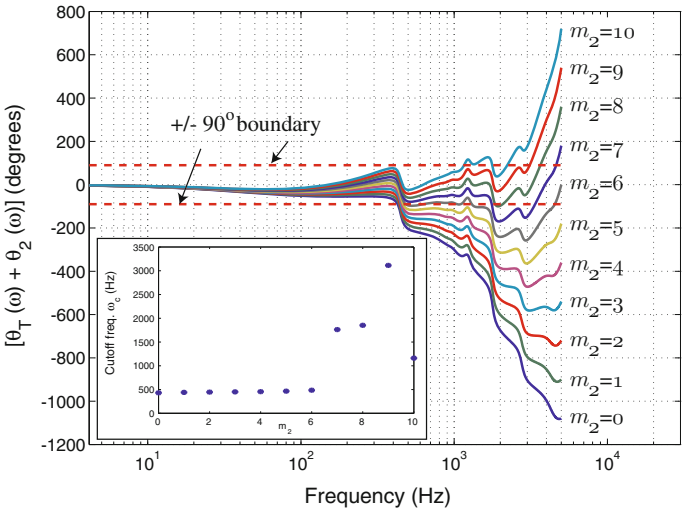


Fig. 7.29 The phase response of the closed-loop feedback system without RC and added phase lead $\theta_2(\omega)$, stability condition Eq. (7.35). The inset plot shows the cutoff frequency versus the phase lead parameter m_2. As m_2 increases, the frequency range for stability increases. A maximum is reached when $m_2 = 9$

reduced the convergence rate. The steady-state tracking error was minimally affected by the RC gain and the phase lead through m_2.

The scan rate can be improved by increasing the $\pm 90°$ crossover frequency by adding phase lead through the parameter m_2. The inset in Fig. 7.29 shows the $\pm 90°$ crossover frequency versus the phase lead parameter m_2.

With the addition of phase lead, such as $m_2 = 7$, the $\pm 90°$ crossover frequency was increased to approximately 2,000 Hz. Therefore, the low-pass filter's cutoff frequency can be improved to raise the RC's bandwidth permitting tracking of higher frequency components. Subsequently, the RC gain Eq. (7.34) can be increased. For example with $m_2 = 7$, $k_{rc} = 1.1$, and simulation results are shown in Fig. 7.30a2, b2 that demonstrate improvement in the convergence rate and reduced tracking error compared to the previous case without phase lead z^{m_2}. As indicated in the inset plot in Fig. 7.29, higher values of m_2 show no improvement in the crossover frequency.

Simulations were done with $k_{rc} = 0.4$ to verify the stability of the closed-loop system with RC for different low-pass filter cutoff frequencies and values of m_2. The results are summarized in Table 7.3. Comparing the inset plot in Fig. 7.29 and the summary in Table 7.3, with $m_2 = 0$ the closed-loop RC system is stable when the low-pass filter frequency is below the crossover frequency of 486 Hz. As the cutoff frequency increases, for example at 500 Hz and above, the RC system is unstable. But the stability can be achieved by adding phase lead through m_2 as shown by the results in Table 7.3.

Finally, by adding phase lead using z^{m_1} in the RC loop, for example $m_1 = 6$, the maximum tracking error, defined as

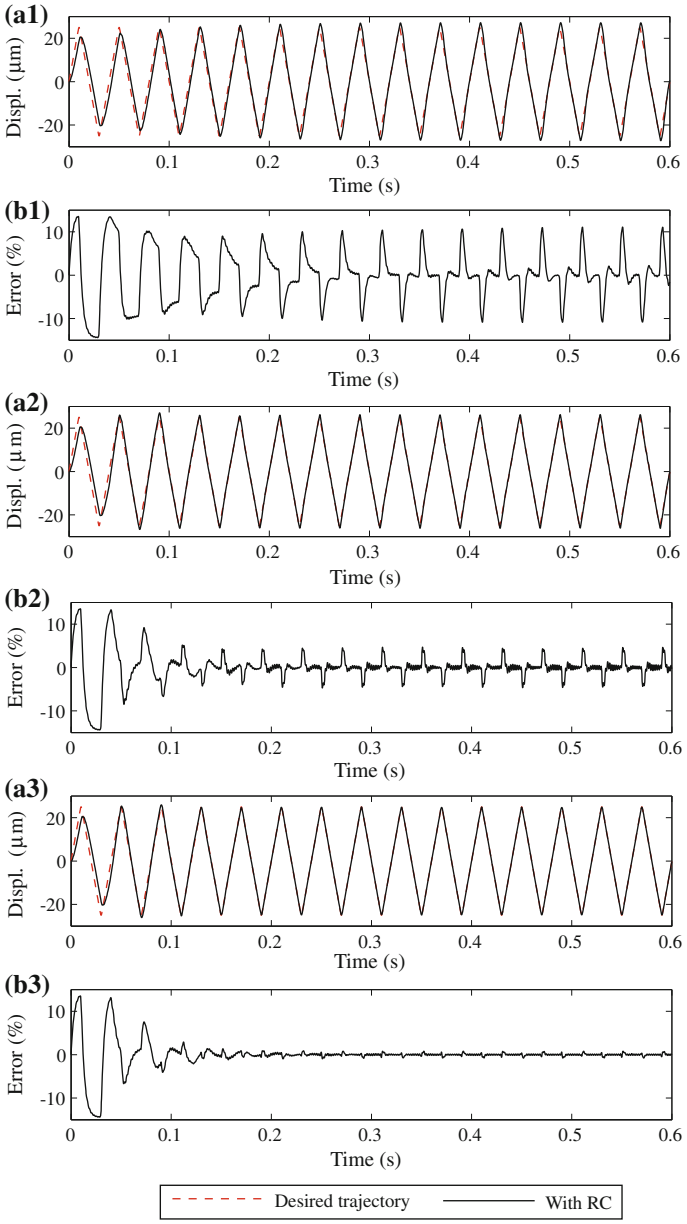


Fig. 7.30 Simulation results showing the tracking performance and error for scanning at 25 Hz, where **a1** and **b1** belong to RC with $k_{rc} = 0.40$ and no phase lead; **a2** and **b2** belong to RC with phase lead $m_2 = 7$ and $k_{rc} = 1.1$; **a3** and **b3** belong to RC with phase leads $m_1 = 6$, $m_2 = 7$ and $k_{rc} = 1.1$

Table 7.3 Stability of RC system for different low-pass filter cutoff frequencies and phase lead z^{m_2}

Phase lead m_2	Low-pass filter $Q(z)$'s cutoff frequency (Hz)				
	250	500	1000	2000	4000
0	Stable	Unstable	Unstable	Unstable	Unstable
2	Stable	Unstable	Unstable	Unstable	Unstable
4	Stable	Stable	Unstable	Unstable	Unstable
6	Stable	Stable	Stable	Unstable	Unstable
8	Stable	Stable	Stable	Stable	Stable

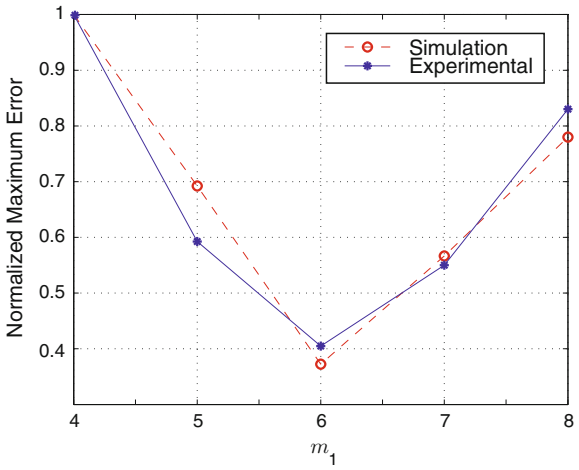


Fig. 7.31 Maximum error versus phase lead parameter m_1. For the experiments, $m_1 = 6$ gave smallest error

$$e_{\max}(\%) = \left[\frac{\max|y-r|}{\max(y)-\min(y)}\right] \times 100\,\%, \tag{7.43}$$

where y and r are the measured and reference outputs, respectively, was substantially reduced from 11.96 and 5.32 % [Fig. 7.30a2, b2] to 0.97 % of the total range (50 μm) as illustrated in Fig. 7.30a3, b3.

The optimum value of the phase lead m_1 was determined by looking at the maximum error versus different m_1 values. The simulation results are shown in Fig. 7.31, plotted as normalized maximum error versus m_1, along with experimental results, which will be discussed in the following section. As shown in the figure, the optimum value is $m_1 = 6$ and this value was also used in the experiments discussed below.

In the experiment, the reference signal was a 25-μm triangle wave at 5, 10, and 25 Hz. The reference trajectory was passed through a two-pole zero-phase-shift filter with cutoff frequency 250 Hz to remove high frequency components before applying it to the closed-loop system. Triangle scan signals are typically used for

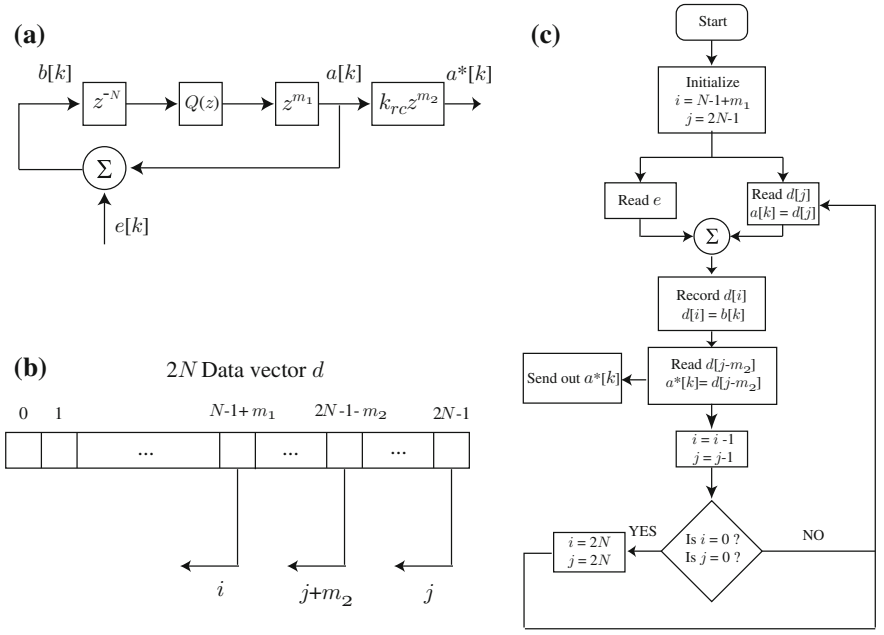


Fig. 7.32 Digital implementation of repetitive control. **a** Equivalent discrete-time block diagram of the RC loop. **b** Linear data vector for implementing the one-period delay and the phase lead compensators. **c** The flow diagram for implementing the RC loop

AFM imaging, and they were filtered to avoid exciting high-frequency dynamics. The cutoff frequency for the low-pass filter $Q(z)$ in the RC loop was set at 250 Hz. Due to hardware limitations where the sampling frequency was 10 kHz, $m_2 = 0$ was chosen to give a maximum scan frequency of 25 Hz. The RC gain was chosen as $k_{\text{rc}} = 0.40$ and this value satisfied the condition given by Eq. (7.34).

Let N be an integer value representing the delay period, the ratio of signal period T_p to the sampling period T_s. Figure 7.32a shows the equivalent discrete-time block diagram for the RC loop, where z^{-N} is a delay of period N. The two phase lead compensators, z^{m_1} and z^{m_2}, had leads of $m_1 = 6$ and $m_2 = 0$. Both the delay and phase leads were implemented using a linear data vector d as shown in Fig. 7.32b with $2N$ elements. Two counters i and j were used, one controlled the location where incoming data was stored to the data vector and the other controlled the location where data was read and sent. The difference in the indices i and j determines the overall delay $-N + m_1 + m_2$, and since $N >> m_1 + m_2$, then the delay implementation is causal. The flow diagram for the RC implementation with respect to the linear data vector d is shown in Fig. 7.32c. Upon reaching the end of the array at $i = 0$ and $j = 0$, both indices were reset to $2N - 1$ and the process was repeated.

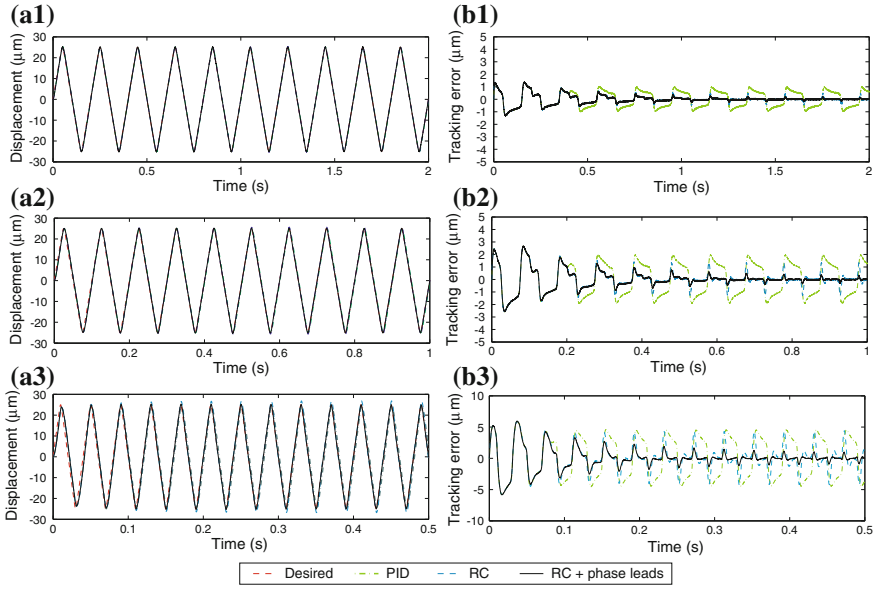


Fig. 7.33 Experimental tracking response and error for PID (*dash-dot*), RC (*dashed line*), and RC with phase lead compensation [$m_1 = 6$ and $m_2 = 0$] (*solid line*) for 5 Hz (**a1** and **b1**), 10 Hz (**a2** and **b2**), and 25 Hz (**a3** and **b3**) scanning

7.10.6 Experimental Results and Discussion

The tracking results for the PID, regular RC, and the RC with the phase lead compensators for ±25-μm scanning at 5, 10, and 25 Hz are presented in Fig. 7.33 and Table 7.4. The steady-state tracking errors, measured at the last two cycles, are reported as a percentage of the range of motion. In particular, the maximum error Eq. (7.43) and the root-mean-squared error defined as

$$e_{\text{rms}}(\%) = \left[\frac{\sqrt{\frac{1}{T}\int_0^T \left[y(t) - r(t)\right]^2 dt}}{\max(y) - \min(y)} \right] \times 100\,\% \tag{7.44}$$

are reported.

Because the action of the repetitive controller is delayed by one scan period, the tracking response for the first period is similar for the PID, RC, and RC with phase lead compensation as shown in Fig. 7.33. However, after the first period, the RC begins to take action as illustrated by reducing tracking error from one cycle to the next. On the other hand, the tracking error of the PID controller persists from one cycle to the next.

The 5 Hz scanning results shown in Fig. 7.33a1, b1 and Table 7.4 demonstrate that the regular RC controller reduced maximum tracking error from 2.01 to 0.96 %

Table 7.4 Tracking results for ±25-μm range

Controller	5 Hz		10 Hz		25 Hz	
	$e_{max}(\%)$	$e_{rms}(\%)$	$e_{max}(\%)$	$e_{rms}(\%)$	$e_{max}(\%)$	$e_{rms}(\%)$
PID	2.01	1.28	3.99	2.61	9.16	6.61
RC	0.96	0.21	2.74	0.79	8.86	3.69
RC + phase leads	0.43	0.08	0.46	0.10	1.78	0.57

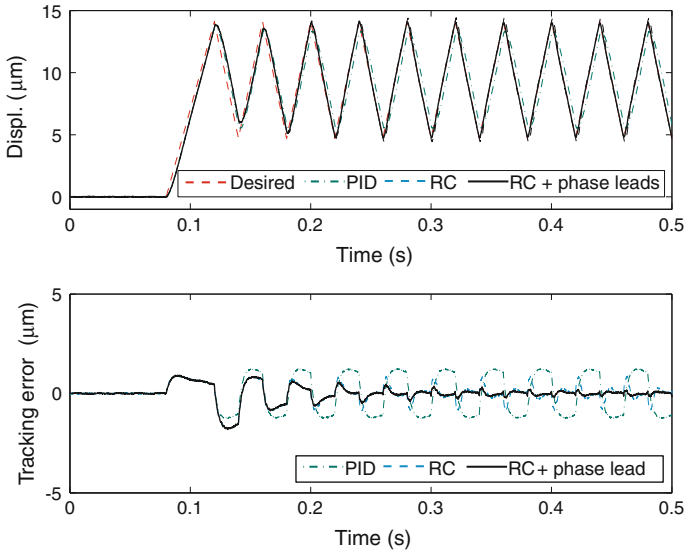


Fig. 7.34 Tracking results for offset triangle scan at 25 Hz

compared to the PID controller, a 52 % reduction. By using RC with the phase lead compensation, an additional 55 % improvement in tracking performance was achieved. In this case, the maximum tracking error is 0.43 %.

At 25 Hz, the tracking error of PID was unacceptable large at 9.16 %. In fact, for AFM scanning operations the maximum tracking error should be less than a few percent. The results in Table 7.4 show that the regular plug-in RC controller was not able to improve the tracking performance at 25 Hz. However, the RC with phase lead compensation gave lower maximum tracking error at 1.78 %. Therefore, the RC with phase lead compensation enables precision tracking at higher scan rates. The optimum value of the phase lead via m_1 was chosen using the simulation results in Fig. 7.31. The simulation results were validated in the experiments as shown in the figure, where $m_1 = 6$ gave the lowest steady-state tracking error.

Finally, scanning offset from the piezoactuator's center position is demonstrated as shown in Fig. 7.34. For this offset scanning operation, the PID controller accounted for the low frequency dynamics such as creep and the RC was used for tracking the

periodic trajectory. The tracking results in Fig. 7.34 show that the RC was effective at minimizing the tracking error.

7.11 Summary

Feedback controllers can be straightforward to design and naturally compensate for many sources of positioning error and nonlinearity. The foremost disadvantage is the need for a position sensor and the possibility of instabilities if plant uncertainty is not taken in account.

This chapter considers three simple controller designs: PI control, inverse control, and IRC damping control. The integral controller was simplest to design and implement but provided the lowest closed-loop bandwidth. An inverse controller (notch filter) can provide much greater bandwidth when the dynamics are well known. However, if the resonance frequency is expected to vary by more than a few percent, the controller must be designed conservatively which can limit the achievable performance.

Integral resonance control (IRC) is a new control strategy that damps the system resonance rather than inverting it. The foremost advantages are simplicity, robustness, and insensitivity to variations in the resonance frequencies. In the experimental comparison, where the resonance frequency varied by 19 %, the settling time of the IRC controller with one-fifth that of the inverse controller.

When the reference trajectory is periodic, RC can significantly improve the tracking performance of a feedback loop. A repetitive controller was combined with a PID feedback system for precise tracking of periodic trajectories with disturbance rejection. Experimental results demonstrate the effectiveness of the RC approach. With a 25 Hz triangular reference signal, the maximum tracking error was less than 2 % using the improved RC technique compared to 9.16 % with standard PID control.

References

Abramovitch DY, Andersson SB, Pao LY, Schitter G (2007) A tutorial on the mechanisms, dynamics, and control of atomic force microscopes. In: Proceeding of American control conference, New York City, pp 3488–3502

Abramovitch DY, Hoen S, Workman R (2008) Semi-automatic tuning of PID gains for atomic force microscopes. In: American control conference, Seattle, pp 2684–2689

Ahn H-S (2003) Design of a repetitive control system for a piezoelectric actuator based on the inverse hysteresis model. In: The fourth international conference on control and automation, pp 128–132

Ando T, Kodera N, Uchihashi T, Miyagi A, Nakakita R, Yamashita H, Matada K (2005) High-speed atomic force microscopy for capturing dynamic behavior of protein molecules at work. e-J Surf Sci Nanotechnol 3:384–392

Ando T, Uchihashi T, Fukuma T (2008) High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes. Prog Surf Sci 83(7–9):337–437

Aphale SS, Bhikkaji B, Moheimani SOR (2008) Minimizing scanning errors in piezoelectric stack-actuated nanopositioning platforms. IEEE Trans Nanotechnol 7(1):79–90
Aphale SS, Fleming AJ, Moheimani SOR (2007) Integral control of resonant systems with collocated sensor-actuator pairs. IOP Smart Mater Struct 16:439–446
Aridogan U, Shan Y, Leang KK (2009) Design and analysis of discrete-time repetitive control for scanning probe microscopes. ASME J Dyn Syst Meas Control 131:061103 (12 p)
Arimoto S, Kawamura S, Miyazaki F (1984) Bettering operation of robots by learning. J Robot Syst 1(2):123–140
Barrett RC, Quate CF (1991) Optical scan-correction system applied to atomic force microscopy. Rev Sci Instrum 62(6):1393–1399
Bhikkaji B, Moheimani SOR (2008) Integral resonant control of a piezoelectric tube actuator for fast nano-scale positioning. IEEE Trans Mechatron 13(5):530–537
Broberg HL, Molyet RG (1994) A new approach to phase cancellation in repetitive control. In: IEEE industry applications society annual meeting, vol 3, pp 1766–1770
Chen CJ (1992) Electromechanical deflections of piezoelectric tubes with quartered electrodes. Appl Phys Lett 60(1):132–134
Chen S-L, Hsieh T-H (2007) Repetitive control design and implementation for linear motor machine tool. Int J Mach Tools Manuf 47(12–13):1807–1816
Chew KK, Tomizuka M (1990) Digital control of repetitive errors in disk drive systems. IEEE Control Syst Mag 10(1):16–20
Choi GS, Lim YA, Choi GH (2002) Tracking position control of piezoelectric actuators for periodic reference inputs. Mechatronics 12(5):669–684
Clayton GM, Tien S, Leang KK, Zou Q, Devasia S (2009) A review of feedforward control approaches in nanopositioning for high-speed SPM. J Dyn Syst Meas Control 131:061 101(1–19)
Costa-Castello R, Grino R, Fossas E (2004) Odd-harmonic digital repetitive control of a single-phase current active filter. IEEE Trans Power Electron 19(4):1060–1068
Croft D, Shed G, Devasia S (2001) Creep, hysteresis, and vibration compensation for piezoactuators: atomic force microscopy application. ASME Trans, J Dyn Syst Meas Control 123:35–43
Eielsen AA, Burger M, Gravdahl JT, Pettersen KY (2011) Pi^2-controller applied to a piezoelectric nanopositioner using conditional integrators and optimal tuning. In: Proceeding of IFAC World Congress, vol 18, Milano
Fanson JL, Caughey TK (1990) Positive position feedback control for large space structures. AIAA J 28(4):717–724
Fantner GE, Hegarty P, Kindt JH, Schitter G, Cidade GAG, Hansma PK (2005) Data acquisition system for high speed atomic force microscopy. Rev Sci Instrum 76(2):026 118-1–026 118–4
Fleming AJ, Aphale SS, Moheimani SOR (2010) A new method for robust damping and tracking control of scanning probe microscope positioning stages. IEEE Trans Nanotechnol 9(4):438–448
Fleming AJ, Behrens S, Moheimani SOR (2002) Optimization and implementation of multi-mode piezoelectric shunt damping systems. IEEE/ASME Trans Mechatron 7(1):87–94
Fleming AJ, Kenton BJ, Leang KK (2010) Bridging the gap between conventional and video-speed scanning probe microscopes. Ultramicroscopy 110(9):1205–1214
Fleming AJ, Moheimani SOR (2006) Sensorless vibration suppression and scan compensation for piezoelectric tube nanopositioners. IEEE Trans Control Syst Technol 14(1):33–44
Fleming AJ, Wills AG (2009) Optimal periodic trajectories for band-limited systems. IEEE Trans Control Syst Technol 13(3):552–562
Francis BA, Wonham WM (1976) The internal model principle of control theory. Automatica 12(5):457–465
Franklin GF, Powell JD, Emami-Naeini A (2006) Feedback control of dynamic systems, 5th ed. Prentice Hall, Upper Saddle River
Griffith JE, Miller GL, Green CA, Grigg DA, Russell PE (1990) A scanning tunneling microscope with a capacitance based position monitor. J Vac Sci Technol B: Microelectron Nanometer Struct 8(6):2023–2027

Hara S, Yamamoto Y, Omata T, Nakano M (1988) Repetitive control system: a new type servo system for periodic exogenous signals. IEEE Trans Autom Control 33(7):659–668
Humphris ADH, Miles MJ, Hobbs JK (2005) A mechanical microscope: high-speed atomic force microscopy. Appl Phys Lett 86:034 106-1–034 106–3
Inoue T, Nakano M, Iwai S (1981) High accuracy control of a proton synchrotron magnet power supply. In: Proceeding of 8th IFAC World Congress, vol 20, pp 216–221
Kenton BJ, Leang KK (2012) Design and control of a three-axis serial-kinematic high-bandwidth nanopositioner. IEEE/ASME Trans Mechatron 17(2):356–369
Kim B-S, Tsao T-C (2004) A performance enhancement scheme for robust repetitive control system. ASME J Dyn Syst Meas Control 126(1):224–229
Leang KK, Devasia S (2007) Feedback-linearized inverse feedforward for creep, hysteresis, and vibration compensation in afm piezoactuators. IEEE Trans Control Syst Technol 15(5):927–935
Leang KK, Devasia S (2006) Design of hysteresis-compensating iterative learning control for piezo positioners: application to atomic force microscopes. Mechatronics 16(3–4):141–158
Lee H-J, Saravanos DA (1998) The effect of temperature dependent material properties on the response of piezoelectric composite materials. J Intell Mater Syst Struct 9(7):503–508, 1998
Lee C, Salapaka S (August 2009) Fast robust nanopositioning: a linear-matrix-inequalities-based optimal control approach. IEEE/ASME Trans Mechatron 14(4):414–422
Li CJ, Li SY (1996) To improve workpiece roundness in precision diamond turning by in situ measurement and repetitive control. Mechatronics 6(5):523–535
Li Y, Ang KH, Chong G (2006) Pid control system analysis and design. IEEE Control Syst 26(1):32–41
Lowrie F, Cain M, Stewart M, Gee M (1999) Time dependent behaviour of piezo-electric materials. National physical laboratory, Technical report, 151
Main JA, Garcia E (1997) Piezoelectric stack actuators and control system design: strategies and pitfalls. AIAA J Guidance Control Dyn 20(3):479–485
Merry RJE, Ronde MJC, van de Molengraft R, Koops KR, Steinbuch M (2011) Directional repetitive control of a metrological afm. IEEE Trans Control Syst Tech 19(6):1622–1629
Moore KL, Dahleh M, Bhattacharyya SP (1992) Iterative learning control: a survey and new results. J Robot Syst 9(5):563–594
Radmacher M (1997) Measuring the elastic properties of biological samples with the afm. IEEE Eng Med Biol 16:47–57
Ratcliffe JD, Lewin PL, Rogers E (2005) Stable repetitive control by frequency aliasing. In: Intelligent control systems and optimization, pp 323–326
Rost MJ, Crama L, Schakel P, van Tol E, van Velzen-Williams GBEM, Overgauw CF, ter Horst H, Dekker H, Okhuijsen B, Seynen M, Vijftigschild A, Han P, Katan AJ, Schoots K, Schumm R, van Loo W, Oosterkamp TH, Frenken JWM (2005) Scanning probe microscopes go video rate and beyond. Rev Sci Instrum 76(5):053 710-1–053 710-9
Salapaka SM, Salapaka MV (2008) Scanning probe microscopy. IEEE Control Syst Mag 28(2):65–83
Salapaka S, Sebastian A, Cleveland JP, Salapaka MV (2002) High bandwidth nano-positioner: a robust control approach. Rev Sci Instrum 75(9):3232–3241
Schitter G (2009) Improving the speed of AFM by mechatronic design and modern control methods. Tech Mess 76(5):266–273
Schitter G, Åström KJ, DeMartini BE, Thurner PJ, Turner KL, Hansma PK (2007) Design and modeling of a high-speed AFM-scanner. IEEE Trans Control Syst Technol 15(5):906–915
Sebastian A, Pantazi A, Moheimani SOR, Pozidis H, Eleftheriou E (2008) A self servo writing scheme for a MEMS storage device with sub-nanometer precision. In: Proceeding of IFAC World Congress, Seoul, pp 9241–9247
Sebastian A, Salapaka S (2005) Design methodologies for robust nano-positioning. IEEE Trans Control Syst Technol 13(6):868–876
Shan Y, Leang KK (2012a) Accounting for hysteresis in repetitive control design: nanopositioning example. Automatica 48(8):1751–1758

Shan Y, Leang KK (2012b) Dual-stage repetitive control with Prandtl-Ishlinskii hysteresis inversion for piezo-based nanopositioning. Mechatronics 22:271–281
Shan Y, Leang KK (2013) Mechanical design and control for high-speed nanopositioning: serial-kinematic nanopositioners and repetitive control for nanofabrication. IEEE Control Syst Mag (Special Issue on Dynamics and Control of Micro and Naoscale Systems) 33(6):86–105
Steinbuch M, Weiland S, Singh T (2007) Design of noise and period-time robust high-order repetitive control, with application to optical storage. Automatica 43(12):2086–2095
Tamer N, Dahleh M (1994) Feedback control of piezoelectric tube scanners. In: Proceeding of American control conference, Lake Buena Vista, pp 1826–1831
Teo YR, Fleming AJ (2014) A new repetitive control scheme based on non-causal fir filters. In: Proceeding of American control conference, Portland
Tien S, Zou Q, Devasia S (2005) Iterative control of dynamics-coupling-caused errors in piezoscanners during high-speed afm operation. IEEE Trans Control Syst Tech 13(6):921–931
Tomizuka M, Tsao TC, Chew KK (1998) Discrete time domain analysis and synthesis of repetitive controllers. In: American control conference, pp 860–866
Wu Y, Zou Q (2007) Iterative control approach to compensate for both the hysteresis and the dynamics effects of piezo actuators. IEEE Trans Control Syst Technol 15(5):936–944
Wu Y, Zou Q (2009) Robust inversion-based 2-dof control design for output tracking: Piezoelectric-actuator example. IEEE Trans Control Syst Technol 17(5):1069–1082
Zhou K, Wang D, Zhang B, Wang Y, Ferreira JA, de Haan SWH (2007) Dual-mode structure digital repetitive control. Automatica 43:546–554
Zhou K, Doyle JC (1998) Essentials of robust control. Prentice-Hall, Upper Saddle River

Chapter 8
Force Feedback Control

Up to this point, nanopositioning controllers have used displacement sensors as the feedback variable. The major drawbacks of typical displacement sensors are the limited bandwidth associated measurement noise. In this chapter the actuator load force of a nanopositioning stage is used as a feedback variable to achieve both tracking and damping. The load force is measured with a small piezoelectric transducer placed between the actuator and moving platform. Compared to a standard position sensor, the load force sensor is simple, low-cost, compact and extremely sensitive. In addition, the resulting system also exhibits a zero-pole ordering that allows a simple integral controller to achieve both damping and tracking.

8.1 Introduction

From the discussion in Chap. 7 it should be clear that feedback controllers can provide good performance at low-frequencies, however, the maximum gain and closed-loop bandwidth are severely limited by the presence of a lightly damped mechanical resonance. Damping controllers can provide a substantial improvement but the tracking controller bandwidth is still restricted by low stability margins. A further limitation of present techniques is the high sensor-induced noise which places a penalty on positioning resolution as bandwidth is increased.

In this chapter, a new method for feedback control of nanopositioning systems is presented. As shown in Fig. 8.1 a measurement of the force applied to the moving platform is utilized as a feedback variable for both tracking and damping control. A major benefit of this arrangement is that the resulting system exhibits a zero-pole ordering, meaning that the resonant zeros of the system appear lower in frequency than the resonant poles. Section 8.2 also presents a new modeling technique for piezoelectric actuators. Rather than modeling piezoelectric actuators as displacement actuators, they are modeled as force actuators. This technique provides a more intuitive understanding of actuator dynamics and is simpler to apply.

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_8,

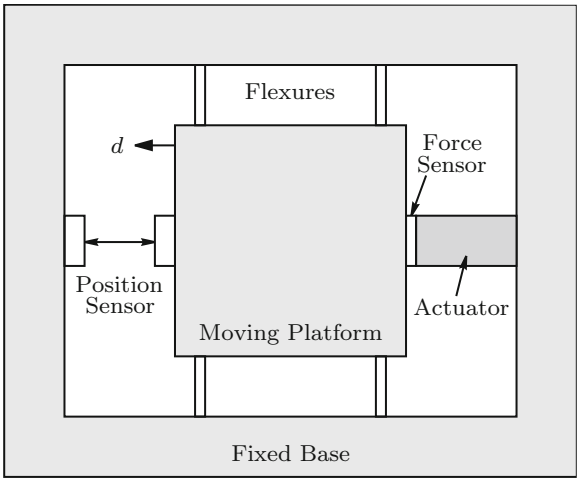


Fig. 8.1 A single degree-of-freedom positioning stage. The actuator generates a force which causes the platform to displace laterally. The force sensor measures actuator load while the position sensor measures platform displacement

In Sect. 8.3 the unique properties of the system described in Sect. 8.2 are exploited to provide damping control. A simple integral controller is shown to provide damping performance without any limitations on gain. The system is guaranteed to be stable with a theoretically infinite gain-margin and 90° phase-margin.

In addition to damping control, the controller described in Sect. 8.3 can be extended to provide tracking control without loss of performance or stability margins. As the noise generated by a piezoelectric force sensor is much lesser than a capacitive or inductive position sensor, the closed-loop positioning noise is also substantially reduced. The performance of the proposed techniques are demonstrated experimentally in Sect. 8.5.

The increased bandwidth and resolution offered by the proposed technique, combined with the simple implementation and high level of robustness, will allow nanopositioning systems to be employed in a new range of high-speed applications. For example, due to the performance penalties associated with closed-loop control, high-speed scanning probe microscopes currently use open-loop nanopositioners (Ando et al. 2005; Schitter et al. 2007; Humphris et al. 2005; Rost et al. 2005). Due to the simplicity and bandwidth of the proposed technique, such applications can now utilize closed-loop control with the associated benefits of improved linearity, less vibration and rejection of disturbance.

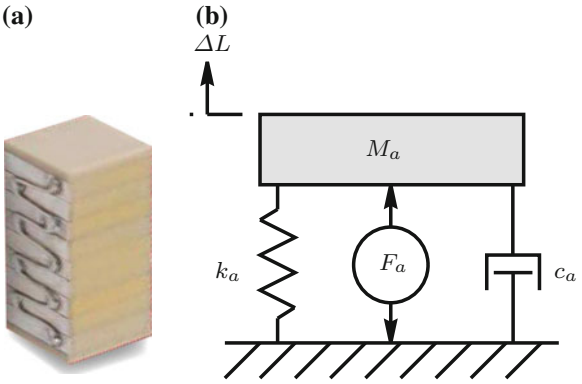


Fig. 8.2 **a** A Noliac monolithic stack actuator represented in **b** by a voltage dependent force F_a, stiffness k_a, effective mass M_a and damping coefficient c_a

8.2 Modeling

In this section, a model is derived for the single degree-of-freedom lateral positioning platform illustrated in Fig. 8.1. In this device, the force developed by a piezoelectric actuator displaces the central platform. The flexures represent the stiffness introduced by guiding flexures and mechanical linkages that are often present between the actuator and platform. Although the model presented is simple, it adequately represents the dominant dynamics exhibited by many nanopositioning geometries.

8.2.1 Actuator Dynamics

A typical multi-layer monolithic stack actuator is pictured in Fig. 8.2a. The actuator experiences an internal stress in response to an applied voltage. This stress is represented by the voltage dependent force F_a and is related to free displacement by

$$\Delta L = \frac{F_a}{k_a} \tag{8.1}$$

where ΔL is the change in actuator length (in m) and k_a is the actuator stiffness (in N/m).

The developed force F_a is most easily related to applied voltage by beginning with the standard expression for unrestrained linear stack actuators (Adriaens et al. 2000),

$$\Delta L = d_{33} n V_a, \tag{8.2}$$

where d_{33} is the piezoelectric strain constant (in m/V), n is the number of layers, and V_a is the applied voltage. Combining Eqs. (8.1) and (8.2) yields an expression for developed force as a function of applied voltage.

$$F_a = d_{33} n k_a V_a. \tag{8.3}$$

The force equation can also be derived from the stress-charge form of the piezoelectric constituent equations (IEEE 1988)

$$T = d_{33} c^E E, \tag{8.4}$$

where T is the stress (in N/m^2), c^E is Young's elastic modulus under constant electric field (in N/m^2) and E is the applied electric field (in V/m). The developed force F_a is proportional to stress T and the surface area A (in m^2) by $F_a = TA$. Also, the electric field is equal to the applied voltage V_a divided by the layer thickness t, i.e., $E = V_a/t$. Taking this into account, the developed force is

$$F_a = \frac{d_{33} c^E A V_a}{t}. \tag{8.5}$$

This can be simplified by recognizing that the number of layers n is equal to the length L divided by layer thickness t, i.e., $n = L/t$. The elasticity c^E can also be replaced by stiffness, which is related to elasticity by

$$k_a = \frac{c^E A}{L}. \tag{8.6}$$

The resulting expression for developed force is again

$$F_a = d_{33} n k_a V_a. \tag{8.7}$$

That is, the ratio of developed force to applied voltage is $d_{33} n k_a$ N/V. In following sections, this constant will be denoted g_a where

$$F_a = g_a V_a \text{ and } g_a = d_{33} n k_a.$$

Compared to standard modeling techniques (Adriaens et al. 2000), which are based on displacement, the above method results in an expression for generated force. This approach provides an intuitive understanding of the actuator mechanics and significantly simplifies the modeling of interconnected structures as the generated actuator force is independent of load force and stiffness. The ease of combining the actuator and structural models when using developed force rather than displacement will become clear in Sect. 8.2.4.

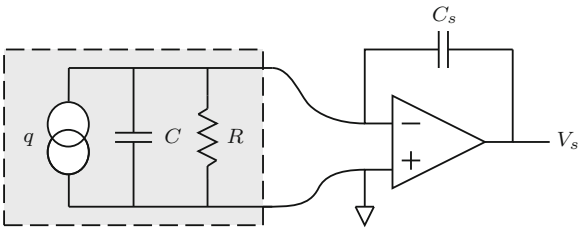


Fig. 8.3 The electrical model of a piezoelectric force sensor is shown in *gray*. The developed charge q is proportional to the strain and hence force experienced by the sensor. The op-amp charge amplifier produces an output voltage V_s equal to $-q/C_s$

8.2.2 Sensor Dynamics

Although the load force F_s can be measured in a number of ways, in this application it is desirable to minimize the additional mass and compliance associated with the sensor. In such scenarios, piezoelectric transducers are an excellent choice. They provide high sensitivity and bandwidth with low-noise at high frequencies.

If a single wafer of piezoelectric material is sandwiched between the actuator and platform, the amount of generated charge per unit area D (in C/m^2) is given by the standard strain-charge form of the piezoelectric constituent equations (IEEE 1988)

$$D = d_{33} T. \tag{8.8}$$

The generated charge is then

$$q = d_{33} F_s. \tag{8.9}$$

If an n-layer piezoelectric transducer is used as a force sensor, the generated charge is

$$q = n d_{33} F_s. \tag{8.10}$$

The electrical model of a piezoelectric force sensor and charge measurement circuit is shown in Fig. 8.3. In this circuit, the output voltage V_s is equal to

$$V_s = -\frac{q}{C_s} = -\frac{n d_{33} F_s}{C_s}, \tag{8.11}$$

that is, the scaling between force and voltage is $-\frac{n d_{33}}{C_s}$ V/N.

Piezoelectric force sensors can also be calibrated using voltage rather than charge measurement. In this case the generated charge is deposited on the transducer's internal capacitance. As the terminal voltage is non-zero, the dynamics of the sensor are slightly altered. In effect, the transducer is marginally stiffened (Liu et al. 2007). However, as the stiffness of the sensor is already substantially greater than that of

the actuator and flexures, this effect is negligible. The open-circuit voltage of a piezoelectric force sensor is

$$V_s = \frac{nd_{33}F_s}{C}, \tag{8.12}$$

where C is the transducer capacitance defined by $C = n\epsilon_T A/h$ and A, h and ϵ_T are the area, thickness and dielectric permittivity under constant stress. The scaling factor between force and measured voltage is $\frac{nd_{33}}{C}$ V/N. In following sections, this sensor constant will be denoted g_s, i.e.,

$$V_s = g_s F_s, \quad \text{and } g_s = \frac{nd_{33}}{C}. \tag{8.13}$$

8.2.3 Sensor Noise

Due to the high mechanical stiffness of piezoelectric force sensors, thermal or Boltzmann noise is negligible compared to the electrical noise arising from interface electronics. As piezoelectric sensors have a capacitive source impedance, the sensor noise density $N_{Vs}(\omega)$ is due primarily to current noise i_n reacting with the capacitive source impedance, i.e.,

$$N_{Vs}(\omega) = i_n \frac{1}{C\omega}, \tag{8.14}$$

where N_{Vs} and i_n are the spectral densities, measured in V and A/$\sqrt{\text{Hz}}$ respectively. In Chap. 13, $N_{Vs}(\omega)$ would be denoted $\sqrt{S_{Vs}(\omega)}$ to indicate that it is a spectral density, not power spectral density. As all of the noise quantities in this chapter are spectral densities, the simplified notation will be used.

Note that the high-pass filter arising from the transducers leakage resistance has been ignored as this pole is approximately canceled by the 1/f corner frequency[1] of the current noise density i_n (Fleming et al. 2008).

In addition to noise, piezoelectric force sensors also exhibit other non-ideal characteristics. These include temperature dependence and a small amount of non-linearity. A thorough treatment of these topics is beyond the scope of this chapter. However, if such characteristics must be avoided, dedicated piezoelectric sensor compositions are available with extremely high linearity and essentially no temperature dependence, e.g., Quartz or Gallium Phosphate.

[1] The power spectral density of an electronic device is approximately constant above the 1/f corner frequency, while below this frequency, it is approximately proportional to the inverse of frequency (Horowitz and Hill 1989).

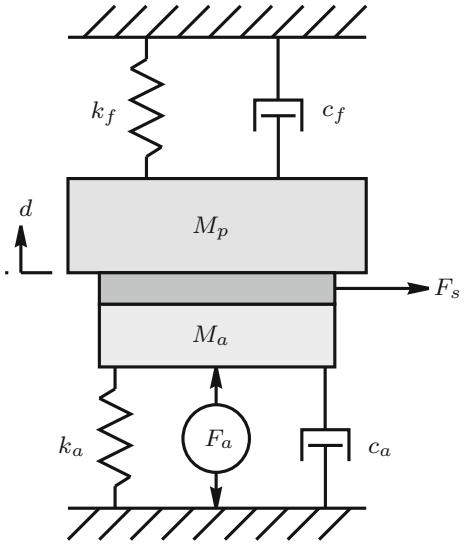


Fig. 8.4 Mechanical diagram of a single-degree-of-freedom positioning stage. F_s is the measured force acting between the actuator and platform mass in the vertical direction

8.2.4 Mechanical Dynamics

The mechanical diagram of a single axis positioner is shown in Fig. 8.4. The developed actuator force F_a results in a load force F_s and platform displacement d. The stiffness and damping coefficient of the flexures and actuator are denoted k_f, c_f, and k_a, c_a respectively.

The dynamics of the suspended platform are governed by Newton's second law,

$$(M_a + M_p)\ddot{d} = F_a - k_a d - k_f d - c_a \dot{d} - c_f \dot{d}, \tag{8.15}$$

where M_a, and M_p are the effective mass of the actuator and mass of the platform. As the actuator and flexure are mechanically in parallel with the suspended platform, the masses, stiffness and damping coefficients can be grouped together, that is

$$M = M_a + M_p, \tag{8.16}$$

$$k = k_a + k_f \text{ and} \tag{8.17}$$

$$c = c_a + c_{f.}$$

The equation of motion is then

$$M\ddot{d} + kd + c\dot{d} = F_a, \tag{8.18}$$

and the transfer function from actuator force F_a to platform displacement d is

$$\frac{d}{F_a} = \frac{1}{Ms^2 + cs + k}. \tag{8.19}$$

Including the actuator gain, the transfer function from applied voltage to displacement can be written

$$G_{dVa} = \frac{d}{V_a} = \frac{g_a}{Ms^2 + cs + k} \tag{8.20}$$

The load force F_s is also of interest, this can be related to the actuator force F_a by applying Newton's second law to the actuator mass,

$$M_a\ddot{d} = F_a - k_a d - c_a\dot{d} - F_s. \tag{8.21}$$

This results in the following transfer function between the applied force F_a and measured force F_s,

$$\frac{F_s}{F_a} = 1 - (M_a s^2 + c_a s + k_a)\frac{d}{F_a} \tag{8.22}$$

$$= \frac{M_p s^2 + c_f s + k_f}{Ms^2 + cs + k}. \tag{8.23}$$

By including the actuator and sensor gains g_a and g_s, the system transfer function from the applied voltage to measured voltage can be found,

$$G_{VsVa} = \frac{V_s}{V_a} = g_a g_s \frac{M_p s^2 + c_f s + k_f}{Ms^2 + cs + k}. \tag{8.24}$$

The two system transfer functions G_{dVa} and G_{VsVa}, will be used in the following sections to simulate the performance of feedback control systems. As both of these transfer functions have the same input V_a and poles, it is convenient to define a single-input two-output system G that contains both of these transfer functions,

$$G = \begin{bmatrix} G_{dVa} \\ G_{VsVa} \end{bmatrix}. \tag{8.25}$$

8.2.5 System Properties

The transfer function G_{VsVa} (8.24) can be rewritten

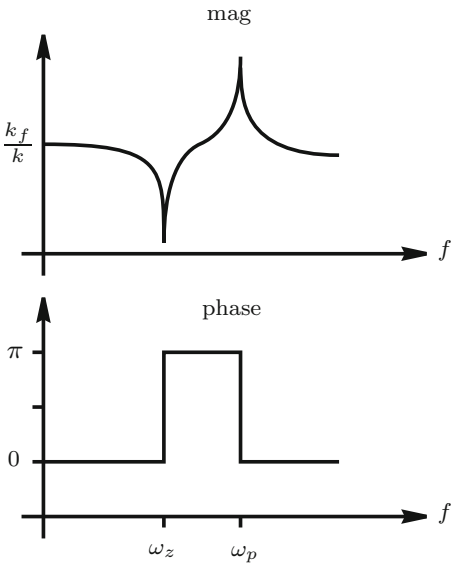


Fig. 8.5 Magnitude and phase response of F_s/F_a (8.22)

$$G_{VsVa} = g_a g_s \frac{M_p}{M} \frac{s^2 + \frac{c_f}{M_p}s + \frac{k_f}{M_p}}{s^2 + \frac{c}{M}s + \frac{k}{M}}. \tag{8.26}$$

This transfer function consists of a pair of resonant poles and zeros at frequencies ω_z and ω_p,

$$\omega_z = \sqrt{\frac{k_f}{M_p}},\ \omega_p = \sqrt{\frac{k}{M}} = \sqrt{\frac{k_a + k_f}{M_a + M_p}}.$$

In general, the resonance frequency of the zeros will appear below the poles. The condition for this to occur is:

$$\begin{aligned} \omega_z &< \omega_p \\ \frac{k_f}{M_p} &< \frac{k_a + k_f}{M_a + M_p} \\ M_a k_f &< k_a M_p. \end{aligned} \tag{8.27}$$

As the actuator mass M_a and flexural stiffness k_f are significantly lesser than the actuator stiffness k_a and platform mass M_p, the resonant zeros will always occur below the resonance frequency of the poles. This characteristic is shown in the frequency response of F_s/F_a in Fig. 8.5.

Table 8.1 Example system parameters

Parameter	Symbol	Value
Platform mass	M_p	100 g
Actuator mass	M_a	2 g
Actuator area	A	5×5 mm
Actuator length	L	10 mm
Young's modulus	c^E	50 GPa
Charge constant	d_{33}	300×10^{-12} C/N
Actuator stiffness	k_a	125 N/μm
Flexure stiffness	k_f	50 N/μm
Actuator layers	n	200
Actuator damping	c_a	100 N/ms^{-1}
Flexure damping	c_f	100 N/ms^{-1}

8.2.6 Example System

For the sake of demonstration and to assess the validity of assumptions in the following sections, an example system will be considered. The system is a single dimensional positioning stage as illustrated in Figs. 8.1 and 8.4. The actuator is a 10 mm long PZT linear actuator with 200 layers. Force sensing is provided by a single PZT wafer of the same area. The dimensions and physical properties of the system are listed in Table 8.1.

The actuator and sensor gains are

$$g_a = 7.5 \text{ N/V, and } g_s = 0.19 \text{ V/N,} \tag{8.28}$$

which results in an open-loop static displacement sensitivity $G_{dVa}(0)$ of

$$G_{dVa}(0) = \frac{g_a}{k} = 43 \text{ nm/V.} \tag{8.29}$$

The full scale displacement is 8.5 μm at 200 V and the system resonance frequencies are

$$\omega_p = 6.3 \text{ kHz, and } \omega_z = 3.6 \text{ kHz.} \tag{8.30}$$

The open-loop frequency response is plotted in Fig. 8.8.

8.3 Damping Control

The technique of Integral Force Feedback (IFF) has been widely applied for augmenting the damping of flexible structures (Preumont et al. 1992; Preumont 2006; Preumont et al. 2007). The feedback law is simple to implement and under common

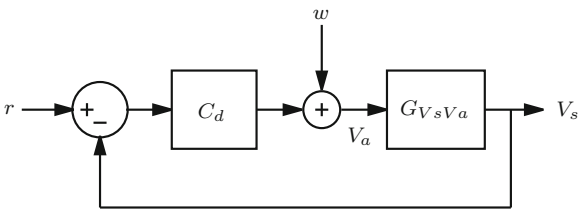


Fig. 8.6 A nanopositioning system G_{VsVa}, with input and output voltages V_a and V_s proportional to applied and measured force, controlled by an integral force feedback (IFF) damping controller $C_d(s)$

circumstances, provides excellent damping performance with guaranteed stability (Preumont 2006). In the following, IFF is applied to augment the damping of nanopositioning systems.

The feedback diagram of an IFF damping controller is shown in Fig. 8.6.

A key observation of the system G_{VsVa} is that its phase response lies between 0 and 180°. This is a general feature of flexible structures with inputs and outputs proportional to applied and measured force (Preumont 2006). A unique property of such systems is that integral control can be directly applied to achieve damping, i.e.,

$$C_d(s) = \frac{\alpha}{s} \tag{8.31}$$

where α is the controller gain. As the integral controller has a constant phase lag of 90°, the loop-gain phase lies between −90 and 90°. That is, the closed-loop system has an infinite gain-margin and phase-margin of 90°. Simplicity and robustness are two outstanding properties of systems with IFF.

A solution for the optimal feedback gain α has already been derived in reference Preumont (2006). These results can be directly adapted for the system considered in this work. The method makes the valid assumption that system damping coefficients are small and can be neglected. A further valid simplification is that the actuator mass M_a is negligible compared to the platform mass M_p. With these assumptions, the optimal feedback gain $\alpha^\star$ and corresponding maximum closed-loop damping ratio $\xi^\star$ are

$$\alpha^\star = \frac{\omega_p\sqrt{\omega_p/\omega_z}}{g_s g_a}, \text{ and} \tag{8.32}$$

$$\xi^\star = \frac{\omega_p - \omega_z}{2\omega_z} \tag{8.33}$$

An expression for the closed-loop poles can also be adapted from Preumont (2006). The closed-loop poles are given by the roots of the following equation

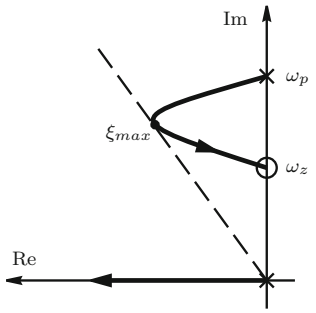


Fig. 8.7 Root-locus of a nanopositioning system G_{VsVa} with integral damping controller C_d

$$1 + \alpha g_s g_a \frac{s^2 + \omega_z^2}{s(s^2 + \omega_p^2)} = 0. \tag{8.34}$$

The corresponding closed-loop root-locus is plotted in Fig. 8.7 (Preumont 2006). Note that the closed-loop poles remain in the left half plane and that the system is unconditionally stable. The root-locus also provides a straight-forward method for finding the optimal feedback gain numerically. This can be useful if the model parameters are unknown, i.e., if the system G_{VsVa} was procured directly from experimental data by system identification. This approach is taken in Sect. 8.5.

For the example system described in Sect. 8.2.6, the optimal gain and maximum damping ratio are computed from Eqs. (8.32) and (8.33), the result is

$$\alpha^\star = 4.0 \times 10^4, \text{ and } \xi^\star = 0.43. \tag{8.35}$$

These values can be checked with a numerical root-locus plot. The numerically optimal gain is 4.07×10^4 which provides a closed-loop damping ratio of 0.45. This correlates closely with the predicted values and supports the accuracy of the assumptions made in deriving the optimal gain.

The simulated open- and closed-loop frequency responses from the disturbance input w to the measured sensor voltage V_s are plotted in Fig. 8.8. Clearly the controller significantly improves the system damping and disturbance rejection at low frequencies.

8.4 Tracking Control

After studying the relationship between force and displacement in the following subsection, three different tracking controller architectures will be discussed.

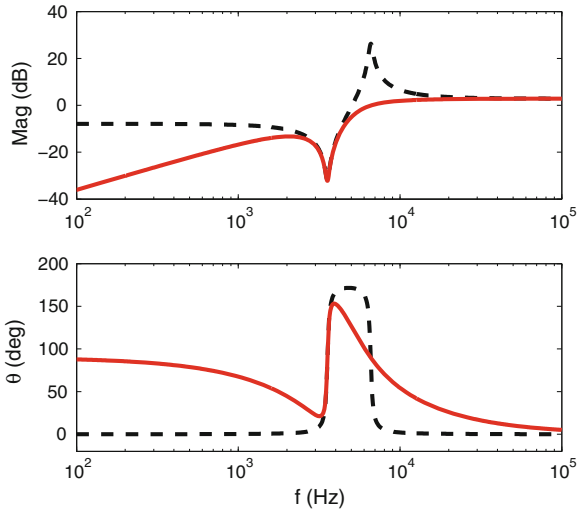


Fig. 8.8 Open-loop (*dashed line*) and closed-loop (*solid line*) frequency response from w to V_s

8.4.1 Relationship Between Force and Displacement

The relationship between measured force and displacement can be found either by applying Newton's second law to the platform mass or by multiplying the two system transfer functions (8.19) and (8.22), i.e.,

$$\frac{d}{F_s} = \frac{d}{F_a}\left(\frac{F_s}{F_a}\right)^{-1} \tag{8.36}$$

$$\frac{d}{F_s} = \frac{1}{M_p s^2 + c_f s + k_f}. \tag{8.37}$$

Thus, the measured voltage V_s is related to displacement by

$$\frac{d}{V_s} = \frac{d}{g_s F_s} = \frac{1/g_s}{M_p s^2 + c_f s + k_f} \tag{8.38}$$

From the transfer function d/V_s (8.38), it can be observed that displacement is proportional to force up until the frequency of the system zeros, $\omega_z = \sqrt{k_f/M_p}$. The scaling factor is $g_{cl} = 1/g_s k_f$ m/V. That is,

$$d \approx g_{cl} V_s = \frac{1}{g_s k_f} V_s \text{ for } \omega < \omega_z. \tag{8.39}$$

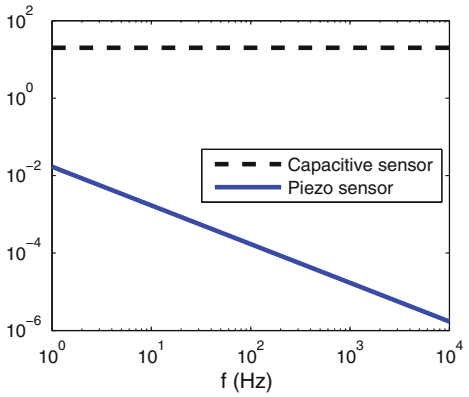


Fig. 8.9 The simulated position noise spectral density (in pm/$\sqrt{\text{Hz}}$) of a state-of-the-art capacitive sensor and the piezoelectric force sensor described in Sect. 8.2.6

Above ω_z, the measured force and voltage is proportional to platform acceleration. The scaling factor is $1/g_s M_p$ m/s/V. That is

$$ds^2 \approx \frac{1}{g_s M_p} V_s \text{ for } \omega > \omega_z. \tag{8.40}$$

As V_s is directly proportional to displacement at frequencies below ω_z, it makes an excellent feedback variable when trajectory tracking is required.

A key benefit of using the piezoelectric force sensor is its extremely low noise density. The approximate position noise density $N(\omega)$ can be found by combining Eqs. (8.14) and (8.39),

$$N(\omega) = i_n \frac{1}{C\omega} \frac{1}{g_s k_f}, \tag{8.41}$$

where i_n is the current noise density of the interface electronics and C is the sensor capacitance. The position noise density of the example system is compared to the noise density of a state-of-the-art capacitive sensor (20 pm/$\sqrt{\text{Hz}}$) in Fig. 8.9. The plot demonstrates the extremely low position noise of the piezoelectric sensor. This simulation uses the current noise density from a general purpose LM833 FET-input op-amp, which is 0.5 pA/$\sqrt{\text{Hz}}$.

In following Sections, $N_{Vs}(\omega)$ and $N_d(\omega)$ will be used to represent the additive sensor noise exhibited by the piezoelectric voltage measurement and capacitive displacement sensor.

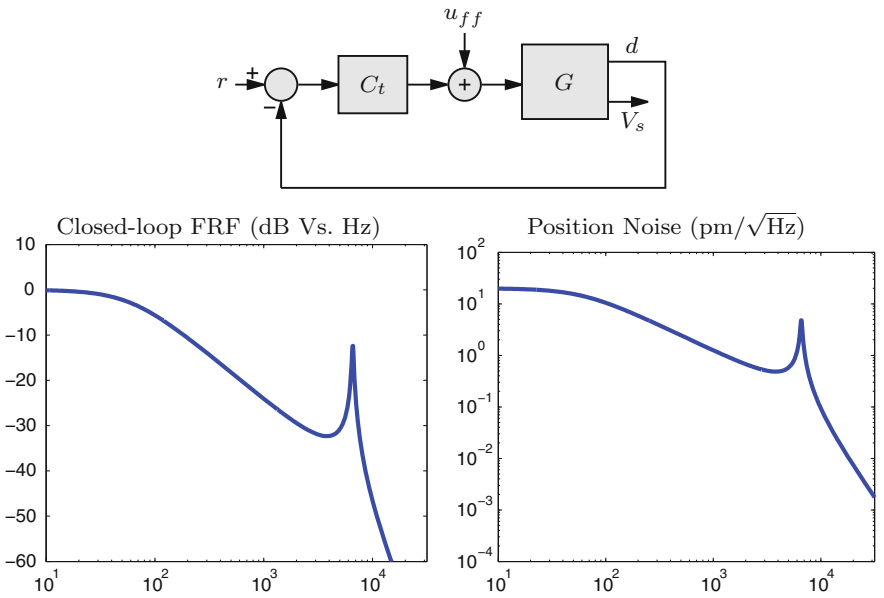


Fig. 8.10 Basic integral control has the benefit of being simple and highly linear due to displacement feedback, however it is also very slow (60 Hz bandwidth) and has low gain margin (5 dB)

8.4.2 Integral Displacement Feedback

The most straight-forward technique for achieving displacement tracking is to simply enclose the system in an integral feedback loop, as pictured in Fig. 8.10. The tracking controller C_t is simply

$$C_t = \frac{\beta}{s}. \tag{8.42}$$

In this strategy, the displacement d must be obtained with a physical displacement sensor such as a capacitive, inductive or optical sensor, see Chap. 5.

As discussed in the Sect. 1.5.1, the foremost limitation of integral tracking controllers is the low gain-margin. For the example system, the bandwidth is limited to only 60-Hz with a 5-dB gain-margin. The gain-margin is also highly sensitive to variations in resonance frequency.

8.4.3 Direct Tracking Control

The low bandwidth of integral tracking controllers can be significantly improved by adding an internal force feedback loop as shown in Fig. 8.11. As the damping controller eliminates the lightly damped resonance, gain-margin is drastically

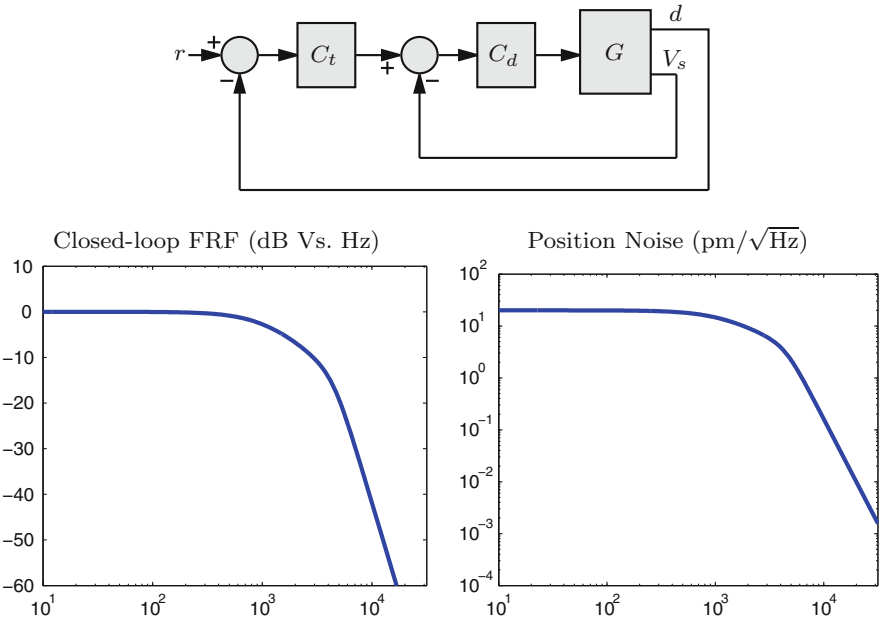


Fig. 8.11 Direct tracking control is faster than integral control (1 kHz bandwidth) and also highly linear due to displacement feedback however, it is also noisy and still limited by gain margin (5 dB)

increased, allowing a proportional increase in tracking bandwidth. This was discussed in Sect. 1.5.1.

To find the closed-loop transfer function, it is first convenient to find the transfer function of the internal loop. That is, the transfer function $\widehat{G_{du}}$ from u to d, this is

$$\widehat{G_{du}} = \frac{G_{dVa}C_d}{1 + C_d G_{VsVa}}. \tag{8.43}$$

The closed-loop response $\widehat{G_{dr}}$ from r to d is then

$$\widehat{G_{dr}} = \frac{C_t\widehat{G_{du}}}{1 + C_t\widehat{G_{du}}}, \tag{8.44}$$

or equivalently,

$$\widehat{G_{dr}} = \frac{G_{dVa}C_tC_d}{1 + G_{dVa}C_tC_d + C_d G_{VsVa}}. \tag{8.45}$$

The frequency response of this transfer function is plotted in Fig. 8.11. Compared to the integral controller with the same gain-margin (5 dB), the bandwidth has been increased from 60 Hz to 1 kHz. Although this is an excellent improvement, the

gain-margin is still sensitive to changes in resonance frequency. In practice, the controller needs to be conservatively designed for stability with the lowest possible resonance frequency.

One disadvantage of increasing closed-loop bandwidth is that position noise is increased. This is illustrated by the wider bandwidth power spectral density plotted in Fig. 8.11. The closed-loop power spectral density $\widehat{N_d}(\omega)$ is obtained from the density of additive sensor noises, $N_d(\omega)$ and $N_{Vs}(\omega)$, and the noise sensitivity of the control loop. As the piezoelectric sensor noise $N_{Vs}(\omega)$ is negligible compared to $N_d(\omega)$, $\widehat{N_d}(\omega)$ can be approximated by

$$\widehat{N_d}(\omega) = \left| \frac{-G_{dVa} C_t C_d}{1 + G_{dVa} C_t C_d + C_d G_{VsVa}} \right| N_d(\omega). \tag{8.46}$$

8.4.4 Dual Sensor Feedback

In the Sect. 8.4.1 it was found that measured force is proportional to displacement at frequencies below the system zeros. A logical progression is to simply apply a reference input r to the force feedback loop and expect displacement tracking at frequencies from DC to ω_z. Unfortunately this is not possible due to the high-pass filter formed by the piezoelectric capacitance and finite input impedance of charge amplifiers and voltage buffers. The measured voltage across a piezoelectric sensor is equal to

$$V_s = V_p \frac{s}{s + 1/R_{\text{in}}C} \tag{8.47}$$

where V_p is the piezoelectric strain voltage, R_{in} is the voltage buffer input impedance and C is the transducer capacitance. The filter is high-pass with a cut-off frequency of $1/R_{\text{in}}C$.

Although the high-pass cut-off frequency can be made extremely low, in the order of 1 mHz, this is not desirable as the settling time becomes extremely long. A preferable solution is to use the displacement measurement d at low frequencies where the piezoelectric force sensor is inaccurate.

The diagram of a dual sensor control loop is contained in Fig. 8.12. This tracking control loop is similar to Fig. 8.6 except for the additional complementary filters F_H and F_L. These complementary filters substitute the displacement measurement d for V_s at frequencies below the crossover frequency ω_c, which in this study is 10 Hz. The simplest choice of complementary filters are

$$F_H = \frac{s}{s + \omega_c}, \text{ and } F_L = \frac{\omega_c}{s + \omega_c}. \tag{8.48}$$

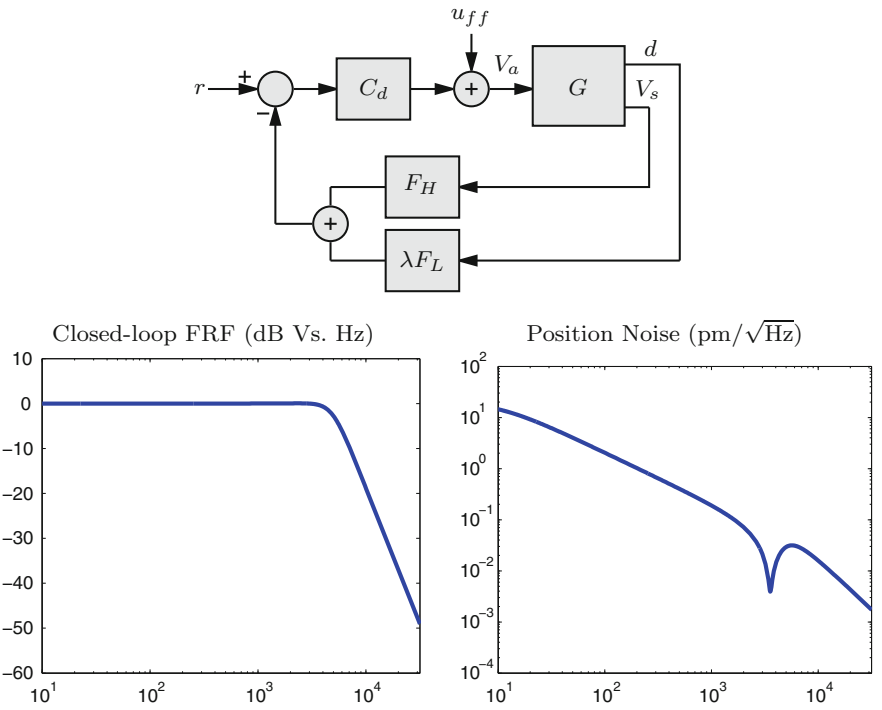


Fig. 8.12 Dual sensor feedback is faster again than direct tracking control (5.1 kHz bandwidth). It also provides unconditional stability and good linearity at low frequencies where displacement feedback is dominant. Although the noise performance is better than direct tracking control, there is still sensor induced noise present at low frequencies

As the measured displacement signal d will have a different sensitivity than V_s, it must be scaled by an equalizing constant λ, as shown in the diagram. The value of λ should be

$$\lambda = \frac{G_{VsVa}(0)}{G_{dVa}(0)} \tag{8.49}$$

If λ is chosen correctly, the closed-loop response $\widehat{G_{dr}}$ is

$$\widehat{G_{dr}} = \frac{G_{dVa}C_d}{1 + C_d G_{VsVa}}. \tag{8.50}$$

As this control loop is unconditionally stable, there is no restriction on the gain of C_d. However, C_d was chosen in the previous section to provide optimal damping performance, this value should be retained. Further increases in C_d are not productive as the disturbance rejection at the resonance frequency will degrade.

The higher gain of the force-feedback loop provides an increase in bandwidth from 1 to 5.1 kHz compared to the direct tracking controller discussed in the previous

subsection. This increase also comes with theoretically infinite gain-margin and 90° phase margin, both of which are immune to variations in resonance frequency.

The closed-loop position noise density of the dual sensor controller is given by

$$\widehat{N_d}(\omega) = \left| \frac{-F_L G_{VsVa} C_d}{1 + G_{VsVa} C_d} \right| N_d(\omega). \tag{8.51}$$

Analogous to the direct tracking controller, position noise due to the piezoelectric force sensor is negligible and can be neglected. As the displacement sensor noise is now filtered by F_L, a significant improvement in noise performance is achieved. This is plotted in Fig. 8.12.

Although physical displacement sensors are much noisier than piezoelectric transducers, they also have better linearity and lower drift (Fleming et al. 2008). The complementary filters F_H and F_L exploit the best aspects of each signal. The wide-bandwidth and low noise of piezoelectric force sensors is exploited above the crossover frequency ω_c, while the physical displacement sensors provide a high level of thermal stability at DC and below the crossover frequency ω_c.

8.4.5 Low Frequency Bypass

If a physical displacement sensor is not available, or the system does not require a high level of DC accuracy, the low frequency displacement can be estimated from the input voltage V_a as shown in Fig. 8.13. This scheme can be viewed as a simple first-order observer that estimates DC position. The signal V_a requires the same sensitivity as V_s so the scaling constant λ is

$$\lambda = G_{VsVa}(0). \tag{8.52}$$

If λ is chosen correctly, the closed-loop response and stability characteristics are the same as that discussed in the previous subsection. The foremost benefit of eliminating the physical displacement sensor is noise reduction. The closed-loop position noise density, plotted in Fig. 8.13, is now

$$\widehat{N_d}(\omega) = \left| \frac{-F_H G_{dVa} C_d}{1 + G_{VsVa} C_d} \right| N_{Vs}(\omega), \tag{8.53}$$

which is orders of magnitude below the other controllers. The force feedback technique with low frequency bypass opens the possibility for nanopositioning systems with large range, wide bandwidth and subatomic resolution. These characteristics are demonstrated experimentally in the following section.

The major penalty from eliminating the physical displacement sensor is that linearity is now dependent only on the piezoelectric force sensor and flexural spring constant k_f, which is less reliable. There is also no control of creep. Although

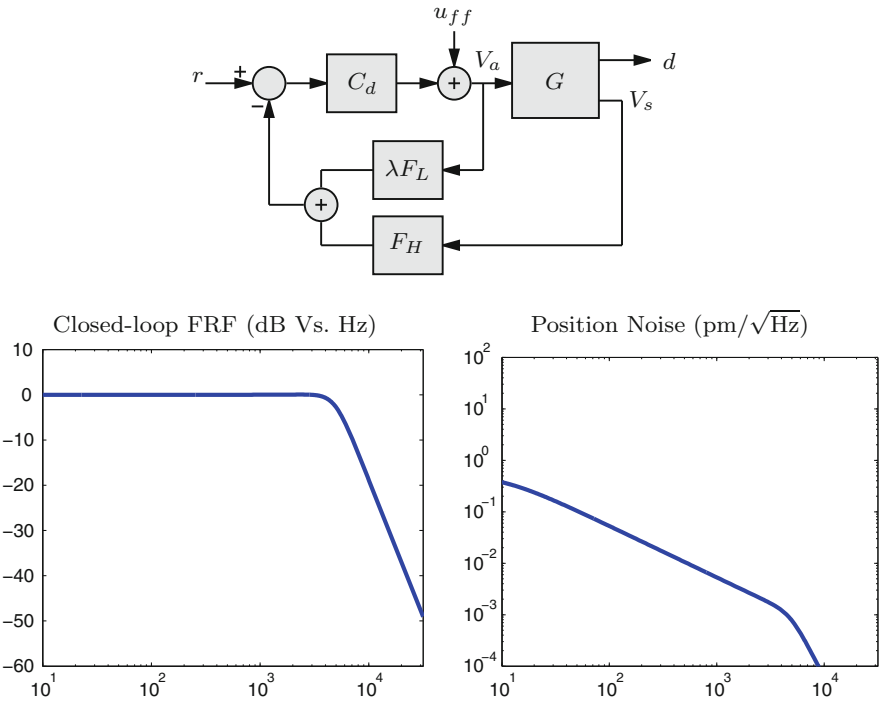


Fig. 8.13 Low frequency bypass provides the same bandwidth as dual-sensor feedback (5.1 kHz). It also provides unconditional stability and is straight-forward to implement. Although the closed-loop noise is extremely low, the absence of a displacement sensor results in the lack of control over low-frequency non-linearity such as creep

these drawbacks may preclude the use of this technique in some applications, other applications requiring subatomic resolution with wide bandwidth will benefit greatly, for example, video speed scanning probe microscopy (Ando et al. 2005; Schitter et al. 2007; Humphris et al. 2005; Rost et al. 2005).

8.4.6 Feedforward Inputs

The feedforward inputs u_{ff} shown in Figs. 8.10, 8.12 and 8.13 can be used to improve the closed-loop response of the system, see Chap. 9. Inversion based feedforward provides the best performance but the additional complexity is undesirable for the analog implementation considered in this work. A basic but effective form of feedforward compensation is to simply use the inverse DC gain of the system as a feedforward injection filter, i.e.,

$$u_{ff} = k_{ff} r. \tag{8.54}$$

This is easily implemented and can provide a reduction in tracking lag.

With a feedforward input, the closed-loop transfer function of the dual-sensor and low-frequency bypass controller is

$$\widehat{G_{dr}} = \frac{k_{ff} G_{dVa} + G_{dVa} C_d}{1 + C_d G_{VsVa}}. \tag{8.55}$$

8.4.7 Higher-Order Modes

So far, only a single-degree-of-freedom system has been considered. Although this is appropriate for modelling the first resonance mode, it does not capture the higher-order modes that occur in distributed mechanical systems. However, such higher order modes are not problematic as they do not disturb the zero-pole ordering of the transfer function from applied actuator voltage to the measured force.

In reference Preumont et al. (2007) it is shown that the transfer function of a generalized mechanical system with a discrete piezoelectric transducer and collocated force sensor is guaranteed to exhibit zero-pole ordering. That is, the transfer function G_{VsVa} will always exhibit zero-pole ordering. As the zero-pole ordering of the system is guaranteed, it follows that the controller discussed in Sect. 8.3 will also guarantee the stability of systems with multiple modes. The zero-pole ordering of an experimental system with multiple modes, and its successful control using the proposed technique, is reported in the following section.

8.5 Experimental Results

8.5.1 Experimental Nanopositioner

In Chap. 4 a high-bandwidth lateral nanopositioning platform was designed for video speed scanning probe microscopy. This device, pictured in Fig. 8.14, is a serial kinematic device with two moving platforms both suspended by leaf flexures and driven directly by 10-mm stack actuators. The displacement is measured with an ADE Tech 2804 capacitive sensor.

The small stage in the center, designed for scan-rates up to 5 kHz, is sufficiently fast with a resonance frequency of 29 kHz. However, the larger stage which provides motion in the adjacent axis is limited by a resonance frequency of 1.5 kHz. As this stage is required to operate with triangular trajectories of up to 100 Hz, active control is required.

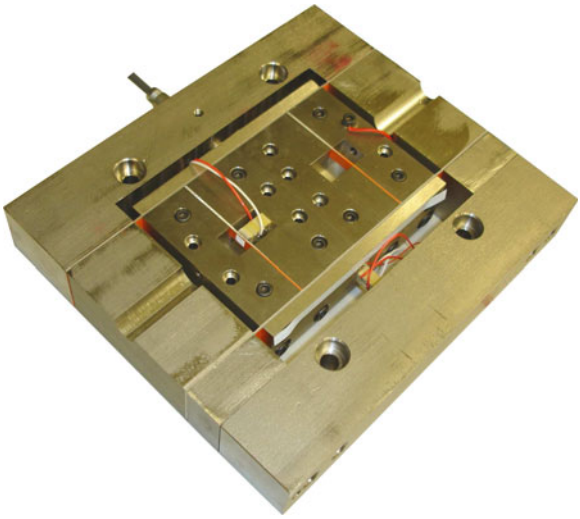


Fig. 8.14 High-speed nanopositioning platform described in Leang and Fleming (2009)

The main application for this nanopositioning device is high speed scanning probe microscopy. In this application, high-resolution and wide bandwidth are the most desirable characteristics. The force-feedback technique with low-frequency bypass, as discussed in Sect. 8.4.5, is the most suitable technique and will be applied here.

The platform under consideration is mechanically similar to the system in Fig. 8.1. The major difference is the existence of higher frequency modes beyond the first resonance frequency. These can be observed in the open-loop frequency response plotted in Fig. 8.17a. Although only a single mode system was previously discussed, the existence of higher order modes is not problematic. The zero-pole ordering and stability properties hold regardless of system order. This topic was discussed in detail in Sect. 8.4.7.

8.5.2 Actuators and Force Sensors

As discussed in Sect. 8.2.2, both piezoelectric plate and stack sensors can be used to measure force. A piezoelectric plate sensor is pictured in Fig. 8.15a. Also shown in Fig. 8.15b is a 10 mm Noliac SCMAP07 actuator connected to a 2 mm Noliac CMAP06 stack force sensor. The metal half-ball is used to eliminate the transmission of torsion and bending moments to the force sensor and moving platform.

For high-speed nanopositioning applications, the force sensor can also be integrated into the actuator. Such an arrangement is pictured in Fig. 8.15c. The actuator is a standard 10 mm Noliac SCMAP07 stack actuator with one of the four internal actuators wired independently for use as a sensor.

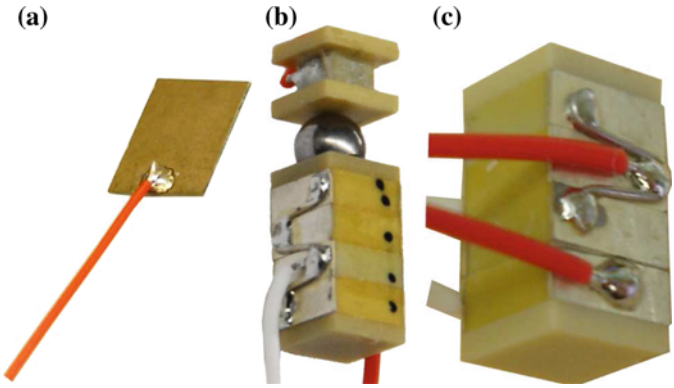


Fig. 8.15 Three types of piezoelectric force sensor, **a** a plate force sensor, **b** a stack actuator with discrete force sensor, and **c** a stack actuator with integrated force sensor

Fig. 8.16 Piezodrive PDL200 voltage amplifier used to drive the actuator

Although integrated sensors are convenient and provide the highest mechanical stiffness, they also have an associated disadvantage. In addition to measuring the applied load force, an integrated sensor also detects contraction of the actuator due to Poisson Coupling as the actuator elongates. This contraction is coupled to the sensor and results in a small additive voltage opposite in polarity to the voltage induced by the load force. This error is small in systems where the flexural stiffness is appropriately matched to the stiffness of the actuator. In nanopositioners with poorly matched actuators, i.e., where the flexural stiffness is much lesser than the actuator stiffness, the error due to Poisson Coupling can be significant. In such cases however, the error can be eliminated using the arrangement shown in Fig. 8.15b.

In the following experiments, the actuator with integrated sensor is utilized. The integrated sensor simplifies the stage assembly and provides the highest mechanical stiffness.

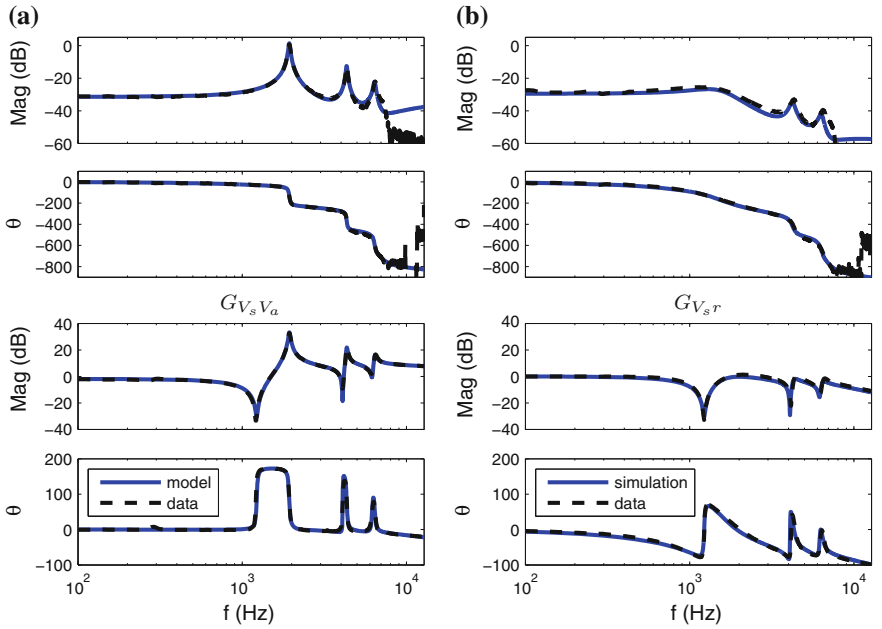


Fig. 8.17 The open- (**a**) and closed-loop (**b**) frequency responses of the nanopositioning system

The actuator was driven with a Piezodrive PDL200 linear amplifier pictured in Fig. 8.16. With the 250 nF load capacitance the PDL200 provides a bandwidth of approximately 30 kHz.

8.5.3 Control Design

To facilitate analysis of the control loop, a model was procured using the frequency domain subspace technique[2] (McKelvey et al. 1996). In Fig. 8.17a the response of a 7th order, single-input, two-output identified model can be verified to closely match the system response.

The optimal control gain was determined using the root-locus technique as $\beta = 7{,}800$. Together with the 1-Hz corner frequency complementary filters, the controller was implemented with an analog circuit. Due to the simplicity of the control loop, analog implementation is straight-forward and has the benefits of avoiding the quantization noise, finite resolution and sampling delay associated with digital controllers.

The closed-loop frequency response is plotted in Fig. 8.17b and reveals significant damping of the first three modes by 24, 9 and 4 dB. In addition to experimental data,

[2] A Matlab implementation of this algorithm is freely available by contacting the first author.

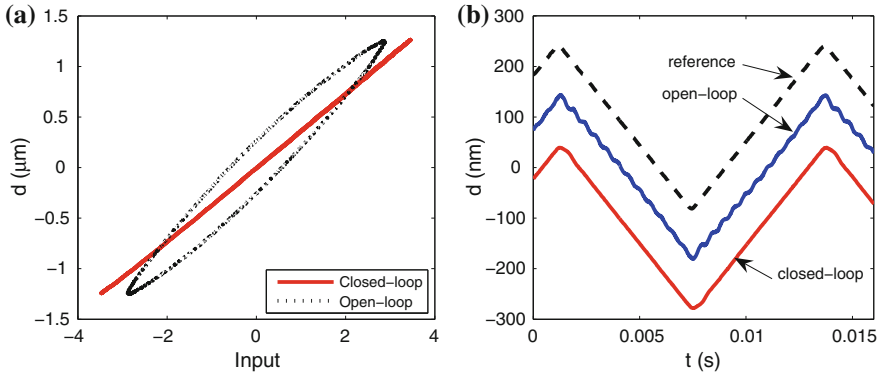


Fig. 8.18 The open- and closed-loop linearity (**a**) and response to an 80-Hz triangle wave (**b**). For the sake of clarity, the displacement curves in figure (**b**) have been offset from each other by 100 nm

the simulated response is also overlain which shows a close correlation. The tracking bandwidth of the closed-loop system is 2.07 kHz, which is higher than the open-loop resonance frequency and significantly greater than the bandwidth achievable with a direct tracking controller, predicted to be 210 Hz with a 5-dB gain-margin.

In Fig. 8.18a, the linearity of the system at 100 Hz is plotted. The large ellipse in the open-loop response is solely due to hysteresis as the system phase response at 100 Hz is negligible. Due to the high loop-gain of the force feedback controller, hysteresis is effectively eliminated, even at 100 Hz.

The time domain response of the closed-loop system to an 80 Hz triangular input is plotted in Fig. 8.18b. Due to the high loop-gain and resonance damping, the closed-loop response exhibits negligible induced vibration and minimal tracking lag.

8.5.4 Noise Performance

A major benefit associated with the piezoelectric force sensor is the extremely low additive noise. To quantify the noise, it was necessary to amplify the sensor output by 10^4 using a circuit of the authors own design. The resulting signal magnitude is then large enough to analyze with an HP-35670A spectrum analyzer. Due to the stochastic nature of the signal, 1,000 FFT averages were required to reduce the measurement variance to an acceptable level. The extremely low noise voltage produced by the piezoelectric sensor also necessitates the quantification of amplifier and instrumentation noise. This noise floor, which sets the limit of detection, was found to be approximately $2\,\text{fm}/\sqrt{\text{Hz}}$ which guarantees the statistical validity of the following measurements.

The spectral densities of the capacitive and piezoelectric sensor noise, scaled to $\text{pm}/\sqrt{\text{Hz}}$ are compared in Fig. 8.19a. At high frequencies, where the impedance of

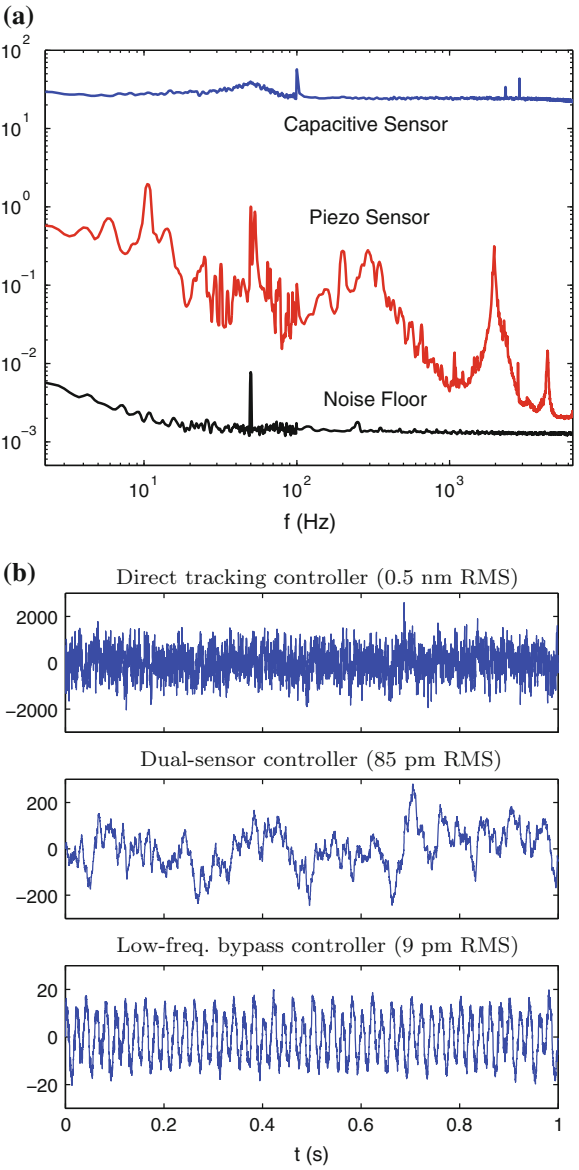


Fig. 8.19 **a** The spectral density of the capacitive sensor, piezo transducer and measuring instruments. **b** The closed-loop position noise of the controllers discussed in Sect. 8.4

the piezoelectric transducer is low, the sensor noise is up to four orders of magnitude lower than the capacitive sensor noise, which is relatively independent of frequency at approximately 26 pm/$\sqrt{\text{Hz}}$. At lower frequencies, the improvement is more modest

(see Sect. 8.4.1). However, even at 1 Hz, the piezoelectric sensor noise is only 2 % of the capacitive sensor's noise, which is 29 pm/$\sqrt{\text{Hz}}$ compared to 0.57 pm/$\sqrt{\text{Hz}}$. In the time domain, the RMS noise of the capacitive sensor is 1.7 nm compared to 9.5 pm for the piezoelectric sensor.

In truth, the piezoelectric sensor noise is even lower than that shown in Fig. 8.19a. The majority of measured noise power is actually due to external interference and mechanical excitation, not random noise. For example, the large peaks at 10 Hz and 2 kHz are due to mechanical and acoustic excitation of the mounting table and nanopositioner resonance. The large noise components at 50 Hz and between 150 and 500 Hz are also exogenous and most likely result from power-line frequency interference and harmonics arising from the use of fluorescent lighting. However, as these noise sources will likely be present in most practical applications, they are included in the following analysis.

The most intuitive method for evaluating closed-loop noise performance is to directly measure the sensor noise and simulate its effect on closed-loop position. The noise sensitivity transfer functions for the direct tracking controller, dual-sensor controller, and low-frequency bypass controller were discussed in Sects. 8.4.3, 8.4.4 and 8.4.5. Based on a 1 s measurement of the capacitive and piezoelectric sensor noise, the resulting closed-loop position noise for each controller is plotted in Fig. 8.19b. As expected, the direct tracking controller is the noisiest as it uses the capacitive sensor signal over its entire closed-loop bandwidth. The dual-sensor controller provides improved noise performance. However, the low-frequency bypass controller, which uses only the piezoelectric force sensor, has an exceptionally low closed-loop noise of only 9 pm RMS. The majority of this noise is clearly due to 50-Hz interference. If this interference were eliminated with comprehensive shielding, the closed-loop position noise could potentially be reduced to just a few picometers.

It should be noted that this analysis has considered only *sensor-induced* noise. That is, the positioning noise resulting from additive sensor noise. In practice, the magnitude of external disturbances will also have a significant impact on the overall positioning resolution, particularly if the sensor noise is reduced to the levels discussed here.

8.6 Chapter Summary

The bandwidth of nanopositioning systems can be significantly increased by damping the mechanical resonances. In previous chapters, this has been achieved with a shunt circuit or displacement sensor and feedback circuit. In this chapter, a force sensor was introduced between the actuator and moving platform. Compared to a standard position sensor, the force sensor is simple, low-cost, compact, and extremely sensitive. A major benefit is that the resulting system exhibits zero-pole ordering that allows a simple integral controller to achieve excellent damping performance and robustness.

In addition to damping control, the force sensor can also be used to estimate the platform displacement. This allows the damping controller to be adapted into an exceptionally high-performance tracking controller without sacrificing stability margins.

As with all piezoelectric sensors, the force sensor exhibits a high-pass characteristic at low-frequencies. This problem is solved by replacing the low-frequency force signal with a physical displacement measurement or displacement estimate based on the open-loop system dynamics.

Simulations on a nanopositioner model demonstrate the effectiveness of the tracking and damping controller. The dual-sensor IFF controller provides a closed-loop bandwidth approaching the open-loop resonance frequency while maintaining an infinite gain-margin and 90° phase-margin. By comparison, a standard integral displacement feedback controller achieves only 5 % of the bandwidth with a gain-margin of only 5 dB.

Experimental application to a high-speed nanopositioner demonstrates the performance and simplicity of force feedback. A bandwidth of 2.07 kHz was achieved from a system with a first resonance frequency of 1.5 kHz. This is an order of magnitude greater than a standard integral tracking controller with a gain margin of 5 dB.

Due to the extremely low noise of piezoelectric force sensors, the low-frequency bypass configuration was able to achieve a closed-loop positioning noise of 9 pm RMS with a full range of 10 μm.

References

Adriaens HJMTA, de Koning WL, Banning R (2000) Modeling piezoelectric actuators. IEEE/ASME Trans Mechatron 5(4):331–341

Ando T, Kodera N, Uchihashi T, Miyagi A, Nakakita R, Yamashita H, Matada K (2005) High-speed atomic force microscopy for capturing dynamic behavior of protein molecules at work. e-J Surf Sci Nanotechnol 3:384–392

Fleming AJ, Wills AG, Moheimani SOR (2008) Sensor fusion for improved control of piezoelectric tube scanners. IEEE Trans Cont Syst Technol 15(6):1265–6536

Horowitz P, Hill W (1989) The art of electronics. Cambridge University Press, Cambridge

Humphris ADL, Miles MJ, Hobbs JK (2005) A mechanical microscope: high-speed atomic force microscopy, Appl Phys Lett 86:034106-1–034106-3

Institute of Electrical and Electronics Engineers Inc. (1988) IEEE standard on piezoelectricity. In: ANSI/IEEE standard 176–1987

Leang KK, Fleming AJ (2009) High-speed serial-kinematic AFM scanner: design and drive considerations. Asian J Cont 11(2):144–153

Liu WQ, Feng ZH, Liu RB, Zhang J (2007) The influence of preamplifiers on the piezoelectric sensors dynamic property. Rev Sci Instrum 78(12):125107(1–4)

McKelvey T, Akcay H, Ljung L (1996) Subspace based multivariable system identification from frequency response data. IEEE Trans Autom Cont 41(7):960–978

Preumont A (2006) Mechatronics. Dynamics of electromechanical and piezoelectric systems. Springer, New York

Preumont A, de Marneffe B, Deraemaeker A, Bossens F (2007) The damping of a truss structure with a piezoelectric transducer. Comput Struct 86:227–239

Preumont A, Dufour JP, Malekian C (1992) Active damping by a local force feedback with piezoelectric actuators. AIAA J Guidance Cont 15(2):390–395

Rost MJ, Crama L, Schakel P, van Tol E, van Velzen-Williams GBEM, Overgauw CF, ter Horst H, Dekker H, Okhuijsen B, Seynen M, Vijftigschild A, Han P, Katan AJ, Schoots K, Schumm R, van Loo W, Oosterkamp TH, Frenken JWM (2005) Scanning probe microscopes go video rate and beyond. Rev Sci Instrum 76(5):053710-1–053710-9

Schitter G, Åström KJ, DeMartini BE, Thurner PJ, Turner KL, Hansma PK (2007) Design and modeling of a high-speed AFM-scanner. IEEE Trans Cont Syst Technol 15(5):906–915

Chapter 9
Feedforward Control

Unlike feedback control, which reacts to the measured tracking error, feedforward control compensates or anticipates for poor performance. A feedforward controller does this by exploiting some information about the system, and thus a well-designed feedforward controller requires sufficient knowledge of the plant dynamics and nonlinearities. In this case, the models are inverted to compensate for positioning errors due to dynamics and hysteresis in nanopositioning systems.

In this chapter, the feedforward control method is introduced. First, a method is described to compensate the effects of linear dynamics, such as induced-structural vibration and the creep effect in piezoactuators. Afterwards, feedforward control for nonlinear behavior such hysteresis is introduced. Experimental results for AFM positioning are presented to illustrate the application of feedforward control.

9.1 Why Feedforward?

Feedforward control is an open-loop approach as depicted in Fig. 9.1. As shown, an inverse model produces the feedforward input u_{ff} that is applied to the positioning system. The accuracy of feedforward control, for example how close the actual output y matches the desired output y_d, depends on the *quality* of the inverse model and whether external disturbances are present. Being an open-loop approach, feedforward control is subject to certain shortcomings, namely lack of robustness. However, for applications such as vibration compensation with a reasonably accurate model of the system dynamics, the advantages of feedforward control outweighs its disadvantages. In particular, feedforward control can provide high-bandwidth positioning, exceeding that of feedback-based methods. Also, feedforward control does not require continuous sensor feedback, and thus sensor-noise related issues can be avoided entirely.

To improve robustness, feedforward can be integrated with feedback control, as well to account for nonlinearity such as hysteresis (Leang and Devasia 2007).

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_9,

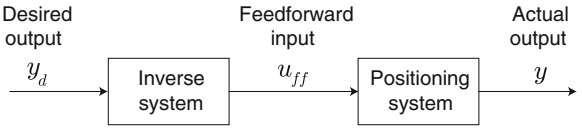


Fig. 9.1 The feedforward control concept

Integrated feedback and feedforward control also eliminates the need for modeling and inverting nonlinear behaviors. Such a task can be difficult and computationally demanding. If iteration is allowed, iterative feedforward techniques as discussed in Sect. 9.5 provides good performance with minimal modeling.

9.2 Modeling for Feedforward Control

As discussed in Chap. 2, Sects. 2.6 and 2.7, the input–output behavior of a nanopositioning system can be quite complex, consisting of structural dynamics and nonlinearities, such as hysteresis. A popular model that describes the dynamics and nonlinearity in a piezoactuator is the cascade model as depicted in Fig. 2.17 (Croft et al. 2001; Tan and Baras 2005), and repeated in Fig. 9.2a for convenience.

This model structure will be assumed and its form will be exploited for feedforward control. In particular, to find the feedforward input for precision output tracking, each submodel is inverted. More specifically, the feedforward control input u_{ff} is obtained by passing the desired output trajectory y_d through the inverse models of the dynamics and hysteresis in reverse order as illustrated in Fig. 9.2b.

The feedforward method is introduced below first to handle the vibrational dynamics and creep effect which are assumed to be linear behaviors.

9.3 Feedforward Control of Dynamics and Hysteresis

9.3.1 Simple DC-Gain Feedforward Control

At frequencies well below the dominant resonant peak, a simple feedforward input u_{ff} can be computed by scaling the desired output trajectory $y_d(t)$ by the inverse of the DC-gain $G(0)$ of the system $G(s)$. For example,

$$u_{ff}(t) = \frac{1}{G(0)} y_d(t). \tag{9.1}$$

Simple DC-gain feedforward is often used when a plant model is difficult and/or expensive to obtain. Also, if the frequency components of the desired trajectory are

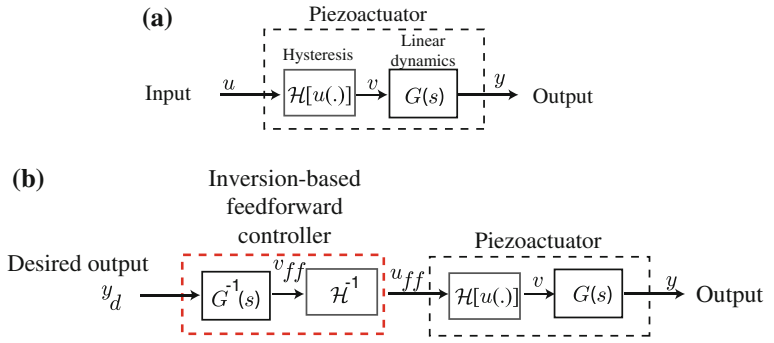


Fig. 9.2 **a** A cascade model structure for hysteresis $\mathcal{H}[u(\cdot)]$, vibrational dynamics, and creep effects $G(s)$ in piezoactuators. **b** An inversion-based feedforward approach to compensate for dynamic and hysteresis effects. The feedforward control input u_{ff} is obtained by passing the desired output trajectory y_d through the inverse models of the hysteresis and dynamics in reverse order

well below the dominant resonances, the need for inverting high-frequency dynamics for feedforward is less important. It is noted that simple DC-gain feedforward offers good performance at low frequency and over small displacement ranges, and also under the assumption that the DG-gain remains stable.

9.3.2 An Inversion-Based Feedforward Approach for Linear Dynamics

9.3.2.1 Frequency-Domain Approach

At higher frequencies, the effects of dynamics must be considered in the feedforward control input. In this case, the problem becomes inverting a dynamics model $G(s)$ to find the feedforward input $u_{ff}(t)$ over a specific frequency range. For piezoactuators, the dynamics include vibration and the creep effect (Croft et al. 2001).

Let $G(s)$ be a transfer function model that captures the linear dynamics of the piezoactuator. In the frequency domain, the feedforward input that accounts for the dynamics is given by

$$u_{ff}(j\omega) = G^{-1}(j\omega)y_d(j\omega), \tag{9.2}$$

where $G^{-1}(j\omega)$ is the inverse dynamics model.

In Eq. (9.2), the Fourier transform of the desired output trajectory $y_d(j\omega)$ and a plant model $G(j\omega)$ are needed to determine the feedforward input. If a measured frequency response function $G(j\omega)$ is available, for example measured by a dynamic signal analyzer, then the experimental data can be used directly to compute the

feedforward input $u_{ff}(j\omega)$. Then, the time-domain solution to Eq. (9.2) is found by taking the inverse Fourier transform of $u_{ff}(j\omega)$.

One key feature of this feedforward technique is that it can be applied to nonminimum-phase systems (Bayo 1987; Devasia et al. 1996; Zou and Devasia 1999). Although the dynamic effects are specifically addressed in this section, the approach can be combined with alternative feedforward or feedback methods that compensate for hysteresis when the range of motion becomes large (Leang and Devasia 2007).

9.3.2.2 Time-Domain Approach

The feedforward input $u_{ff}(t)$ can be computed directly in the time domain as follows. Consider the minimal state-space realization of $G(s)$, given by

$$\dot{x}(t) = Ax(t) + Bu(t), \tag{9.3}$$

$$y(t) = Cx(t), \tag{9.4}$$

where $x(t)$ is the state vector, $u(t)$ is the input, and $y(t)$ is the output, for example, the displacement along one lateral (x or y) axis. To simplify the presentation, the piezoactuator system is assumed to be single-input, single-output (SISO). This control method can be applied to multi-input, multi-output (MIMO) systems. To find the feedforward input $u_{ff}(t)$ that exactly tracks the desired output $y_d(t)$ of the system (9.3), (9.4), we differentiate the output Eq. (9.4) until the input appears explicitly in the expression. Hence,

$$y^{(r)}(t) = CA^r x(t) + CA^{r-1}Bu(t), \tag{9.5}$$

where $CA^{r-1}B \neq 0$, r is the relative degree of the system (9.3), (9.4), and the superscript "(r)" denotes the rth time derivative. For a SISO system, the relative degree r is the difference between the number of poles and zeros of $G(s)$. Thus, the inverse feedforward input $u_{ff}(t)$ that tracks the desired trajectory $y_d(t)$ can be obtained directly from (9.5) by replacing $y(t)$ with the desired output $y_d(t)$, that is,

$$u_{ff}(t) = (CA^{r-1}B)^{-1}\left[y_d^{(r)}(t) - CA^r x_{\text{ref}}(t)\right]. \tag{9.6}$$

The inverse feedforward input (9.6) shows that finding the inverse input $u_{ff}(t)$ is equivalent to finding the reference states $x_{\text{ref}}(t)$. In other words, a bounded solution for $x_{\text{ref}}(t)$ is needed.

Under a state transformation, a portion $\xi_d(t)$ of the reference states $x_{\text{ref}}(t)$ is specified by the desired output and its derivatives, up to $r-1$ derivatives. Thus, for a given desired trajectory, $\xi_d(t)$ is known. Then it remains to find the unknown reference states $\eta(t)$ to determine the feedforward input (9.6).

The unknown reference states $\eta(t)$ are found by solving the associated dynamics for a given desired output trajectory $y_d(t)$. The inverse input (9.6) is substituted back into (9.3), (9.4) and then rewritten in the transformed coordinate $[\xi_d,\ \eta]^T$. The unknown reference state equation becomes (Zou and Devasia 1999)

$$\dot{\eta}(t) = \hat{A}_\eta \eta(t) + \hat{B}_\eta \mathbf{Y}_d(t), \tag{9.7}$$

where $\mathbf{Y}_d(t)$ is the vector consisting of the desired output $y_d(t)$ and its derivatives up to the rth order. The details about $\hat{A}_\eta$ and $\hat{B}_\eta$ can be found in Zou and Devasia (1999). Equation (9.7) constitutes the *internal dynamics* of system (9.3), (9.4).

It can be shown, for example in Isidori (1995), that the poles of the internal dynamics (9.7) are exactly the zeros of (9.3). Therefore, if the system is nonminimum phase, then the internal dynamics (9.7) are unstable, and the goal is to find a bounded solution to the internal dynamics $\eta(t)$. This objective is addressed by the stable inversion theory (Devasia et al. 1996; Zou and Devasia 1999).

Stable inversion of unstable internal dynamics is based on the concept of noncausality. The internal dynamics (9.7) of a system that has no zeros on the imaginary axis can be decoupled into the stable σ_s and unstable σ_u dynamics through a state transformation, that is,

$$\dot{\sigma}_s = A_s \sigma_s(t) + B_s \mathbf{Y}_d(t), \tag{9.8}$$

$$\dot{\sigma}_u = A_u \sigma_u(t) + B_u \mathbf{Y}_d(t). \tag{9.9}$$

See Devasia (1997) for systems that have pure imaginary zeros.

The stable internal dynamics (9.8) are associated with the minimum-phase zeros, that is, the eigenvalues of A_s in (9.8) lie in the open left-half complex plane. Likewise, the unstable internal dynamics (9.9) are associated with the nonminimum-phase zeros, that is, the eigenvalues of A_u in (9.9) are on the open right-half complex plane. Then the stable solution to the unstable part of the internal dynamics can be solved by flowing the dynamics backwards in time,

$$\sigma_u(t) = -\int_t^\infty e^{\bar{A}_u(t-\tau)} \hat{B}_u \mathbf{Y}_d(\tau) d\tau. \tag{9.10}$$

Therefore, (9.10) implies that, to obtain the current value of the internal dynamics as well as the current value of the inverse input, the desired output trajectory must be specified in advance; thus, the stable inversion is noncausal. In many applications, such as the lateral scanning trajectory for AFM imaging, the desired trajectory is known a priori. For applications in which the desired trajectory is not completely known in advance, a preview-based stable inversion approach can be used (Zou and Devasia 1999, 2007). Basically, the preview-based approach computes the inverse input using the future desired trajectory within a finite time window. Finite preview of the desired trajectory is feasible in many applications. For example, in AFM-based

nanomanipulation and nanofabrication, it may be required to drive the AFM-probe to follow a real-time, user-specified trajectory. Therefore, finite preview of the future desired trajectory is available, and the preview-based inversion technique is applicable. In short, this technique tracks the user's motion with a delay time that equals the preview time. This delay is usually acceptable in nanomanipulation applications.

9.3.3 Frequency-Weighted Inversion: The Optimal Inverse

The inversion-based method presented above may yield excessively large inputs when the system has lightly damped system zeros. These large inputs can saturate the voltage amplifiers that drive the piezoactuator, or, even worse, depole the piezoactuator. Additionally, large model uncertainties around the resonant peaks or lightly damped zeros can cause significant error in computing the feedforward input. These model uncertainties thus lead to a lack of robustness when the inversion-based feedforward method is used. The following optimal inversion approach is used to account for these issues. Specifically, an optimal feedforward input is obtained by minimizing the quadratic cost function (Dewey et al. 1998)

$$J(u) = \int_{-\infty}^{\infty} \left\{ u^*(j\omega) R(j\omega) u(j\omega) + \left[x(j\omega) - x_d(j\omega) \right]^* Q(j\omega) \left[x(j\omega) - x_d(j\omega) \right] \right\} d\omega, \tag{9.11}$$

where '*' denotes the conjugate transpose, and $R(j\omega)$ and $Q(j\omega)$ are nonnegative, frequency-dependent real-valued weights on the input energy and the tracking error, respectively. The optimal feedforward input $u_{ff,opt}$ that minimizes (9.11) is

$$u_{ff,opt}(j\omega) = \left[\frac{G^*(j\omega) Q(j\omega)}{R(j\omega) + G^*(j\omega) Q(j\omega) G(j\omega)} \right] y_d(j\omega). \tag{9.12}$$

By choosing the frequency-dependent weights $R(j\omega)$ and $Q(j\omega)$, it is possible to systematically consider the effects of the input magnitude and the model uncertainties. For instance, the input energy weight $R(j\omega)$ can be chosen to be much larger than the tracking error weight $Q(j\omega)$ at frequencies where large model uncertainties exist or around lightly damped zeros. For details and implementation issues, see Zou (2008) and Zou and Devasia (2004).

9.3.4 Application to AFM Imaging

The inversion-based feedforward method is applied to AFM imaging to illustrate its application. The subject AFM system is described in Chap. 3. Specifically, the vibrational dynamics model $G_v(s)$ of the piezoscanner are inverted to find a feed-

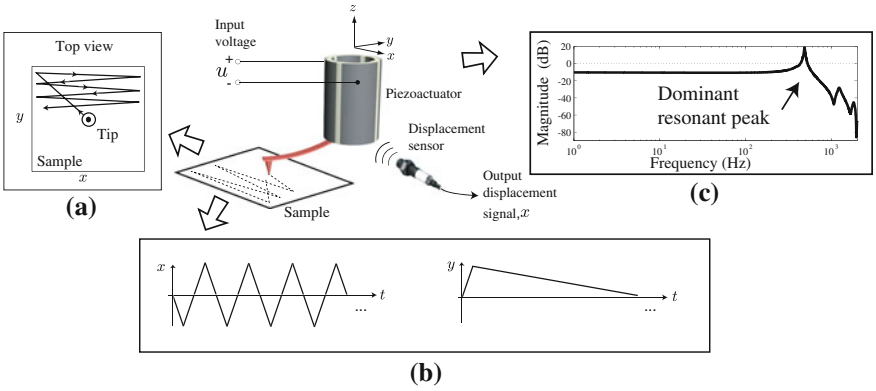


Fig. 9.3 The movement of the piezoactuator and cantilever in atomic force microscope (AFM) imaging. **a** A *top view* of the AFM scan path that shows the tip's path during imaging. The tip starts at the center of the sample and then moves to the *upper left-hand corner*. From the corner, the tip rasters back and forth across the sample in the x direction. At the same time the tip moves slowly in the y direction during imaging. **b** The lateral x and y scan paths versus time. The movement of the piezoactuator in the x direction is significantly faster than the movement in the y direction. **c** The frequency response of the piezoactuator dynamics in the x direction, where the input is the applied voltage u and the output is the displacement signal x. The frequency response shows a sharp resonant. The sharp resonant peak limits the open-loop operation of the AFM to low frequencies

forward input $u_{ff}(t)$ that tracks a given desired trajectory $x_d(t)$, that is, the desired trajectory along the fast-scanning x-axis. In the experiments, the range of motion is 10 μm, less than 5 % of the maximal range. Over this range, the hysteresis effect is negligible. The desired scan frequency is chosen greater than 1 Hz to avoid the creep effect in this first example. It is noted that vertical motion control is not considered here, but rather the focus is on controlling the lateral motion of the piezoactuator for AFM imaging.

The fast scanning axis in the x direction is at least 100 times faster than the motion in the y axis during AFM imaging. For instance, a 100×100 pixel image implies that the AFM probe rasters back and forth across the sample 100 times per image acquired (see Fig. 9.3 for the scan pattern). Therefore, the fast scanning motion in the x direction excites the mechanical resonances of the piezoactuator, causing the output to oscillate. The oscillations subsequently cause unwanted ripple-like distortion to appear in the AFM image.

To compensate for the dynamic effects, the inversion-based approach is used to determine a feedforward input to be applied to the piezoactuator. Figure 9.4 shows the feedforward control scheme, and the AFM imaging results over the small range for without feedforward compensation (left image) and with (right image) feedforward compensation. The inversion process for a prespecified desired trajectory is directly implemented in frequency domain using the fast Fourier transform (FFT) algorithm in Matlab (see Eq. 9.2). The left image shows ripples caused by the vibrational dynamics for a 30-Hz scan. Lightly colored vertical bands are evident of the vibration effects.

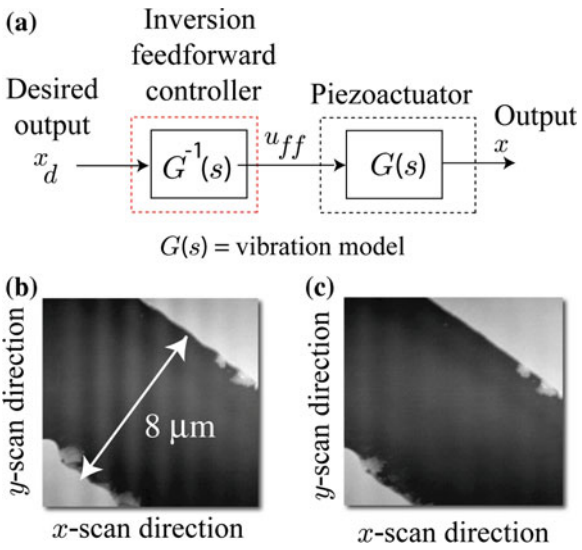


Fig. 9.4 Feedforward control of the linear vibrational dynamics to achieve high-speed positioning over small range. The block diagram in (**a**) shows the feedforward control scheme, where the linear vibrational dynamics model $G(s)$ is inverted to compensate for vibration effects. The atomic force microscope images are acquired without feedforward compensation in (**b**) and with feedforward compensation in (**c**). The feedforward input reduces the ripples caused by vibration

When the feedforward input u_{ff} is applied, the image shows significantly fewer ripples. In particular, the edges that separate the light and dark regions in the image show less oscillations in their appearance. Artifacts caused by minute particles on the sample's surface along the black/white edges and in the lower left-hand corner can be seen in both images.

At low frequency, the creep effect can be compensated for using the inversion-based approach. Creep causes the displacement of the piezoactuator in the AFM to slowly drift with time, especially when the scanning motion is offset from the nominal position. Figure 9.5a shows a 1 Hz scanning motion for the AFM piezoactuator with creep. The creep effect is modeled using spring-damper elements as described in Chap. 2 and the transfer function $G_{cx}(s)$ is given by Eq. (2.7). The creep effect is compensated for using the inversion-based approach and the result is shown in Fig. 9.5b.

9.4 Feedforward and Feedback Control

Modeling and inverting the dynamic and hysteresis effects are effective methods for precision positioning in piezoactuators in AFM (Croft et al. 2001; Zou and Devasia 2004). However, because the approach exploits knowledge of the

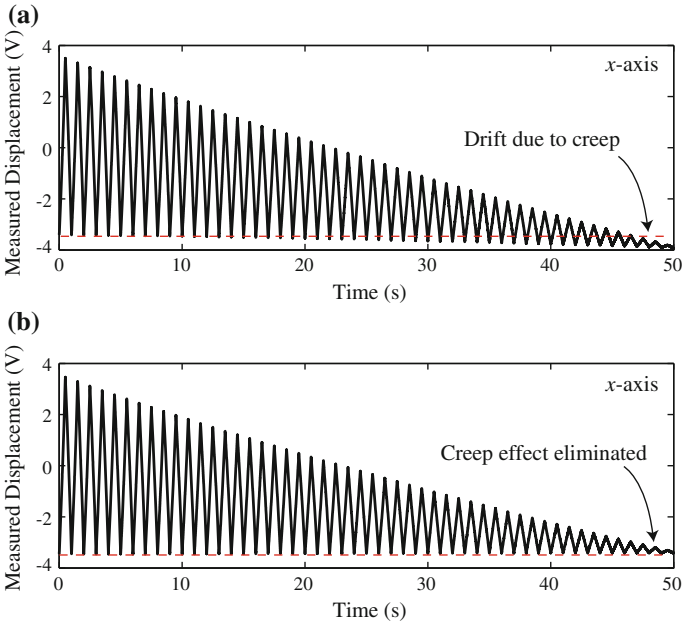


Fig. 9.5 Drift due to creep effect: **a** uncompensated; **b** compensated

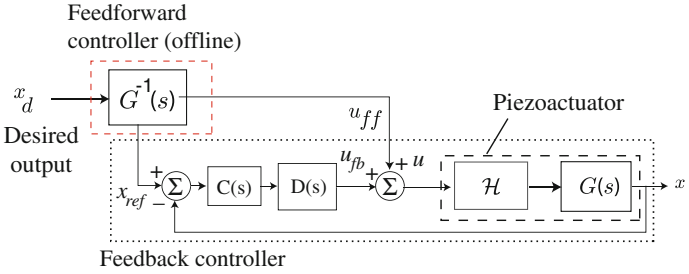


Fig. 9.6 Integrated feedforward and feedback controller for dynamics $G(s)$ and hysteresis $\mathcal{H}$. The integrated controller achieves high-speed positioning over large range. The block diagram shows a feedback controller for minimizing hysteresis and a feedforward controller for compensating the vibrational dynamics

piezoactuator behavior, the modeling process can be time-consuming, particularly when both the inverse dynamics and inverse hysteresis are used. If a simpler method to account for hysteresis is preferred over control performance, then high-gain feedback control can be used to linearize the nonlinear behavior of the piezoactuator. In this case, the vibrational dynamics are modeled, inverted, and combined with the feedback controller as shown in Fig. 9.6. However, when feedback is used, piezoactuators often exhibit low gain margin and can cause instability. For example, the frequency response of the piezoactuator depicted in Fig. 2.16, Chap. 2, shows a

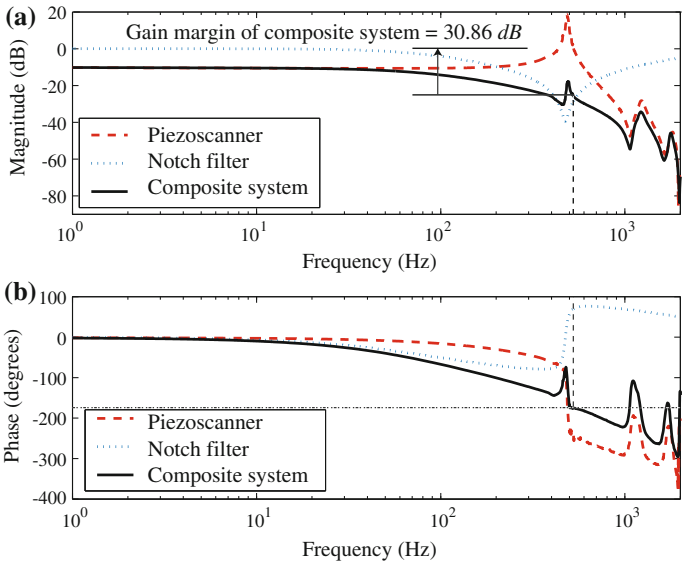


Fig. 9.7 Measured frequency response **a** magnitude vs. frequency **b** phase vs. frequency of the piezoactuator (*dashed line*), the notch filter (*dotted line*), and the notch filter cascaded with the piezoscanner (*solid line*). The measured gain margin of the original system is −17.05 dB, whereas the gain margin of the composite system is 30.86 dB

−17.05-dB gain margin. This low gain margin is attributed to the low structural damping and higher order dynamics (poles) that combine to pull the system's phase response below the −180° mark. Therefore, the feedback gain is severely limited, and a high-gain closed-loop system can potentially become unstable.

Gain margin can be improved by cascading the piezoactuator with a notch filter $D(s)$, which cancels the effect of the sharp resonant peak (Leang and Devasia 2007). For example, a notch filter of the following form

$$D(s) = k_D \frac{(s - 2\pi z_1)(s - 2\pi z_2)}{(s - 2\pi p_1)(s - 2\pi p_2)} \left(\frac{V}{V}\right), \tag{9.13}$$

where $k_D = 2.22, z_1 = -5 + j475, z_2 = -5 - j475, p_1 = -100$, and $p_2 = -5{,}000$ is used to bring the gain margin from −17.05 dB to over 30 dB as shown in Fig. 9.7. In the design of the notch filter $D(s)$, the zeros were chosen to suppress the effect of the dominant resonant peak of the piezoactuator (at 486 Hz). The modification compensated for the significant decrease in phase (180°) caused by the resonant poles. The zeros of the notch filter $D(s)$ were placed at 475 Hz to achieve high gain margin for the composite system despite small changes in the location of the resonance frequency of the open-loop system. To ensure that $D(s)$ was proper, a pair of poles were added to the notch filter at 100 and 5,000 Hz, and the poles helped to attenuate high frequency noise. The notch filter was realized using analog op-amp

circuits (e.g., Lam 1979, pp. 394–399) and its measured frequency response is shown by the dotted line in Fig. 9.7, together with the superimposed frequency response of the original system (dashed line) for comparing the old and new gain margins.

With the improved gain margin, traditional PD, PI, or PID controllers can be combined with the feedforward controller for high-speed precision positioning. The feedback controller provides robustness and minimizes hysteresis and creep effect. The feedforward controller is then designed to account for the vibrational dynamics.

9.4.1 Application to AFM Imaging

To illustrate both hysteresis and dynamics compensation, AFM imaging experiments were done to compare the performance of (1) high-gain feedback and (2) high-gain feedback with inverse feedforward input. At low frequency (1 Hz), open-loop imaging in Fig. 9.8a reveals distortion in AFM imaging due to mainly hysteresis. The application of feedback control with a notch filter shows that hysteresis is minimized in Fig. 9.8b. However, as the scanning frequency increases to 30 Hz, significant image distortion due to vibration under feedback control occurs. The ripples in the image shown in Fig. 9.8c are caused by vibration effect. The effect was minimized by augmenting feedforward input as shown in Fig. 9.8d. Therefore, the use of feedback with feedforward input computed from the linear dynamics model avoids the need to model/invert the complex nonlinear piezo-dynamics. Additionally, feedback provides robustness to parameter variation. The imaging result in Fig. 9.8d shows that the integrated approach provides a means of achieving precision positioning over a wider range of scan rates and displacements.

9.5 Iterative Feedforward Control

Simple linear feedforward control can be used to enhance the open-loop (and closed-loop) response of nanopositioning controllers. Iterative feedforward techniques are a second class of feedforward control that can be used in place of a feedback loop. The major benefit is nearly perfect tracking, but after some time for iterations and in the case of model-based iterative feedforward, more complicated DSP.

Rather than model and invert the dynamics and nonlinearities of a positioning system for feedforward control, if iterations can be used, the feedforward input can be found using iterative techniques. This approach is commonly referred to as iterative learning control (ILC). Some immediate advantages of ILC is minimal system information is needed for good tracking, and if an inverse model of the system dynamics is available and incorporated into the update law, the rate of convergence improves dramatically. The ILC framework is based on the observation that if the system's operating conditions remain the same during each operation, then the errors in the output response repeat. The objective is to make use of the information from

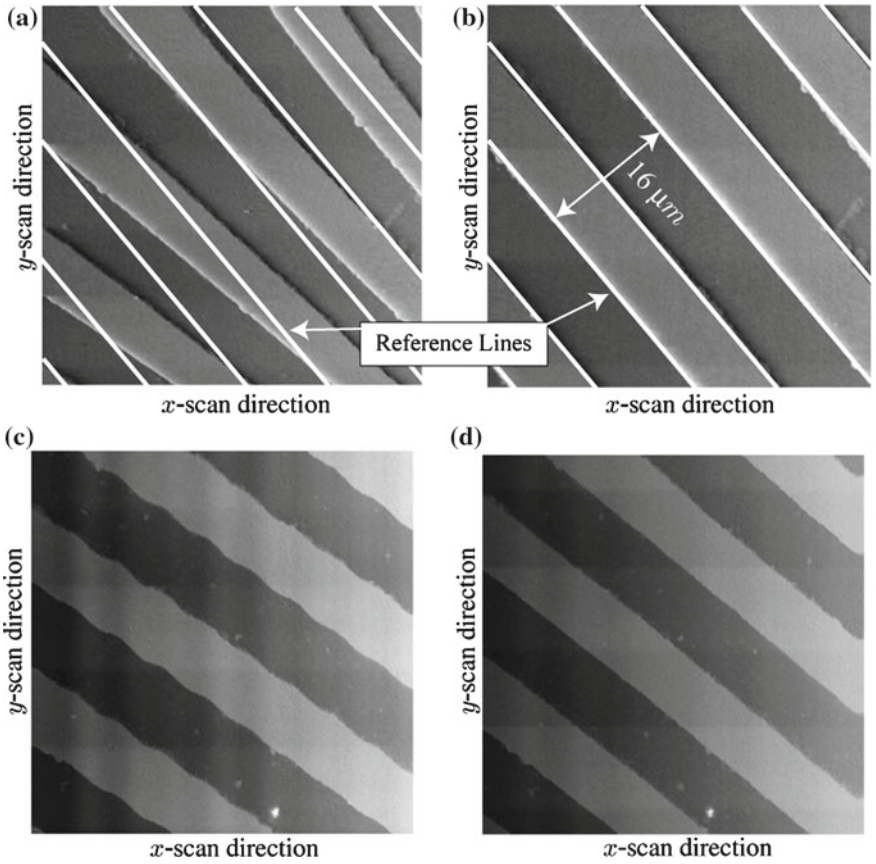


Fig. 9.8 AFM imaging results of calibration sample: Slow speed scanning (1 Hz) **a** open-loop, without feedback compensation and **b** with high-gain feedback compensation; High-speed scanning (30 Hz), **c** high-gain feedback control and **d** high-gain feedback with feedforward input

previous operating trials to improve the response in the next iteration; and as a result, the performance of a system can be improved through iteration. A block diagram of the control scheme is shown in Fig. 9.9, where y_d is the desired output, and u_k and y_k are the input and output at the kth trial, respectively. The task is to design a recursive algorithm that generates an input for the next step, i.e., u_{k+1}, such that the performance of the system is better than the previous step. As a requirement, the system to be controlled must operate repetitively over a finite time interval $I = [0, t_f]$.

ILC should not be confused with the feedback-based approach known as repetitive control (RC). Though they share the common trait of exploiting repetition to improve performance, they are fundamentally different. First, the feedback update in ILC is in the iteration domain k, whereas in RC the feedback is continuous in time like a feedback controller second, an ILC controller does not affect the stability of a

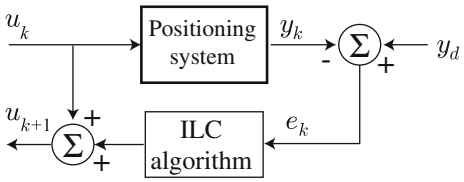


Fig. 9.9 Block diagram of ILC scheme. The iteration number is denoted by k

system in the same sense that an RC would. The instability of ILC essentially means the algorithm does not converge from one iteration to the next. Finally, the concept of RC is based on the Internal Model Principle (Francis and Wonham 1976; Inoue et al. 1981; Hara and Yamamoto 1988).

The ILC method was first proposed by Uchiyama (1978)[1] in the late 1970s and further developed by Arimoto et al. (1984) and Craig (1984) in the mid-1980s. Early contributions of modified ILC schemes were investigated by many others including Kawamura et al. (1988), Atkeson and McIntyre (1986) and Bondi et al. (1988). Since the work of Arimoto's group, the ILC methodology has been studied for a variety of systems from linear (Sugie and Ono 1991) to nonlinear nonminimum phase plants (Ghosh and Paden 2001) and a thorough treatment of the subject can be found from references Moore et al. (1992) and Chen and Wen (1999). In practice, the ILC methodology is a convenient solution for eliminating repeating errors. The approach has been applied to robotics (Atkeson and McIntyre 1986), internal combustion engines (Hoffmann et al. 2003) and permanent magnet motors (Tan et al. 2001), for example. Herein, the ILC method is described for high-performance nanopositioning.

There are numerous applications in nanopositioning where ILC excels, such as AFM imaging (Croft et al. 2001) and nanomanufacturing. These applications require the nanopositioner to operate repetitively, e.g., the back and forth lateral (x and y) scanning movements. As such, ILC can be used to eliminate errors due to hysteresis as well as the affects of vibration and creep. Because ILC requires minimal system information, it reduces the complexity of computing compared to inversion-based feedforward (Croft et al. 2001). Even if the operation is not repetitive, the ILC method can still be used. For example, ILC can be used off-line to *learn* the feedforward input and then the input can be applied to the piezo positioner.

9.5.1 The ILC Problem

The ILC method is presented with slight abuse in notation for convenience. Let T_s be an operator that maps an input u to an output y. The operator T_s can be thought of as a dynamical system, or a hysteretic system, representing, for example, the behavior of a nanopositioner. Given a desired output trajectory $y_d(t)$ defined over the fixed time interval I, the objective is to find an input $u_d(t)$ by repetitively applying the

[1] The work was not well known at the time because it was written in Japanese.

following iterative learning control algorithm (ILCA):

$$u_{k+1} = T_u u_k + T_e y_d - T_e y_k, \tag{9.14}$$

where T_u and T_e are casual operators (Moore et al. 1992). In Eq. (9.14), u_{k+1} is the input for the next trial, and y_d and y_k are the desired and current output, respectively. The task is to determine the conditions such that as the number of iterations $k \to \infty$, $u_k \to u_d$, and the input u_d satisfies

$$y_d = T_s u_d, \tag{9.15}$$

for all $t \in I$. As shown in Moore et al. (1992), for a linear system the ILCA Equation (9.14) converges if

$$||T_u - T_e T_s||_i < 1. \tag{9.16}$$

For example in continuous-time, Arimoto et al. (1984) showed that for a linear time-invariant system of the form,

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + B\mathbf{u}(t), \tag{9.17}$$
$$\mathbf{y}(t) = C\mathbf{x}(t), \tag{9.18}$$

the ILCA given by

$$u_{k+1}(t) = u_k(t) + \rho[\dot{y}_d(t) - \dot{y}_k(t)], \quad \text{for } t \in I, \tag{9.19}$$

converges provided that $u_0(t)$ is continuous on I, $y_d(t)$ is continuously differentiable on I, and

$$||I - CB\rho||_\infty < 1, \tag{9.20}$$

where $||\mathbf{z}||_\infty$ is the standard infinity norm of a vector $\mathbf{z}$. The constant ρ is called the *iteration gain*. Furthermore, convergence of the output requires the initial condition be reset at the start of each trial, that is, $y_k(0) = y_d(0)$, for all positive integers $k \in \mathbb{Z}^+$.

It is insightful to note that in condition (9.20) a constant ρ exists provided the matrix CB has full rank. For a relative degree one system ($CB \neq 0$), this condition is easily met. What this means is the input u appears explicitly in the first-derivative of the output, resulting in direct feed-through or transmission (Sugie and Ono 1991),

$$\dot{\mathbf{y}}(t) = C\dot{\mathbf{x}}(t) = CA\mathbf{x}(t) + CB\mathbf{u}(t). \tag{9.21}$$

The simple ILCA Equation (9.19) is a typical proportional or P-type ILCA (Sugie and Ono 1991; Saab 1994), exploiting the relative degree of the system. For such

a scheme with fixed ρ, the only information needed for convergence is the sign of the product CB. Therefore, ILC requires minimal system information. In connection with ILCA Eq. (9.14), let $T_e = I$ and for a fixed T_e = constant, condition (9.16) is satisfied provided the phase of T_s is known. This can be interpreted as knowing the direction in which the input should be applied to reduce the tracking error for the next iteration, a concept similar to Arimoto et al.'s assumption for knowing the sign of the CB term. As long as the input is pointing *away* from the direction of increasing error, the input update law will converge.

In fact, for a linear single-input single-output (SISO) system, the following conclusion can be drawn based on Sugie and Ono's work (1991). First, it is assumed that:

1. The linear system has a well defined relative degree r. The relative degree in this case is simply the difference between the order of the numerator and denominator of the system transfer function;
2. The first $r-1$ derivatives of the output satisfies

$$\begin{aligned} y_k(0) &= y_d(0), \\ \dot{y}_k(0) &= \dot{y}_d(0), \\ \vdots &= \vdots \\ y_k^{(r-2)}(0) &= y_d^{(r-2)}(0), \\ y_k^{(r-1)}(0) &= y_d^{(r-1)}(0), \end{aligned}$$

   for $k = \mathbb{Z}^+$; and
3. The input $u_0 \in C^0(I)$ and $y_d \in C^{(r)}(I)$.

Then, the following ILCA

$$u_{k+1}(t) = u_k(t) + \rho e_k^{(r)}(t), \tag{9.22}$$

where $e(t) \triangleq y_d(t) - y_k(t)$ and $e^{(r)}(t) \triangleq \frac{d^r}{dt^r}[e(t)]$, converges uniformly in t if $x_k(t_0) = x_d(t_0)$ for all $k \in \mathbb{Z}^+$ and the iteration gain ρ satisfies

$$||I - CA^{r-1}B\rho||_\infty < 1. \tag{9.23}$$

In principle, the ILCA given by Eq. (9.22) is applicable to right invertible systems (Sugie and Ono 1991). Such systems also include certain classes of nonlinear and time-varying systems.

9.5.2 Model-Based ILC

By exploiting more information about the system, model-based ILC provides higher performance, in particular, improved the rate of convergence. One of the earlier

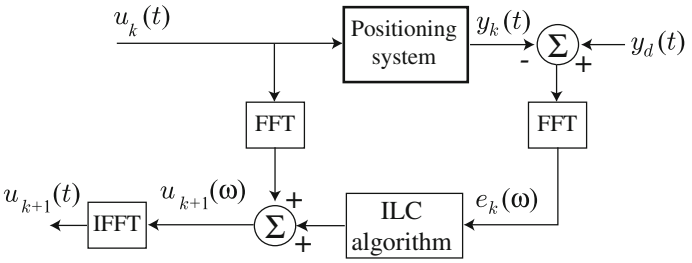


Fig. 9.10 Block diagram of frequency-domain ILC scheme

works on model-based ILC was by Atkeson and McIntyre in (1986), where the ILC feedforward input was injected into a feedback loop to control a robotic arm. The input update law for a linear system in the frequency domain takes the form

$$u_{k+1}(\omega) = u_k(\omega) + G^{-1}(\omega)\big[y_d(\omega) - y_k(\omega)\big], \tag{9.24}$$

where $G^{-1}(\omega)$ is the inverse system. With a perfect model $G(\omega)$, it is easy to see that perfect tracking is achieved in one step by noting that $u_k(\omega) = G^{-1}(\omega)y_k(\omega)$. However, in practice modeling errors exist, and thus the ILCA Equation (9.24) is modified to reflect this fact (Tien et al. 2005)

$$u_{k+1}(\omega) = u_k(\omega) + \rho(\omega)\hat{G}^{-1}(\omega)\big[y_d(\omega) - y_k(\omega)\big], \tag{9.25}$$

where $\hat{G}^{-1}(\omega)$ is an approximate inverse of the system and $\rho(\omega)$ is a frequency-dependent iteration gain. Figure 9.10 shows the block diagram of the frequency-domain implementation of ILCA Equation (9.25). First, the input $u_k(t)$ and tracking error $e_k(t)$ are Fourier transformed, then the ILCA is applied, producing the updated input $u_{k+1}(\omega)$. The time-domain input $u_{k+1}(t)$ is obtained by inverse Fourier transform. The input is applied to the system and the process is repeated.

ILCA Equation (9.25) converges when the difference in the phase between model and actual system dynamics is less than 90°, i.e., $|\Delta\theta(\omega)| \leq \pi/2$, and

$$0 < \rho(\omega) < \frac{2\cos(\Delta(\omega))}{A(\omega)}, \tag{9.26}$$

where $A(\omega)$ is the difference in the magnitude response between the model $G(\omega)$ and actual system dynamics.

At frequencies where the phase difference $\Delta(\omega)$ is greater than 90°, a sign change in $\cos(\Delta(\omega))$ from positive to negative occurs, and thus, requires a sign change of iteration gain, from positive to negative, for convergence. Therefore, the iteration gain should be chosen as

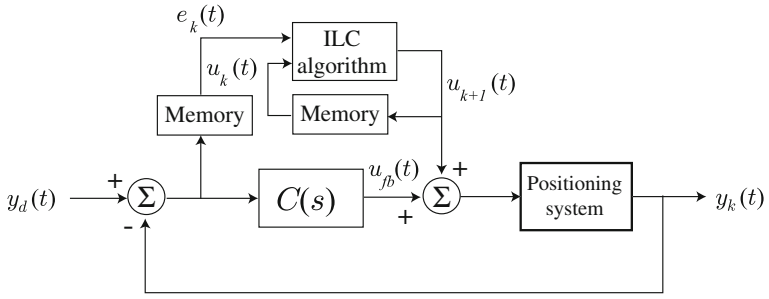


Fig. 9.11 Block diagram of integrated feedback and ILC

$$0 < \rho(\omega) < \frac{2\cos(\Delta(\omega))}{A(\omega)}, \text{ for } \cos(\Delta(\omega)) > 0;$$
$$\rho(\omega) < \frac{2\cos(\Delta(\omega))}{A(\omega)} < 0, \text{ for } \cos(\Delta(\omega)) < 0; \tag{9.27}$$

Both regular and model-based ILC have been applied to nanopositioning systems (Lee et al. 2000; Choi et al. 2002; Bristow et al. 2008; Tien et al. 2005; Kim et al. 2008). The method has also been integrated with feedback control to account for hysteresis effects (Bristow et al. 2008; Wu et al. 2008). In this case, the feedback is designed to 'linearize' the system's behavior and a feedforward input generated by an ILCA is injected downstream into the system's input as shown in Fig. 9.11.

The enhanced performance of model-based ILC requires a relatively accurate system model. Recently, a simple ILC algorithm was proposed which eliminates the need for a model (Kim et al. 2008; Li and Bechhoefer 2008). First, let $u_0(\omega) = y_d(\omega)/G(0)$, where $G(0)$ is the DC gain of the system. Then the input update law is given by

$$u_{k+1}(\omega) = \frac{u_k(\omega)}{y_k(\omega)} y_d(\omega), \tag{9.28}$$

for $k \in \mathcal{Z}^+$, where $u_{k+1}(\omega) = 0$ if $y_d(\omega) = 0$.

9.5.3 Nonlinear ILC

The ILC algorithms presented where based on linear models. Piezoactuators in nanopositioners exhibit nonlinearity such as hysteresis. Although the ILC approach is well-suited for precise control of linear dynamics, the challenge is developing a convergence criteria for hysteresis. The difficulty in proving convergence of ILC algorithms for hysteretic systems arises due to two main reasons: (i) branching effects and (ii) nonlinearity of each branch (Brokate and Sprekels 1996). The latter issue can be addressed by standard ILC methods. For example, the convergence of ILC on a

single branch was shown in Hu et al. (2004), in which the hysteresis nonlinearity was modeled as a single branch (using a polynomial). Alternatively, a functional approach was proposed for systems that satisfy the incrementally strictly increasing operator (ISIO) property (Venkataraman and Krishnaprasad 2000); however, the branching effect in hysteresis results in loss of the ISIO property (Leang and Devasia 2003). The reason branching causes problems in proving convergence is because branching prevents the ILC algorithm from predicting the direction in which the input needs to be changed based on a measured output error. For example, the input error can grow from one iteration to the next.

The inability, to predict the direction in which the input needs to be changed for reducing the output error in hysteretic systems, can be overcome if the input–output behavior is restricted to belong on one single hysteresis branch. This observation that the direction can be determined from the output error on a single branch was used to prove the convergence for an ILC algorithm for hysteretic systems in Leang and Devasia (2006). First, the desired output trajectory is partitioned into monotonic sections (several branches). The algorithm is applied to each section until a desired tracking precision is achieved.

In particular, the ILCA of the following form can be used to compensate for hysteresis (Leang and Devasia 2006):

$$u_{k+1}(t) = u_k(t) + \rho[y_d(t) - y_k(t)], \forall t \in I. \tag{9.29}$$

The ILCA Equation (9.29) converges if the desired trajectory $v_d(t)$ is continuous and monotonic over the finite time interval I. The iteration gain ρ for convergence is based on the parameters of the hysteresis model, such as the Preisach hysteresis model. It is pointed out that a sufficiently small iteration gain can be found provided the output different can be bounded above and below by the input difference (Leang and Devasia 2006).

9.5.3.1 Application to AFM Imaging

The ILCA Equation (9.29) was applied to compensate for hysteresis in AFM imaging. The details of the experiment can be found in Leang and Devasia (2006). A flow chart for the experimental implementation of the algorithm is shown in Fig. 9.12. The desired trajectory is a typical triangular scan path for the x-axis.

The algorithm was applied to individual monotonic partitions as outlined in the flow chart in Fig. 9.12. The ILC tracking results for the first and second branch are shown in Fig. 9.13. For the first branch ($m = 1$), after $k_1 = 40$ iterations, the magnitude of the maximum tracking error, $e_{\max}$, is 0.24 % (Fig. 9.13d1). This error corresponds to ± 15.5 mv, or ± 216 nm (which is approximately the noise level of the sensor measurement at 15 mv). Additionally, the root-mean-square error, e_{rms}, is 0.087 %, where $T = 0.5$ s (Fig. 9.13e1). Likewise, after $k_2^* = 40$ iterations, the tracking error on the second branch reduces to $e_{\max} = 0.26\,\%$ and $e_{\text{rms}} = 0.10\,\%$ (Fig. 9.13d2 and e2). By comparison, without ILC compensation, the maximum error

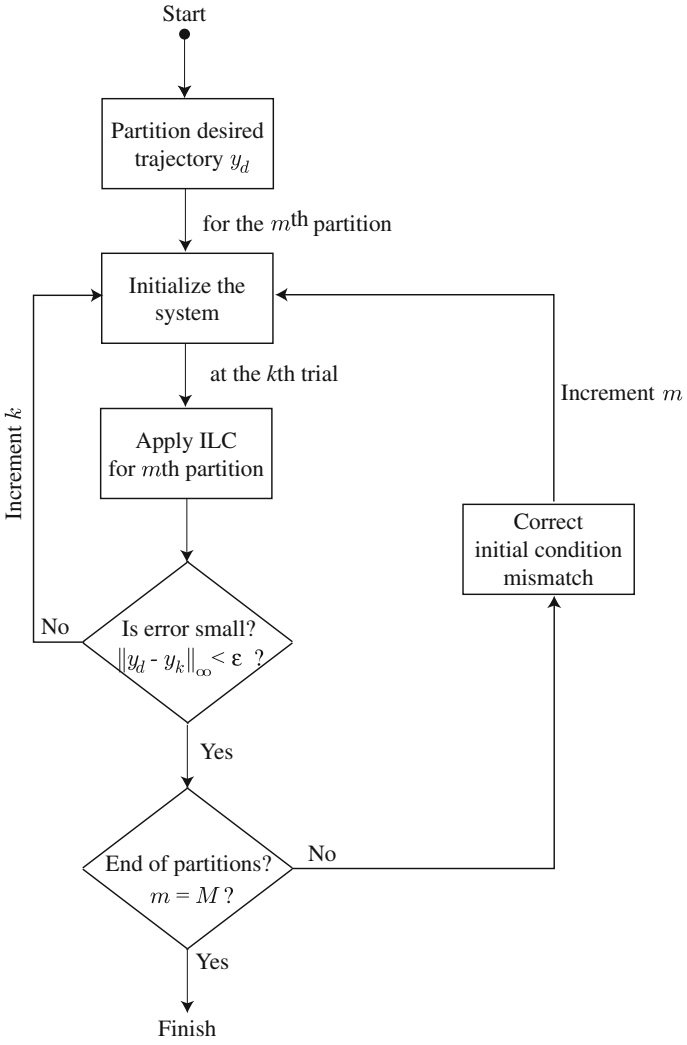


Fig. 9.12 The flow chart for implementing the multi-branch ILC algorithm

due to hysteresis for the first and second branch are 7.15 and 6.66 %, respectively. The results show that the ILC method compensates for hysteresis effect by reducing the tracking error to the noise level of the sensor measurement. In particular, the maximum error $e_{\max}$ is reduced by over 96 %. Additionally, the output error decays rapidly to the noise level as indicated in Fig. 9.13d1, e1, d2 and e2. Therefore, the results show that ILC achieves high-precision positioning for piezo-based systems; the error reduces to the noise level of the sensor measurement.

The inputs found using the ILCA were applied to AFM imaging a calibration sample. A calibration sample consisting of parallel markings with a 16 μm pitch

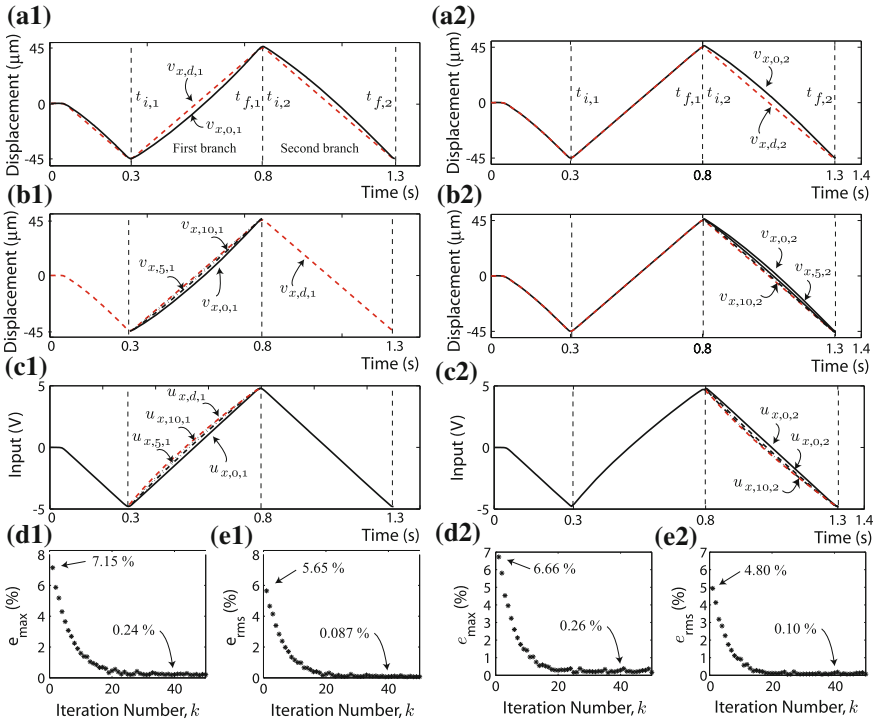


Fig. 9.13 Experimental AFM output tracking results for the ILC method applied to the x-axis, where **a1** through **e1**shows tracking performance for first monotonic section of the desired trajectory; and **a2** through **e2** shows the performance for the second section

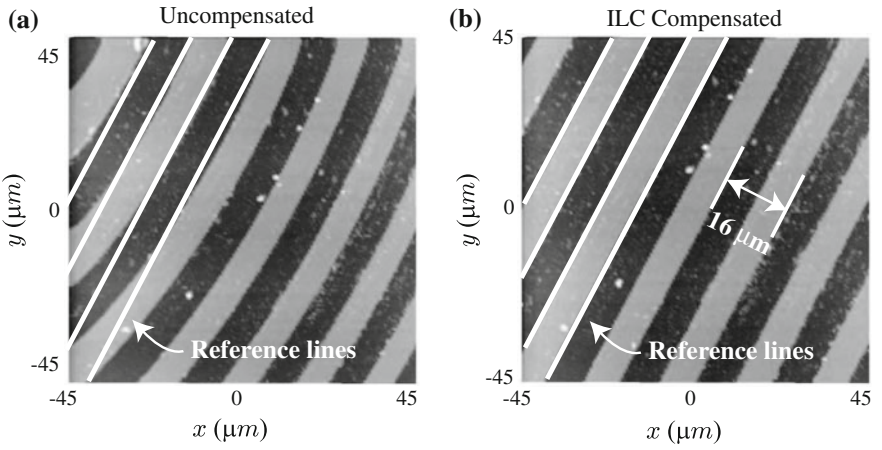


Fig. 9.14 Atomic force microscope imaging results. The sample is a 16 μm-pitch encoder grating (calibration sample). **a** Uncompensated image and **b** ILC compensated image

was imaged using and experimental AFM system. The imaging process is initiated by gradually moving the sample close to the probe until a desired (*nominal*) probe-to-sample distance (distance between the AFM-probe tip and the sample surface) is achieved. Then the AFM-probe tip is scanned across the sample surface. During AFM imaging, the effects of the probe-to-sample distance is measured as the probe is scanned across the sample's surface. In particular, the displacement of the AFM-probe (cantilever) is measured using an optical sensor and the measurements are used to construct an image of the sample topography. An image of the surface topology is obtained by plotting the measured cantilever displacement versus the desired v_x and v_y-position of the AFM probe—this mode of operation is called the constant-height contact mode (for other AFM modes of operation, see, e.g., Binnig (1992)).

The imaging results are shown in Figs. 9.14a and b. Figure 9.14a is an image without ILC compensation and it shows the effect of hysteresis. For example, the features are significantly distorted due to hysteresis; specifically, the parallel features appear curved and they vary in width—the features are separated by 16 μm as shown in Fig. 9.14b. Such distortions give an inaccurate representation of the sample surface. However, by applying the input found using the ILC algorithm, the distortions can be corrected as shown in Fig. 9.14b. In the figure, reference lines are superimposed on the image to illustrate the improvement in precision achieved by using ILC. By compensating for hysteresis error, the corrected-image more accurately represents the actual surface topology compared to the image with distortions when ILC is not used.

9.5.4 Conclusions

This chapter described an inversion-based feedforward approach to compensate for dynamic and hysteresis effects in piezoactuators with application to AFM technology. To handle the coupled behavior of dynamics and hysteresis, a cascade model was presented to enable the application of inversion-based feedforward control. The dynamics, which include vibration and creep, are modeled using linear transfer functions. A frequency-based method is used to invert the linear model to find an input that compensates for vibration and creep. The inverse is noncausal for nonminimum-phase systems. Similarly, the hysteresis is handled by a nonlinear ILC approach that exploits the Preisach hysteresis model. Finally, feedforward control is combined with feedback control to compensate for the linear dynamics to achieve high-bandwidth positioning.

References

Arimoto S, Kawamura S, Miyazaki F (1984) Bettering operation of robots by learning. J Robot Syst 1(2):123–140

Arimoto S, Kawamura S, Miyazaki F (1984) Bettering operation of dynamic systems by learning: a new control theory for servomechanism or mechatronics systems. In: Proceedings of American control conference, pp 1064–1069
Atkeson CG, McIntyre J (1986) Robot trajectory learning through practice. In: IEEE Int Conf Robot Autom, pp 1737–1742
Bayo E (1987) A finite-element approach to control the end-point motion of a single-link flexible robot. J Robot Syst 4:63–75
Binnig G (1992) Force microscopy. Ultramicroscopy 42–44:7–15
Bondi P, Casalino G, Gambardella L (1988) On the iterative learning control theory for robotic manipulators. IEEE J Robot Autom 4(1):14–22
Bristow DA, Dong J, Alleyne AG, Ferriera P, Salapaka S (2008) High bandwidth control of precision motion instrumentation. Rev Sci Instr 79:103704
Brokate M, Sprekels J (1996) Hysteresis and phase transitions. Springer, New York
Chen Y, Wen C (1999) Iterative learning control: convergence, robustness and applications. Springer, New York
Choi GS, Lim YA, Choi GH (2002) Tracking position control of piezoelectric actuators for periodic reference inputs. Mechatronics 12(5):669–684
Craig JJ (1984) Adaptive control of manipulators through repeated trials. In: Proceedings of American control conference, pp 1566–1573
Croft D, Shed G, Devasia S (2001) Creep, hysteresis, and vibration compensation for piezoactuators: atomic force microscopy application. Trans ASME J Dyn Syst Meas Cont 123:35–43
Devasia S, Chen D, Paden B (1996) Nonlinear inversion-based output tracking. IEEE Trans Autom Cont 41(7):930–942
Devasia S (1997) Output tracking with nonhyperbolic and near nonhyperbolic internal dynamics: helicopter hover control, AIAA. J Guidance Cont Dyn 20(3):573–580
Dewey JS, Leang KK, Devasia S (1998) Experimental and theoretical results in output-trajectory redesign for flexible structures. ASME J Dyn Syst Meas Cont 120(4):456–461
Francis BA, Wonham WM (1976) The internal model principle of control theory. Automatica 12(5):457–465
Ghosh J, Paden B (2001) Iteartive learning control for nonlinear nonminimum phase plants. ASME J Dyn Syst Meas Cont 123:21–30
Hara S, Yamamoto Y, Omata T, Nakano M (1988) Repetitive control system: a new type servo system for periodic exogenous signals. IEEE Trans Autom Cont 33(7):659–668
Hoffmann W, Peterson K, Stefanopoulou AG (2003) Iterative learning control for soft landing of electromechanical valve actuator in camless engines. IEEE Trans Cont Syst Tech 11(2):174–184
Hu M, Du H, Ling S-F, Zhou Z, Li Y (2004) Motion control of an electrostrictive actuator. Mechatronics 14(2):153–161
Inoue T, Nakano M, Iwai S (1981) High accuracy control of a proton synchrotron magnet power supply. In: Proceedings of 8th world congress IFAC, vol 20. pp 216–221
Isidori A (1995) Nonlinear control systems, 3rd ed. Springer, New York
Kawamura S, Miyazaki F, Arimoto S (1988) Realization of robot motion based on a learning method. IEEE Trans Syst Man Cybern 18(1):126–134
Kim K-S, Zou Q, Su C (2008) A new approach to scan-trajectory design and tracking: AFM force measurement example. ASME J Dyn Syst Meas Cont 130:051005-1–051005-10
Kim K-S, Lin Z, Shrotriya P, Sundararajan S, Zou Q (2008) Iterative control approach to high-speed force-distance curve measurement using AFM: time dependent response of PDMS. Ultramicroscopy 108:911–920
Lam HY-F (1979) Analog and digital filters: design and realization. Prentice-Hall, New York
Leang KK, Devasia S (2003) Iterative feedforward compensation of hysteresis in piezo positioners. In: IEEE 42nd conference on decision and controls, invited session on nanotechnology: control needs and related perspectives, pp 2626–2631
Leang KK, Devasia S (2006) Design of hysteresis-compensating iterative learning control for piezo positioners: application to atomic force microscopes. Mechatronics 16(3–4):141–158

Leang KK, Devasia S (2007) Feedback-linearized inverse feedforward for creep, hysteresis, and vibration compensation in afm piezoactuators. IEEE Trans Cont Syst Technol 15(5):927–935
Lee TH, Tan KK, Lim SY, Dou HF (2000) Iterative learning control of permanent magnet linear motor with relay automatic tuning. Mechatronics 10:169–190
Li Y, Bechhoefer J (2008) Feedforward control of a piezoelectric flexure stage for AFM. In: American control conference. Seattle, pp 2703–2709
Moore KL, Dahleh M, Bhattacharyya SP (1992) Iterative learning control: a survey and new results. J Robot Syst 9(5):563–594
Saab SS (1994) On the p-type learning control. IEEE Trans Autom Cont 39(11):2298–2302
Sugie T, Ono T (1991) An iterative learning control law for dynamical systems. Automatica 27(4):729–732
Tan KK, Lee TH, Zhou HX (2001) Micro-positioning of linear-piezoelectric motors based on a learning nonlinear PID controller. IEEE/ASME Trans Mechatron 6(4):428–436
Tan X, Baras JS (2005) Adaptive identification and control of hysteresis in smart materials. IEEE Trans Autom Cont 50(6):827–839
Tien S, Zou Q, Devasia S (2005) Iterative control of dynamics-coupling-caused errors in piezoscanners during high-speed AFM operation. IEEE Trans Cont Syst Tech 13(6):921–931
Uchiyama M (1978) Formation of high-speed motion pattern of a mechanical arm by trial. Trans Soc Instrum Cont Eng 14(6):706–712
Venkataraman R, Krishnaprasad PS (2000) Approximate inversion of hysteresis: theory and numerical results. In: Proceedings of 39th IEEE conference on decision and control, pp 4448–4454
Wu Y, Zou Q, Su C (2008) A current cycle feedback iterative learning control approach to AFM imaging. In: American control conference, pp 2040–2045
Zou Q, Devasia S (1999) Preview-based stable-inversion for output tracking. ASME J Dyn Syst Meas Cont 121(4):625–630
Zou Q, Devasia S (2004) Preview-based optimal inversion for output tracking: application to scanning tunneling microscopy. IEEE Cont Syst Technol 12(3):375–386
Zou Q, Devasia S (2007) Preview-based stable-inversion for output tracking of nonlinear nonminimum-phase systems: the vtol example. Automatica 41(1):117–127
Zou Q (2009) Optimal preview-based stable-inversion for output tracking of nonminimum-phase linear systems. Automatica 45(1):230–237

Chapter 10
Command Shaping

The speed of an electromechanical scanner is limited by its first resonance frequency. To maximize scan speed, input signals are required that contain negligible frequency components near, or above the first resonance frequency. Such signals are usually obtained by low-pass filtering the desired scan trajectory. However, this introduces curvature and ripple into linear (constant velocity) scan regions.

In this chapter, input signals are designed with guaranteed linear regions and minimal harmonic components above a chosen frequency. The proposed scanning trajectories are proven by simulation and experiment to induce less vibration than existing techniques.

10.1 Introduction

Many scientific and industrial machines contain mechanical scanners driven with periodic trajectories. For example, beam steering scanners (Potsaid et al. 2007), manufacturing robots, cam motion generators, and scanning probe microscopes (Meyer et al. 2004). In this chapter, without knowledge of system dynamics, periodic input signals are designed to maximize the speed and accuracy of band-limited scanners. The focus is on design of input signals for scanning probe microscope nanopositioning stages, as reviewed in Zou et al. (2004), Abramovitch et al. (2007) and Devasia et al. (2007).

10.1.1 Background

The foremost difficulty associated with high-speed scanners is illustrated in Fig. 10.1. Here, the system G represents a mechanical scanner driven with a triangular signal r. In this example, the mechanical system G is a unity-gain second-order low-pass

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_10,

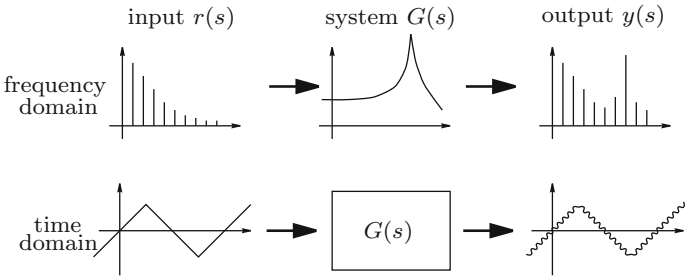


Fig. 10.1 A triangular scanning signal distorted by a typical mechanical system

system with resonance frequency ω_n and damping ratio ξ_n, that is

$$G(s) = \frac{\omega_n^2}{s + 2\omega_n \xi_n s + \omega_n^2}. \tag{10.1}$$

When G is excited by an input with significant frequency content at, or near, the resonance frequency, this content is amplified and appears as output ripple. For systems with settling time shorter than the scan period, resonance excitation appears after high-frequency events such as the peak of a sharp waveform. In addition to resonance excitation, frequency components of the input above the resonance frequency are attenuated and shifted in phase by 180°.

If we quantify the tracking error e as the difference between input and output, i.e.,

$$e(t) = r(t) - y(t), \tag{10.2}$$

the error can be expressed in the Laplace domain as

$$e(s) = r(s)\,(1 - G(s))\,. \tag{10.3}$$

Thus, at frequencies where $r(s)$ is significant and $G(s)$ is not close to unity, the error is significant. There are three possible means for reducing error: inverting $G(s)$ (or otherwise filtering $r(s)$); reducing $G(s)$ to unity where $r(s)$ is significant; and reducing $r(s)$ to zero where $G(s)$ is not unity. The characteristics of each approach are discussed in the following:

10.1.1.1 Inverting *G*

Inversion of G is a commonly applied technique that can provide good performance if the plant model or its frequency response is known with high accuracy. When the input is periodic, inversion is easily accomplished by multiplying the Fourier coefficients of the input by the inverse frequency response.

The foremost problem with inversion is the lack of robustness to changes in plant dynamics, especially if the system is resonant (Zhao and Jayasuriya 1994). Perfect inversion can also result in large amplitudes if the system response is small or zero at harmonics of the input (Dewey et al. 1998). Large signal amplitudes can cause actuator saturation and exacerbate amplitude-dependent nonlinearity such as hysteresis.

The main attraction of inversion-based control is its simplicity and ease of implementation, particularly in high-speed applications. With consideration of plant uncertainty, a significant improvement in imaging speed was achieved in Schitter and Stemmer (2004) and Croft et al. (2001). Another inversion based technique (Dewey et al. 1998) avoids large amplitudes by trading off tracking performance for reduced input energy. A related work (Perez et al. 2004) generates optimal output trajectories with minimal input energy and was successfully applied to an STM scanner.

Iterative inversion is a more elaborate technique that requires a sensor, but overcomes many limitations of linear inversion and can provide excellent performance when no exogenous disturbance is present. Although such techniques originally required a reference model (Wu and Zou 2007), in 2008, both Kim and Zou (2008) and Li and Bechhoefer (2008) presented techniques that operate without any prior system knowledge. Iterative techniques however, require time to converge, can generate large input signals, and require digital signal processing hardware.

Compared to feedback control, it is difficult or impossible to use non-iterative feedforward compensation for accurate inversion of nonlinearity such as hysteresis. There is also no immunity to exogenous disturbance, offset, and gain drift.

10.1.1.2 Controlling *G*

Controlling *G* is a popular method for linearizing electromechanical systems at low-frequencies. Proportional-Integral (PI) controllers, with and without notch filters for gain-margin improvement are commonly used, for a review see Zou et al. (2004), Abramovitch et al. (2007) and Devasia et al. (2007). If sufficient sensor bandwidth is available, feedback control can also be used to damp mechanical resonance. For this purpose, Positive Position Feedback (PPF) control and variants are straightforward to implement and perform well (Fanson and Caughey 1990; Aphale et al. 2007, 2008). The major disadvantages of feedback are: the addition of sensor-induced noise, limited bandwidth, and tracking lag.

The addition of a feedforward controller can significantly improve the bandwidth and tracking lag of feedback systems without compromising stability or induced noise (Leang and Devasia 2007; Pao et al. 2007). However, due to the nature of feedforward control, immunity to hysteresis and disturbance is not improved and performance robustness can be reduced (Devasia 2002).

If only attenuation of mechanical resonance is required, the technique of shunt damping can be employed as an alternative to sensor-based feedback control (Fleming and Moheimani 2006; Aphale et al. 2007). Shunt damping can provide attenuation of mechanical resonance without contributing sensor-induced noise.

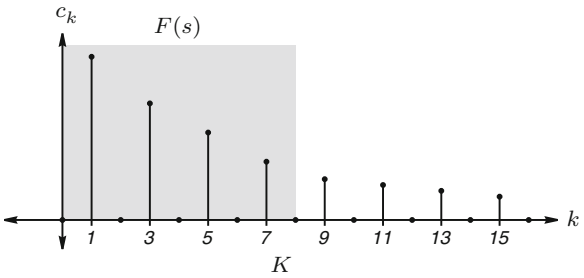


Fig. 10.2 The Fourier coefficients of a triangle wave c_k filtered by $F(j\omega)$. Harmonics k from nine onwards are removed

10.1.1.3 Reducing the Magnitude of $r(s)$

Reducing the magnitude of $r(s)$ towards zero at frequencies near and above the resonance is a simple, practical, and popular technique for minimizing induced vibration. The most obvious technique for reducing high-frequency content in $r(s)$ is to simply low-pass filter the signal. For periodic signals, this can be performed perfectly in the frequency domain by multiplying the Fourier coefficients of the reference signal with the filter magnitude specification, then applying the inverse Fourier Transform.

The greatest disadvantage of low-pass filtering is the ripple introduced into linear (constant velocity) regions of the scan. As an explanation, consider the Fourier coefficients c_k of a periodic triangle wave shown in Fig. 10.2. If the filter $F(s)$ is designed to pass the first K harmonics and attenuate the remainder, the filtered triangle wave $y(t)$ can be viewed as the original ideal trajectory $r(t)$, minus an error signal $e(t)$, i.e.,

$$y(t) = r(t) * F(t) = r(t) - e(t). \tag{10.4}$$

Conceptually, the error signal $e(t)$ is the *rippled part* of $y(t)$. In the frequency domain, $e(s)$ comprises the frequency components removed from $r(s)$ by $F(s)$, i.e.,

$$\begin{aligned} e(s) &= r(s) - y(s) \\ &= r(s) - F(s)r(s) \\ &= r(s)\left(1 - F(s)\right). \end{aligned} \tag{10.5}$$

More exactly, the Fourier coefficients of $e(t)$ are those of the original triangle above $k = K$. That is, if e_k is the Fourier coefficients of $e(t)$,

$$e_k = \begin{cases} 0 & \text{when } -K < k < K \\ c_k & \text{otherwise} \end{cases} \tag{10.6}$$

The power P_e in the error signal $e(t)$ can be quantified using Parseval's equality,

$$P_e = \sum_{k=-\infty}^{\infty} |c_k|^2 . \tag{10.7}$$

As a consequence of this equality, the error becomes larger as signal bandwidth is reduced. This contradicts the original goal of low-pass filtering, to reduce scan error. Furthermore, as the filter $F(s)$ becomes more efficient, i.e., provides faster roll-off and better attenuation; the error also increases.

To eliminate the ripple and curvature introduced by frequency domain filtering, time domain signal shaping was developed. This allows critical parts of the trajectory to be retained while corners and turnaround points are smoothed to reduce high-frequency content. The most straightforward signal shaping method is the minimum acceleration technique. This involves replacing the turning points of a trajectory with a smooth quadratic curve. Although this minimizes inertial force, it does not lead to optimal tracking performance. Minimum acceleration signals were used by Rost and colleagues to achieve SPM imaging rates of up to 200 frames per second (Rost et al. 2005).

Better performance than the minimum acceleration signal can be achieved by convolving the desired trajectory with a signal that minimizes induced vibration (Masterson et al. 2000; Singhose et al. 1995; Singer and Seering 1990). Such techniques have found broad industrial application in manufacturing machinery. The foremost reported disadvantages of convolution techniques are: the significant filter length (signal delay), sensitivity to resonance frequency variation (Vaughan et al. 2008), and increased control signal magnitude (Masterson et al. 2000). A performance comparison of convolution-based techniques can be found in Vaughan et al. (2008). Design tools for convolution-based input shaping can be obtained commercially from Convolve, Inc. Armonk, NY.

In addition to the many industrial applications, convolution-based input shaping has also been employed in nanopositioning applications. In Schitter et al. (2006) the triangular trajectory of an AFM scanner was shaped to reduce vibration. The shaped triangle wave contains a flat section at each signal apex that persists for half the resonance period. The shaped-triangle technique can provide excellent performance if the resonance frequency is exactly known and the mechanical system is second order. Unfortunately, the performance degrades if the resonance frequency is not exactly known or if the system order is greater than two. This technique is compared to others in Sect. 10.6.

10.1.2 The Optimal Periodic Input

In Fleming and Wills (2009), a new method was proposed for designing periodic input trajectories for mechanical systems. The method optimizes a desired trajectory

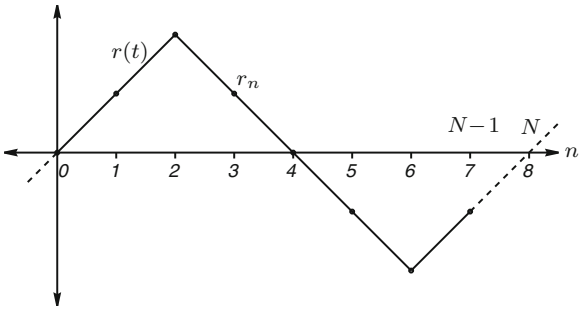


Fig. 10.3 A periodic signal $r(t)$ and its samples r_n

based on frequency domain and/or time domain cost-functions. A key feature is that certain parts of the trajectory can be fixed. For scanning applications, the proposed technique can be used to design input signals with perfectly linear (constant velocity) regions and minimal signal power above a chosen frequency. Comparison with other techniques shows a significant reduction in tracking error.

The proposed technique is most closely related to the convolution techniques discussed in the previous section. The resulting optimal signals are similar in appearance to minimum acceleration signals but provide much improved performance. Unlike feedforward and feedback techniques, a parametric model or sensor is not required and the implementation is straightforward.

In the following section, the signal optimization scheme is described. This is followed by a range of cost functions in Sect. 10.3 that minimize properties such as acceleration and signal power. These can be used to generate signals with fixed and free regions that are optimal with respect to the chosen cost function. The frequency-weighted-power cost function is discussed in Sect. 10.5 as a technique for generating input signals for low-bandwidth positioning stages. The performance with respect to other techniques is evaluated by simulation in Sect. 10.6 and experiment in Sect. 10.7. A summary of results and conclusions follow in Sect. 10.8.

10.2 Signal Optimization

In this section, the signal optimization problem is defined and solved. The method begins with an ideal scanning trajectory, this is split into regions that are fixed, and regions that can be modified. The variable parts are then redesigned to minimize a quadratic cost function. In the next section, cost functions are described for various time and frequency domain objectives.

Consider the triangular waveform $r(t)$ plotted in Fig. 10.3. The samples of $r(t)$ are denoted $r_n = r(\Delta n)$ where Δ is the sampling interval, $n \in \{0, 1, 2, \ldots, N-1\}$ and N are the number of samples per period. In the illustration, the sampling frequency $F_s = \frac{1}{\Delta}$ is equal to 8 times the triangle frequency F_T.

The samples of $r(t)$ over one period can be written in vector notation:

$$r = \begin{bmatrix} r_1 \\ r_2 \\ r_2 \\ \vdots \\ r_{N-1} \end{bmatrix} = \begin{bmatrix} r(0) \\ r(\Delta) \\ r(2\Delta) \\ \vdots \\ r((N-1)\Delta) \end{bmatrix}. \tag{10.8}$$

This notation will be used throughout the remainder of this chapter. That is, the vector of samples of one period of a waveform $x(t)$ will be denoted x, where $x \in \mathcal{R}^{N\times 1}$.

A new signal y is sought that is equal to r at an arbitrary set of sample indices S and free to vary elsewhere. The free part of the signal is varied to minimize the quadratic cost $y^T Hy$. That is, we seek y that is the solution to

$$\begin{aligned} y &= \arg\min_x x^T Hx, \\ \text{subject to } x_k &= r_k \quad k \in S, \end{aligned} \tag{10.9}$$

where $x \in \mathcal{R}^{N\times 1}$ and $H \in \mathcal{R}^{N\times N}$. Problem (10.9) is equivalent to the linearly constrained convex quadratic optimization problem (Fletcher 1987)

$$\begin{aligned} y &= \arg\min_x x^T Hx + 2f^T x, \\ \text{subject to } Ax &= r(S), \end{aligned} \tag{10.10}$$

where A is the selection matrix representing S and $r(S)$ is a row vector containing the samples of r_n indexed by the values of S.

The solution to problem (10.10) can be stated in matrix form as Fletcher (1987)

$$\begin{bmatrix} H & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} -f \\ r(S) \end{bmatrix}, \tag{10.11}$$

where λ are the Lagrange multipliers (Fletcher 1987).

A solution to (10.11) may be obtained by

$$\begin{bmatrix} y \\ \lambda \end{bmatrix} = \begin{bmatrix} H & A^T \\ A & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} -f \\ r(S) \end{bmatrix}, \tag{10.12}$$

provided the above matrix inverse exists.

To this end, we observe that A has full row rank (since it is constructed as rows of the identity matrix) and $AA^T = I$, so that A^T forms a basis for the row space of A. Let Z be defined as the matrix formed from the rows of the identity matrix that are not present in A, e.g., if $N = 5$ and $S = \{2, 3\}$, then

$$A = \begin{bmatrix} 0\,1\,0\,0\,0 \\ 0\,0\,1\,0\,0 \end{bmatrix},$$
$$Z = \begin{bmatrix} 1\,0\,0\,0\,0 \\ 0\,0\,0\,1\,0 \\ 0\,0\,0\,0\,1 \end{bmatrix}.$$

Note that $AZ^T = 0$ and that the rows of Z form a basis for the null space of A. Therefore, according to Fletcher (1987) (pp. 231–237), the inverse in (10.12) exists and problem (10.10) has a unique minimizer if ZHZ^T is positive definite. While this condition may be difficult to prescribe, it is easily checked. Indeed, for all the examples presented here, this condition was satisfied.

10.3 Frequency Domain Cost Functions

The weighting matrix H can be chosen so that the quadratic cost $x^T Hx$ represents a wide variety of frequency domain cost functions, for example, frequency-weighted-power. Techniques for selecting H follow.

10.3.1 Background: Discrete Fourier Series

The discrete Fourier series c_k of a periodic signal r_n is described by the analysis function (Proakis and Manolakis 2007)

$$c_k = \frac{1}{N} \sum_{n=0}^{N-1} r_n \, e^{-jn\frac{2\pi k}{N}}. \tag{10.13}$$

The synthesis function is (Proakis and Manolakis 2007)

$$r_n = \sum_{k=0}^{N-1} c_k \, e^{jn\frac{2\pi k}{N}}, \tag{10.14}$$

where $\widehat{\omega} = \frac{2\pi k}{N}$ is the normalized frequency, and $\frac{2\pi}{N}$ is the normalized fundamental frequency. The real frequency in Hertz is related to $\widehat{\omega}$ by $f = \frac{\widehat{\omega}}{2\pi\Delta}$. As an example, the discrete Fourier components of an 8 sample signal are shown in Fig. 10.4.

The discrete Fourier coefficients of r can be written in matrix notation:

$$c = \frac{1}{N} E\, r, \text{ where} \tag{10.15}$$

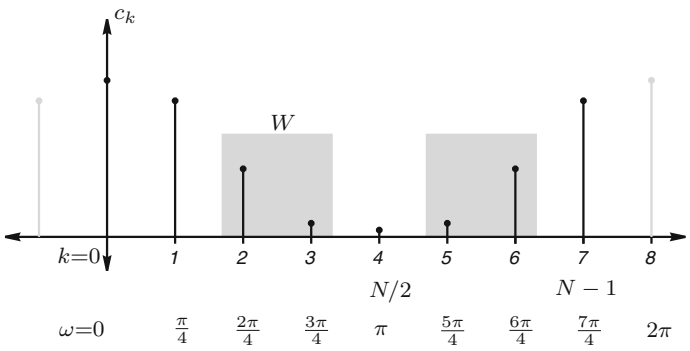


Fig. 10.4 The Fourier components c_k of r

$$c = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{N-1} \end{bmatrix}, \; r = \begin{bmatrix} r_0 \\ r_1 \\ r_2 \\ \vdots \\ r_{N-1} \end{bmatrix} \text{ and}$$

$$E = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ 1 & e^{-j\frac{2\pi 1}{N}} & e^{-j2\frac{2\pi 1}{N}} & \cdots & e^{-j(N-1)\frac{2\pi 1}{N}} \\ 1 & e^{-j\frac{2\pi 2}{N}} & e^{-j2\frac{2\pi 2}{N}} & \cdots & e^{-j(N-1)\frac{2\pi 2}{N}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & e^{-j\frac{2\pi(N-1)}{N}} & e^{-j2\frac{2\pi(N-1)}{N}} & \cdots & e^{-j(N-1)\frac{2\pi(N-1)}{N}} \end{bmatrix}.$$

10.3.2 Minimizing Signal Power

By Parseval's equality, the average power P_r of a discrete time signal r is

$$P_r = \frac{1}{N}\sum_{n=0}^{N-1} |r_n|^2 = \sum_{k=0}^{N-1} |c_k|^2 = \|c_k\|_2^2, \tag{10.16}$$

where the sequence $|c_k|^2$ for $k \in \{0, 1, 2, \ldots, N-1\}$ is the distribution of power as a function of frequency, or the power spectral density. This can be written in vector form:

$$P_r = c^* c \tag{10.17}$$
$$= \frac{1}{N^2} r^T E^* E r,$$

thus, referring to Eq. (10.9), minimum power is achieved when

$$H = \frac{1}{N^2} E^* E. \tag{10.18}$$

10.3.3 Minimizing Frequency Weighted Power

In Fig. 10.4, a frequency dependent weighting W is shown. The power resident in the shaded bandwidth can be calculated by summing only these components. W must be symmetric around π.

We wish to specify a cost function in Eq. (10.9) that represents power above a certain frequency or harmonic. This allows complete freedom in signal power up to the Kth harmonic while imposing a power penalty at higher frequencies. The frequency weighted power P_r^W of r is:

$$P_r^W = \frac{1}{N^2} r^T E^* W E r, \tag{10.19}$$

where $W = diag(Q)$ and

$$Q = \begin{cases} 0 & k \in [0 \ldots K] \\ 1 & k \in [K+1 \ldots N-K-1] \\ 0 & k \in [N-K \ldots N-1] \end{cases},$$

thus, referring to Eq. (10.9), minimum frequency-weighted-power is achieved when

$$H = \frac{1}{N^2} E^* W E. \tag{10.20}$$

It is worth mentioning that frequency-weighted-power signals are not band-limited. Rather, a frequency-weighted-power signal contains the least possible power above a certain frequency with the imposed time domain constraints. If perfect band-limiting is desired, the Fourier coefficients above $k = K$ can be removed via the discrete Fourier Transform and its inverse. The consequences of such filtering, namely the addition of ripple and curvature, are discussed in Sect. 10.1.1.3. The root-mean-square error as a result of filtering is also quantified in (10.7). As the frequency-weighted-power signal contains the least power above the Kth harmonic, if the signal is then band-limited, the resulting signal has the least possible root-mean-square error (10.7). In applications where band-limiting is required, this is an important result. Restated, frequency-weighted-power signals suffer the least possible distortion when perfectly band-limited.

10.3.4 Minimizing Velocity and Acceleration

The use of frequency dependent weighting matrices in Sect. 10.3.3 can also be extended for weighting velocity or acceleration. The Fourier transform of the ith order derivative or integral of $x(t)$ is $(j\omega)^i X(j\omega)$ where i is positive for differentiation and negative for integration.

Rather than calculating the Fourier series of r in Eq. (10.15), we can calculate the Fourier series of its derivatives and integrals. The Fourier coefficients of the differentiated or integrated signal are

$$c = \frac{1}{N} D\, E\, r, \tag{10.21}$$

where E and r are defined in Eq. (10.15), $D = diag(Q)$ and

$$Q = \begin{cases} (jk\frac{F_s}{N})^i & k \in [0 \ldots N/2] \\ (j(N-k)\frac{F_s}{N})^i & k \in [N/2+1 \ldots N-1] \end{cases}, \tag{10.22}$$

This can be simplified to

$$c = \left(\frac{F_s}{N}\right)^i \frac{1}{N} \widetilde{D} E r, \tag{10.23}$$

where $\widetilde{D} = diag(\widetilde{Q})$ and

$$\widetilde{Q} = \begin{cases} (jk)^i & k \in [0 \ldots N/2] \\ (j(N-k))^i & k \in [N/2+1 \ldots N-1] \end{cases}, \tag{10.24}$$

The average power P_r^i in the chosen ith derivative or integral of r is

$$P_r^i = \left(\frac{F_s}{N}\right)^{2i} \frac{1}{N^2} r^T E^* \widetilde{D}^* \widetilde{D} E r. \tag{10.25}$$

Thus, referring to Eq. (10.9), minimum velocity or acceleration is achieved when $i = 1$ or 2, respectively, and

$$H = \left(\frac{F_s}{N}\right)^{2i} \frac{1}{N^2} E^* \widetilde{D}^* \widetilde{D} E. \tag{10.26}$$

Analogous to Sect. 10.3.3, we can also consider a frequency weighted version of P_r^i,

$$P_r^{i,W} = \left(\frac{F_s}{N}\right)^{2i} \frac{1}{N^2} r^T E^* \widetilde{D}^* W \widetilde{D} E r. \tag{10.27}$$

Referring to Eq. (10.9), minimum frequency-weighted velocity or acceleration is achieved when

$$H = \left(\frac{F_s}{N}\right)^{2i} \frac{1}{N^2} E^* \widetilde{D}^* W \widetilde{D} E. \tag{10.28}$$

10.3.5 Single-Sided Frequency Domain Calculations

Real valued signals with an even number of samples have a symmetric spectrum about the Nyquist frequency. The problem size of Eq. (10.9) can be significantly reduced by considering only one half of the spectrum. The signal power is simply twice the sum contained in each half spectrum. That is,

$$P = 2P_{0:N/2} - P_{N/2}, \tag{10.29}$$

where the additional $P_{N/2}$ term is due to the power at the Nyquist rate only occurring once. The error in neglecting this additional term becomes smaller as the number of samples increases. For large N it is sufficient to approximate

$$P = 2P_{0:N/2}. \tag{10.30}$$

Using this simplification, the E, D, $\widetilde{D}$ and W need only be computed for $k = 0$ to $N/2$.

10.4 Time Domain Cost Function

In addition to the frequency domain objectives discussed in the previous section, the quadratic cost in Eq. (10.9) can also represent a function of time. This is useful for incorporating FIR weighting functions used in previous trajectory design techniques. The time domain approach is also numerically robust when specifying optimizations that include a weighting on signal derivatives, for example velocity and acceleration.

The time domain cost function is defined as the output power of an FIR filter whose input is y. That is, we seek to minimize:

$$z_n = \frac{1}{N} \sum_{n=0}^{N-1} \left| B(q^{-1}) y_n \right|^2 \tag{10.31}$$

where $B(q^{-1})$ is an FIR filter of order N_B and length $N_B + 1$.

In matrix form, $z_n = B(q^{-1})y_n$ can be written as

$$z = \mathbf{B}\, y, \text{ where} \tag{10.32}$$

$$z = \begin{bmatrix} z_{0+N_b} \\ z_{1+N_b} \\ z_{2+N_b} \\ \vdots \\ z_{N-1} \end{bmatrix}, y = \begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \\ y_{N-1} \end{bmatrix} \text{ and}$$

$$\mathbf{B} = \begin{bmatrix} b_{N_B} & \cdots & b_1 & b_0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & b_{N_B} & \cdots & b_1 & b_0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & b_{N_B} & \cdots & b_1 & b_0 & 0 & 0 & 0 & 0 \\ & & & & & & & & & \\ 0 & 0 & 0 & 0 & 0 & 0 & b_{N_B} & \cdots & b_1 & b_0 \end{bmatrix}.$$

The power in z is

$$\begin{aligned} \frac{1}{N}\sum_{n=0}^{N-1}\left|B(q^{-1})y_n\right|^2 &= \frac{1}{N}\left\|B(q^{-1})y_n\right\|_2^2 \\ &= \frac{1}{N}z^T z \\ &= \frac{1}{N}y^T\mathbf{B}^T\mathbf{B}y. \end{aligned} \tag{10.33}$$

Thus, referring to Eq. (10.9), the power in z is minimized when

$$H = \frac{1}{N}\mathbf{B}^T\mathbf{B}. \tag{10.34}$$

where $\mathbf{B}$ is the matrix of FIR filter coefficients described in (10.32).

10.4.1 Minimum Velocity

The discrete velocity of y_n is the first-order time derivative

$$\frac{\mathrm{d}y_n}{\mathrm{d}t} = \frac{y_n - y_{n-1}}{\Delta}. \tag{10.35}$$

Thus, the FIR filter that represents differentiation is

$$B(q^{-1}) = \frac{1}{\Delta}(1 - 1q^{-1}). \tag{10.36}$$

This filter can be used in the time domain cost function (10.34) to penalize velocity. The filter coefficients are $b_0 = 1$ and $b_1 = -1$.

10.4.2 Minimum Acceleration

The discrete acceleration of y_n is the second-order time derivative

$$\begin{aligned}\frac{\mathrm{d}^2 y_n}{\mathrm{d}t^2} &= \frac{1}{\Delta}\left(\frac{\mathrm{d}y_n}{\mathrm{d}t} - \frac{\mathrm{d}y_{n-1}}{\mathrm{d}t}\right) \\ &= \frac{(y_n - y_{n-1}) - (y_{n-1} - y_{n-2})}{\Delta^2} \\ &= \frac{y_n - 2y_{n-1} + y_{n-2}}{\Delta^2}.\end{aligned} \tag{10.37}$$

Thus, the FIR filter that represents double differentiation is

$$B(q^{-1}) = \frac{1}{\Delta^2}(1 - 2q^{-1} + 1q^{-2}). \tag{10.38}$$

This filter can be used in the time domain cost function (10.34) to penalize acceleration. The filter coefficients are $b_0 = 1$, $b_1 = -2$ and $b_2 = 1$.

10.4.3 Frequency Weighted Objectives

Analogous to the frequency weighted cost functions in Sect. 10.3, time domain cost functions can also be subjected to frequency domain weightings, however, the process is less direct.

Frequency weighted power can be achieved by using the filter $B(q^{-1})$ to implement the desired frequency weighting. In this case, the quadratic cost H representing power at the output of the filter is described in Eqs. (10.33) and (10.34). If the filter $B(q^{-1})$ has already been utilized, for example to specify velocity or acceleration, a frequency weighting can still be applied by generating a second filter $B_2(q^{-1})$, whose frequency response represents the desired weighting, and convolving the two, i.e.,

$$B(q^{-1}) = B_1(q^{-1}) \otimes B_2(q^{-1}) \tag{10.39}$$

where $B(q^{-1})$ is the filter used in the cost function (10.34), $B_1(q^{-1})$ is the filter used for example to specify velocity, and $B_2(q^{-1})$ is the frequency weighting filter.

10.5 Application to Scan Generation

In periodic scanning applications, it is desirable to scan as quickly as possible without exciting mechanical resonance. In other words, an input signal is required that contains the least possible power at frequencies near and above the first mechanical

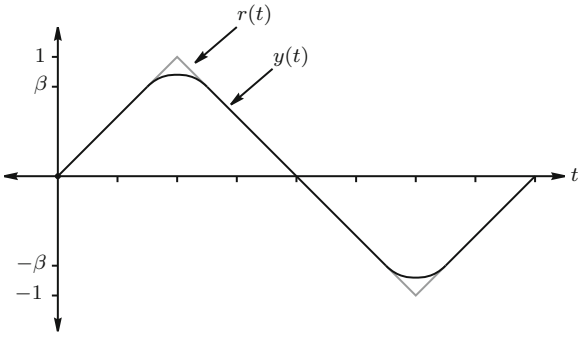


Fig. 10.5 The reference and optimal trajectory, $r(t)$ and $y(t)$. The optimal signal is equal to $r(t)$ when $r(t) < |\beta|$, otherwise there is no restriction

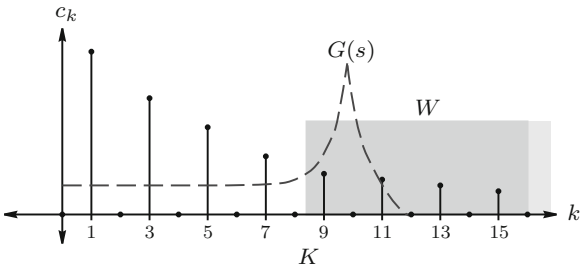


Fig. 10.6 The Fourier components of a triangular scanning signal plotted against harmonic number k. The optimal signal is unrestricted in spectral content between DC and the Kth harmonic. All harmonics greater than K are penalized to avoid excitation of the system G

resonance. This objective is satisfied by the frequency-weighted-power cost function described in Sect. 10.3.3. The resulting trajectory contains the least possible power above a certain frequency while maintaining perfect scanning over a portion of the range.

For triangular and sawtooth scanning waveforms, the linear range is easily specified by a single parameter β. Referring to Fig. 10.5, the optimal trajectory y_k is equal to r_k when $r_k < |\beta|$, otherwise there is no restriction. Using the notation in Sect. 10.2, the previous statement can be rewritten as $y(S) = r(S)$ where S is the set of sample indices for which $r_k < |\beta|$.

To specify the frequency weighting, it is convenient to stipulate the number of unrestricted low-frequency harmonics that may appear in the optimal signal. The spectrum of a triangular scanning signal is shown in Fig. 10.6. The frequency components of the optimal signal are unrestricted between DC and the Kth harmonic. All harmonics greater than K are penalized equally.

A Matlab function that generates and simulates optimal scanning signals, named `generateTriangle`, is available by contacting the first author.

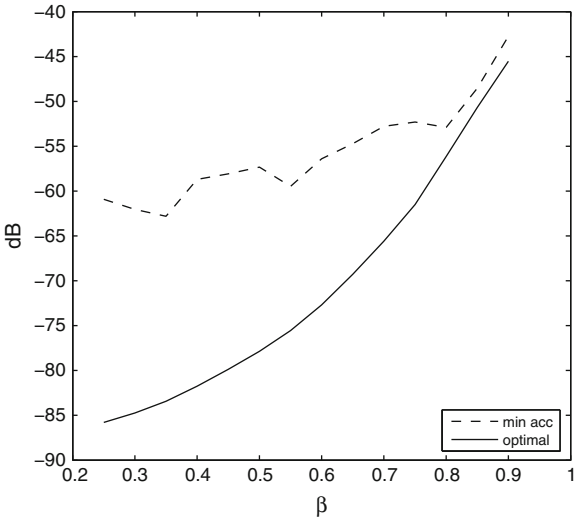


Fig. 10.7 Absolute sum of the first 50 out of bandwidth Fourier coefficients versus scan range β ($K = 9$)

10.5.1 Choosing β and K

When using the frequency weighted power objective, frequency content above the cutoff is minimized by decreasing β and increasing K. If either parameter is fixed, the other can be varied to reduce scan error to an arbitrary value.

Assuming the allowable bandwidth is known, two possible scenarios arise when considering the choice of β and K, these are:

1. The error and scan range are fixed. What is the maximum scan frequency? This is characteristic of most practical circumstances where scan range and precision are more highly valued than frequency. The scan frequency is simply reduced to a point where the number of in-bandwidth harmonics are sufficient to satisfy the error criterion.
2. The error and scan frequency are fixed. What is the maximum scan range β? This case arises in high-speed applications where scan range is sacrificed for increased frequency. Given the number of allowable harmonics, e.g., 3, β is reduced until the error is satisfactory. If the resulting scan range in impractically small, the scan frequency must be revised.

Both these scenarios are easily resolved by plotting the free parameter versus error.

In general-purpose applications where no fixed limit on frequency or scan range exists, some insight can be gained by plotting the high-frequency signal content versus the scan range β and number of harmonics K as shown in Figs. 10.7 and 10.8.

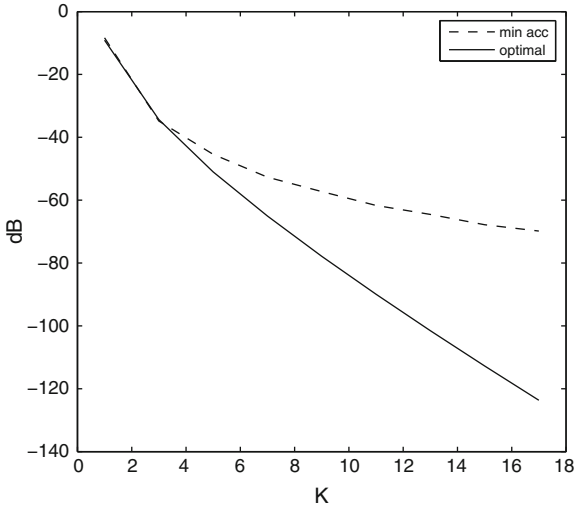


Fig. 10.8 Absolute sum of the first 50 out of bandwidth Fourier coefficients versus the number of included harmonics K ($\beta = 0.5$)

Here, the high-frequency signal power is defined as the absolute sum of the first 50 harmonic components above K.

If scan range is valued highly, a good choice for β is 0.7, which provides approximately the maximum scan range before high-frequency content significantly increases. Beyond $\beta = 0.8$ there is little difference between the optimal, and minimum acceleration signals. If β is chosen fairly large ($\beta \geq 0.7$), the scan error must be minimized by including a large number of harmonics. For example, if the scan frequency is one-twentieth of the mechanical resonance frequency, K can be chosen up to 19. In Fig. 10.8 it is clear that K=19 will provide a very high degree of performance.

If scan speed is highly valued, K must be small. If the scan speed is 10 % the resonance frequency, K must be nine or less. The smallest reasonable value for K is five, which allows only three sine waves in the optimal signal and scan speeds up to 20 % the resonance frequency. In such cases, β must be severely reduced to minimize induced vibration. In Fig. 10.7, reducing β to 0.3 can provide excellent performance at ultra high speed.

The authors recommend two general-purpose choices for β and K:

- $\beta = 0.7$ and $K = 9$. This provides good scan range, operation up to 10 % of the mechanical resonance frequency, and a reasonable minimum of induced vibration. Slower scan speeds with higher K improve performance.
- $\beta = 0.5$ and $K = 5$ or 7. This is more suitable for high performance scanning where scan frequency approaches 20 % of the resonance frequency. Vibration can be reduced by further reducing β to 0.4 or less.

10.5.2 Improving Feedback and Feedforward Controllers

10.5.2.1 Feedback

In addition to improving the performance of open-loop scanners, optimized input signals are also useful as reference commands for feedback control loops. As tracking control loops are typically limited in bandwidth to around one-tenth that of the open-loop system, the frequency content of reference commands must be strictly conserved if tracking error is to be kept low.

Further limitations arise in many electromechanical systems that exhibit nonlinearity such as hysteresis. In these systems, high controller loop-gain is required to attenuate tracking error. In integral control loops, significant loop-gain is only available one-decade below the closed-loop bandwidth. Thus, the system should only be driven by reference commands that contain frequency components significantly lower than the closed-loop bandwidth, which is typically only a fraction of the first resonance frequency. In such cases, an optimized reference trajectory can provide the best utilization of the small bandwidth available.

In more general circumstances, reference commands with lower high-frequency content relax the close-loop bandwidth requirement. This, in turn, requires less controller gain, resulting in greater robustness and less feed-through of sensor noise to the regulated variable.

10.5.2.2 Feedforward

In systems using inversion-based feedforward control, as discussed in Chap. 9, the choice of reference signal is critical. Wide bandwidth input signals have spectral components at frequencies where the inversion filter can be highly sensitive to modeling error (Devasia 2002). Sensitivity to modeling error can be reduced if the reference signal has minimal harmonic content in the bandwidth where inversion is required (Devasia 2002). The frequency-weighted-power signal, discussed in Sect. 10.5, is such an input that contains minimum high-frequency power and can provide the greatest immunity to modeling error.

Frequency-weighted-power signals also minimize control signal magnitude by avoiding frequencies where the plant response is small. This is highly advantageous in iterative systems that achieve near perfect inversion (Wu and Zou 2007). If the internal reference signal contains frequency components at, or near, plant zeros, extremely large inputs are generated in compensation. Frequency-weighted-power signals that contain minimal high-frequency harmonics can greatly reduce this problem.

10.6 Comparison to Other Techniques

As discussed in the Introduction, a number of techniques have been proposed for minimizing induced vibration in mechanical scanners. In this section, these techniques are compared to the frequency-weighted-power signal discussed in the previous section.

A simple scanner model is considered with two resonances, one at 10 Hz and another at 100 Hz. The transfer function is:

$$G(s) = \frac{0.7\omega_1^2}{s^2 + 2\omega_1\xi_1 s + \omega_1^2} + \frac{0.3\omega_2^2}{s^2 + 2\omega_2\xi_2 s + \omega_2^2}, \tag{10.40}$$

where $\omega_1 = 2\pi 10$, $\omega_2 = 2\pi 100$ and $\xi_1 = \xi_2 = 0.01$. The frequency response of $G(s)$ is plotted in Fig. 10.9b.

It is desirable to operate the scanner at one-tenth the resonance frequency, i.e., 1 Hz. The five input signals under consideration are the:

1. *Triangle signal*. A 1 Hz, unity amplitude triangle wave with a linear range of ± 1.
2. *Filtered-triangle*. A triangle signal, noncausally filtered by the minimum order Butterworth frequency response that achieves less than 3 dB ripple below 7 Hz and more than 80 dB attenuation at 9 Hz. The linear range is ± 0.75.
3. *Shaped-triangle*. A triangle signal with 0.05 s flat area at each apex as described in Schitter et al. (2006). This signal provides excellent performance if the resonance frequency is known and the mechanical system is second order. The performance degrades if the resonance frequency is not exactly known or the system order is greater than two. The linear range is ± 0.9.
4. *Minimum acceleration (Min. Acc.)*. The minimum acceleration trajectory with a linear scan range of ± 0.5 ($\beta = 0.5$).
5. *Optimal*. The frequency-weighted-power signal with a linear scan range of ± 0.5 ($\beta = 0.5$) and $K = 7$ as described in Sect. 10.5.

The five input signals under consideration are plotted in Fig. 10.9a. When applied to the example system $G(s)$, the resulting output and corresponding error are shown in Fig. 10.9c, d. To summarize the results, root-mean-square errors are presented in Table 10.1. The frequency-weighted-power signal is observed to outperform other techniques by between 8 and 400 times. In Table 10.1, results from a second simulation where the resonance frequency is reduced by 10 % are also reported. While the shaped-triangle signal performs well in the nominal simulation, it is not robust to changes in the resonance frequency. This is due to its dependency on the scanner resonance. In contrast, the frequency-weighted-power signal is not model-based and performs well when the resonance frequency is not known or prone to variation. Further insight can be gained by considering Fig. 10.9e where the Fourier coefficients of the triangle, minimum-acceleration, and frequency-weighted-power signal are plotted. Clearly, after the seventh harmonic, the frequency-weighted-power coefficients drop to extremely small magnitudes. Hence, variations in system dynamics after the seventh harmonic have little effect on the tracking error.

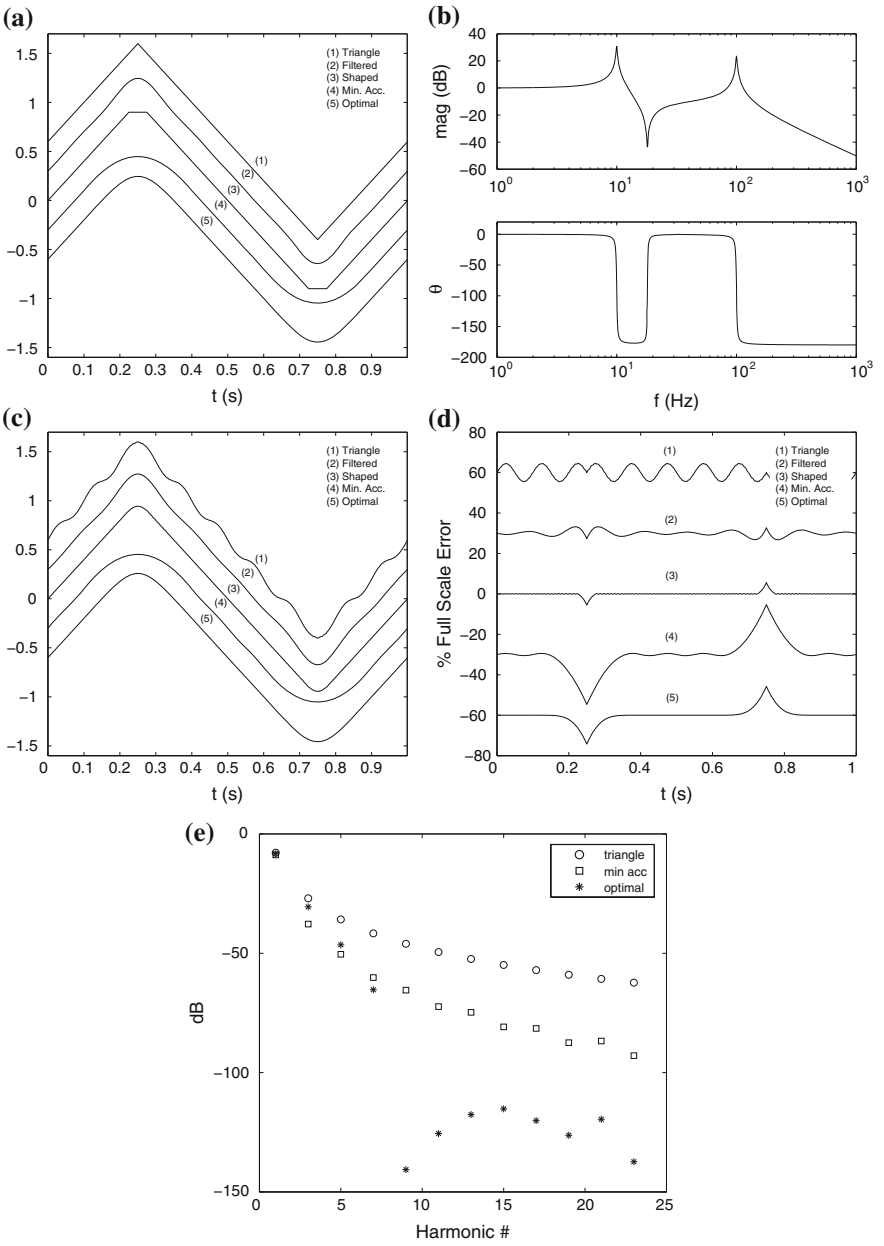


Fig. 10.9 Comparison of different input shaping techniques. **a** Input signals. **b** Frequency response. **c** Resulting displacement. **d** Error. **e** Fourier coefficients

Table 10.1 Simulated root-mean-square error between the outputs and a triangle wave (calculated in the time range where the optimal signal is linear)

Resonance FrEquation	10 Hz (%)	9 Hz (%)
Triangle	3.1	1.2
Filtered-Triangle	0.53	0.77
Shaped-Triangle	0.062	0.46
Min. Acc.	0.39	0.13
Optimal	0.0075	0.011

Two cases are considered, one where the resonance frequency is 10 Hz, and another where the resonance frequency is reduced by 10 % to 9 Hz

10.7 Experimental Application

The P-734 nanopositioner, described in Sect. 3.2.1, is a typical two-axis nanopositioner commonly used in many forms of scanning probe microscopy. Although such devices can achieve high precision with millimeter range motion, the internal displacement amplifiers, large piezoelectric stacks, and large platform mass contribute to a low mechanical resonance frequency. The P-734 stage has a range of 100 microns but a resonance frequency of only 420 Hz. The frequency response of a single axis is plotted in Fig. 10.10b. The unity gain bandwidth extends from DC to around 140 Hz where a phase and magnitude shift of 5° and 1 dB exists. Above this frequency the phase and magnitude response degrade rapidly. To achieve accurate scanning in open-loop, the input signal spectrum should be retained to within 140 Hz.

Without using model-based inversion, the fastest practical scan speed for the platform under consideration is around 20 Hz. In this case, the 3rd, 5th, and 7th harmonics occur at 60, 100, and 140 Hz. An optimal signal can be designed to achieve high scan range with minimal harmonic content above 140 Hz, this implies $\beta = 0.5$ and $K = 7$. With a sampling rate of 20 kHz (1,000 points per period), the 20 Hz optimal input signal can be generated with the command: `generateTriangle(20000,20,0.5,7)`. This signal and the other signals discussed in Sect. 10.6 were applied to develop a scan with 13 micron linear range. As the choice of β and K is identical to that in Sect. 10.6, Fig. 10.9e also pertains to the signals here.

The resulting displacement and difference to an ideal triangle wave is plotted in Fig. 10.10c, d. The performance is summarized in Table 10.2. Although the frequency-weighted-power signal outperforms other techniques, the magnitude of the error is significantly greater than expected from the spectra plotted Fig. 10.9e. The difference is due to the presence of measurement noise and piezoelectric hysteresis that set a minimum bound on the achievable error.

In general, piezoelectric hysteresis will be worsened if the optimization increases peak signal amplitude and *vice-versa*.

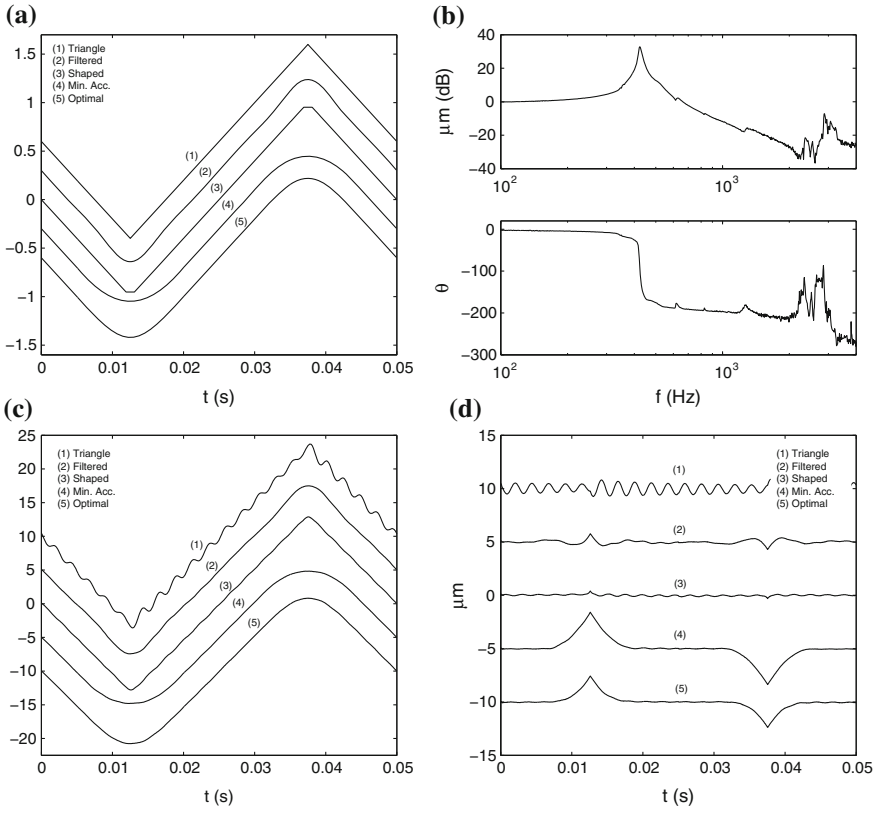


Fig. 10.10 The 20 Hz input signals in (**a**) were experimentally applied to the scanner with frequency response shown in (**b**). The resulting displacement and corresponding error is plotted in (**c**) and (**d**)

Table 10.2 Experimental root-mean-square error between an ideal triangle wave and the measured output (calculated in the time range where the optimal signal is linear)

Resonance Fr	420 Hz (%)	350 Hz (%)
Triangle	2.9	3.87
Filtered-triangle	0.61	0.72
Shaped-triangle	0.56	1.1
Min. Acc.	0.21	0.60
Optimal	0.18	0.22

Two cases are considered, one where the scanner is unloaded, and another where a sample plate is added that reduces the resonance frequency by 16 % to 350 Hz

10.8 Chapter Summary

This chapter describes scanning trajectories for band-limited systems that minimize a frequency or time domain cost function while enforcing linearity over a certain range. Specific cost functions include minimum velocity, acceleration, or power. These are easily combined to achieve multiple objectives, and/or subjected to frequency domain weighting.

The frequency-weighted-power objective was introduced to maximize the scanning performance of band-limited systems. It enforces linearity over a certain range ($\pm\beta$) while minimizing signal power above a chosen frequency. The key advantages of the frequency-weighted-power signal are:

- Perfect linearity over a certain range ($\pm\beta$).
- Minimum frequency content above the chosen Kth harmonic.
- β and K can be varied to achieve arbitrarily low oscillation.
- Simplifies and improves the performance of feedforward and feedback control systems.

The frequency-weighted-power signal outperforms present techniques in simulation and experiment on a standard nano-positioning platform. Even with conservative values of β and K, an order of magnitude improvement in induced oscillation can be achieved. This improvement increases dramatically as scan range is sacrificed, or more harmonics are allowed.

References

Abramovitch DY, Andersson SB, Pao LY, Schitter G (2007) A tutorial on the mechanisms, dynamics, and control of atomic force microscopes. In: Proceedings of American control conference. New York City, pp 3488–3502

Aphale SS, Fleming AJ, Moheimani SOR (2007) Integral control of resonant systems with collocated sensor-actuator pairs. IOP Smart Mater Struct 16:439–446

Aphale SS, Fleming AJ, Moheimani SOR (January 2007) High speed nano-scale positioning using a piezoelectric tube actuator with active shunt control. IET Micro Nano Lett 2(1):9–12

Aphale SS, Bhikkaji B, Moheimani SOR (2008) Minimizing scanning errors in piezoelectric stack-actuated nanopositioning platforms. IEEE Trans Nanotechnol 7(1):79–90

Croft D, Shed G, Devasia S (2001) Creep, hysteresis, and vibration compensation for piezoactuators: atomic force microscopy application. Trans ASME J Dyn Syst Meas Control 123:35–43

Devasia S, Eleftheriou E, Moheimani SOR (2007) A survey of control issues in nanopositioning. IEEE Trans Contro Syst Technol 15(5):802–823

Devasia S (2002) Should model-based inverse inputs be used as feedforward under plant uncertainty? IEEE Trans Autom Control 47(11):1865–1871

Dewey JS, Leang KK, Devasia S (1998) Experimental and theoretical results in output-trajectory redesign for flexible structures. ASME J Dyn Syst Meas Control 120:456–461

Fanson JL, Caughey TK (1990) Positive position feedback control for large space structures. AIAA J 28(4):717–724

Fleming AJ, Moheimani SOR (2006) Sensorless vibration suppression and scan compensation for piezoelectric tube nanopositioners. IEEE Trans Control Syst Technol 14(1):33–44

Fleming AJ, Wills AG (2009) Optimal periodic trajectories for band-limited systems. IEEE Trans Control Syst Technol 13(3):552–562
Fletcher R (1987) Practical methods of optimization. Wiley, Chichester
Kim K, Zou Q (2008) Model-less inversion-based iterative control for output tracking: piezo actuator example. In: American control conference. Seattle, pp 2710–2715
Leang KK, Devasia S (2007) Feedback-linearized inverse feedforward for creep, hysteresis, and vibration compensation in afm piezoactuators. IEEE Trans Control Syst Technol 15(5):927–935
Li Y, Bechhoefer J (2008) Feedforward control of a piezoelectric flexure stage for AFM. In: American control conference. Seattle, pp 2703–2709
Masterson RA, Singhose WE, Seering WP (2000) Setpoint generation for constant-velocity motion of space-based scanners. AIAA J Guidance Contro Dyn 23(5):892–895
Meyer E, Hug HJ, Bennewitz R (2004) Scanning probe microscopy. The lab on a tip. Springer, Heidelberg
Pao LY, Butterworth JA, Abramovitch DY (2007) Combined feedforward/feedback control of atomic force microscopes. In: Proceedings of American control conference. New York, pp 3509–3515
Perez H, Zou Q, Devasia S (March 2004) Design and control of optimal scan trajectories: scanning tunneling microscope example. J Dyn Syst Meas Control 126:187–197
Potsaid B, Wen JT, Unrath M, Watt D, Alpay M (2007) High performance motion tracking control for electronic manufacturing. ASME J Dyn Syst Meas Control 129:767–776 (mirror Scanner)
Proakis JG, Manolakis DG (2007) Digital signal processing, 4th edn. Pearson Education Inc., New Jersey
Rost MJ, Crama L, Schakel P, van Tol E, van Velzen-Williams GBEM, Overgauw CF, ter Horst H, Dekker H, Okhuijsen B, Seynen M, Vijftigschild A, Han P, Katan AJ, Schoots K, Schumm R, van Loo W, Oosterkamp TH, Frenken JWM (2005) Scanning probe microscopes go video rate and beyond. Rev Sci Instrum 76(5):053 710-1-053 710–9
Schitter G, Fantner GE, Thurner PJ, Adams J (2006) Design and characterisation of a novel scanner for high-speed atomic force microscopy. In: IFAC symposium on mechatronic systems. Heidelberg, pp 819–824
Schitter G, Stemmer A (2004) Identification and open-loop tracking control of a piezoelectric tube scanner for high-speed scanning-probe microscopy. IEEE Trans Control Syst Technol 12(3): 449–454
Singer NC, Seering WP (1990) Preshaping command inputs to reduce system vibration. ASME J Dyn Syst Meas Control 112(2):76–82
Singhose W, Singer N, Seering W (1995) Comparison of command shaping methods for reducing residual vibration. In: Proceedings of European control conference, Rome
Vaughan J, Yano A, Singhose W (2008) Performance comparison of robust negative input shapers. In: Proceedings of American control conference. Seattle, pp 3257–3262
Wu Y, Zou Q (2007) Iterative control approach to compensate for both the hysteresis and the dynamics effects of piezo actuators. IEEE Trans Control Syst Technol 15(5):936–944
Zhao Y, Jayasuriya S (1994) Feedforward controllers and tracking accuracy in the presence of plant uncertainties. In: Proceedings of American control conference. Baltimore, pp 360–364
Zou Q, Leang KK, Sadoun E, Reed MJ, Devasia S (2004) Control issues in high-speed AFM for biological applications: collagen imaging example. Asian J Control 6(2):164–176

Chapter 11
Hysteresis Modeling and Control

This chapter focuses on the fundamentals of hysteresis, including modeling and compensation.

11.1 Introduction

Hysteresis, which is a nonlinear behavior between the applied electric field and the mechanical displacement of a piezoelectric actuator, is believed to be caused by irreversible losses that occur when similarly oriented electric dipoles interact upon application of an electric field (Jiles and Atherton 1986). The effect of hysteresis on the displacement of a piezoelectric actuator is more pronounced over large-range motion (Barrett and Quate 1991; Adriaens et al. 2000). The term hysteresis comes from the Greek word "to be late" or "come behind" and it was first coined for application in 1881 by physicist Ewing when he was studying magnetization. Interestingly, the year 1881 was when the Curie Brothers were credited with the discovery of the piezoelectric effect. Hysteresis is often referred to as a lag in the response. An interesting writing on the history of hysteresis can be gleaned from reference (Cross 1988), which describes other systems which exhibit this behavior. The mechanism responsible for hysteresis in piezoelectric transducers is better understood by considering the domain wall analogy for describing hysteresis in magnetic materials (Jiles and Atherton 1986; Cao and Evans 1993). For example, magnetic materials consist of tiny elementary magnetic dipoles. These particles align to an applied field. The analogy to this in piezoelectric materials is the unit cell of the crystal which exhibits an electric dipole. The term *domains of polarization* refers to regions of similarly oriented dipoles, that is, a relatively large region of connected unit cells having similarly oriented net polarization. The imaginary boundary which separates these regions are referred to as *domain walls*. These boundaries grow or shrink depending on the nature of the applied field. For the simple case, an isolated elementary dipole subjected to an applied field will orient itself to the field instantaneously, and therefore

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_11, © Springer International Publishing Switzerland 2014

displays no hysteresis. However, hysteresis is said to arise due to "internal forces," which causes the dipoles to exhibit a preference for their orientation, and hence the motion of the domain walls are retarded by such forces. These internal forces are attributed to material defects and internal friction between dipoles and the domain walls. Although the domain wall analogy was conceived for magnetic materials, it can easily be extended to materials which consist of elementary dipole-like particles, such as piezoelectric materials. Additionally, hysteresis "remembers" the effect of the past, which further complicates the problem in terms of precision control.

11.2 Modeling Hysteresis

Hysteresis is widely accepted as a nonlinear behavior characterized by a nonvanishing input-output loop as the frequency decreases to zero. The behavior sometimes corresponds to energy loss. A wide variety of models have been proposed for hysteresis, but described below are five popular models to describe rate-independent hysteresis. These models have been applied to compensate for the nonlinear effect in PZT ceramic actuators. A detailed discussion of hysteresis can be found in the three-volume collection, "The Science of Hysteresis," edited by Bertiotti and Mayergoyz (Bertotti and Mayergoyz 2006a, b, c).

11.2.1 Simple Polynomial Model

Hysteresis can be modeled with reasonable accuracy with many approaches. The simplest is a polynomial fit of the output versus input data of the form,

$$y = a_n u^n + a_{n-1} u^{n-1} + \cdots + a_1 u + a_0, \tag{11.1}$$

where n is the order of the polynomial, y is the output, u is the input, and the a_i's coefficients can be found using a least-squares algorithm. However, this model ignores the branching (looping) behavior, that is, the ascending and descending branches are the same. Such a model provides reasonable accuracy for electrostrictive actuators as they tend to exhibit significantly less hysteresis compared to piezoelectric actuators (Hu et al. 2004). Inversion of the polynomial hysteresis model for feedforward control is straightforward. A look-up table can be used in this case.

11.2.2 Maxwell Slip Model

A more realistic representation is the Maxwell slip model, a lumped parameter approach (Goldfarb and Celanovic 1997). In this model, the nonlinear input-output

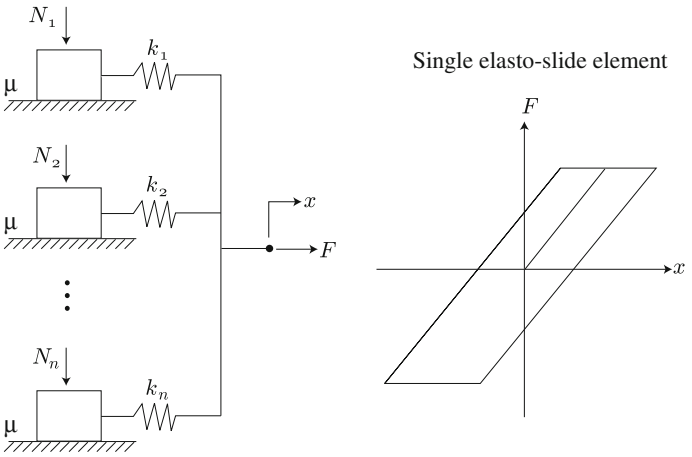


Fig. 11.1 Maxwell slip model for hysteresis

map and looping behavior are caused by a parallel combination of elasto-slide elements as shown in Fig. 11.1. An elasto-slide element consists of a massless linear spring and a massless block that is subjected to Coulomb friction. The constitutive behavior of the ith element is given by

$$F_i = \begin{cases} k_i(x - x_{bi}) & \text{if } |k_i(x - x_{bi})| < f_i \\ f_i \text{sgn}(\dot{x}) & \text{and } x_{bi} = x - \frac{f_i}{k_i}\text{sgn}(\dot{x}) \text{ else,} \end{cases} \tag{11.2}$$

where x is the input displacement, F_i, k_i, f_i, and x_{bi} are the output force, spring stiffness, breakaway force, and block position, respectively, for the ith element. The resultant output force is

$$F = \sum_{i=1}^{n} F_i. \tag{11.3}$$

Starting in a relaxed state, the measured input-output map can be used to determine the individual spring constants and break away forces (hence μ's). The Maxwell slip model has been applied to model and control a piezoelectric stack actuator as described in Goldfarb and Celanovic (1997).

11.2.3 Duhem Model

The rate-independent hysteresis in a PZT actuator can be modeled by a finite-dimensional, differential model which was typically used to model ferromagnetically soft materials (Coleman and Hodgdon 1986). The model is given by

$$\dot{v}(t) = \alpha|\dot{u}(t)|\big[\beta u(t) - v(t)\big] + \gamma\dot{u}(t), \tag{11.4}$$

where $v(t)$ is the output, $u(t)$ is the input, and α, β, γ are positive constants. The parameters of the model can be obtained by a least-squares fit of the measured input-output data. For example, the input $u(t) = A\sin(\omega t)$, with $\omega = 2\pi$ rad/s and $A = 2$ V, was applied to a piezoactuator and the output $v(t)$ was measured. Assuming zero initial conditions, Eq. (11.4) was rewritten in the following matrix form

$$\mathbf{v} = \mathbf{Q}\Theta, \tag{11.5}$$

where $\mathbf{v} = [v(t_1), v(t_2), v(t_3), \cdots]^T$ is the vector of measured outputs at specific time instances; $\Theta = [\alpha\beta, -\alpha, \gamma]^T$ is the vector of unknown parameters; and

$$\mathbf{Q} = \begin{bmatrix} q_1(t_1) & q_2(t_1) & q_3(t_1) \\ q_1(t_2) & q_2(t_2) & q_3(t_2) \\ q_1(t_3) & q_2(t_3) & q_3(t_3) \\ \vdots & \vdots & \vdots \end{bmatrix}, \tag{11.6}$$

where $q_3(t) = u(t)$ and

$$q_1(t) = \int |\dot{u}(t)|u(t)\,\mathrm{d}t; \quad q_2(t) = \int |\dot{u}(t)|v(t)\,\mathrm{d}t. \tag{11.7}$$

The identified parameters were $\alpha = 0.3$, $\beta = 1.2$, and $\gamma = 1.0$. Figure 11.2a compares the normalized measured and model outputs for $u(t) = 2\sin(2\pi t)$ V. The hysteresis curves are compared in Fig. 11.2b. The results show good agreement between the measured and model outputs, where the maximum steady-state error was less than 2 % of the total range.

11.2.4 Preisach Model

A relatively accurate hysteresis model is the Preisach hysteresis model, which was first developed in 1935 for magnetic materials (Preisach 1935). This model has been studied extensively to characterize the rate-independent hysteresis in piezoelectric materials (Ge and Jouaneh 1995; Mayergoyz 1991), as well as many hysteretic systems, such as shape memory alloy devices (Majima et al. 2001). This model can be inverted for feedforward control (Croft et al. 2001); however, finding the forward and inverse model typically involves identifying a large set of parameters.

The Preisach model is a phenomenological description, whereby the output of a hysteretic system is the net effect of elementary relays, which represent the behavior of individual domains within the material. These domains or relays can assume a value of $+1$ or -1 depending on the current and future values of the input. The relay operator $\mathcal{R} : \mathbb{R} \to \{-1, +1\}$ is defined as (Mayergoyz 1991):

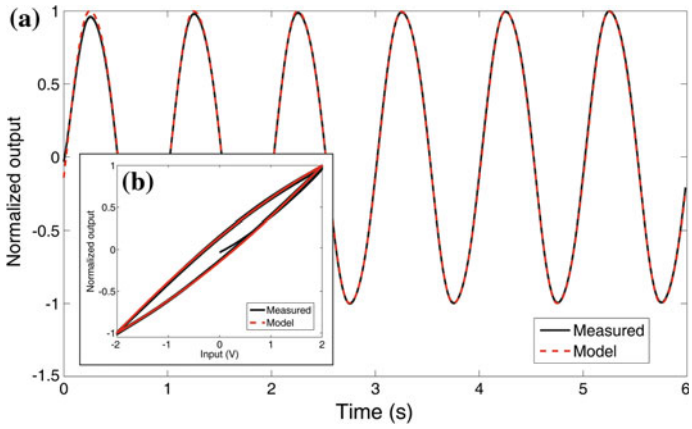


Fig. 11.2 Comparison of measured and Duhem hysteresis model output for a piezoactuator: **a** The normalized outputs versus time and **b** a hysteresis curve. Input was $u(t) = 2\sin(\omega t)$ V, with $\omega = 2\pi$ rad/s

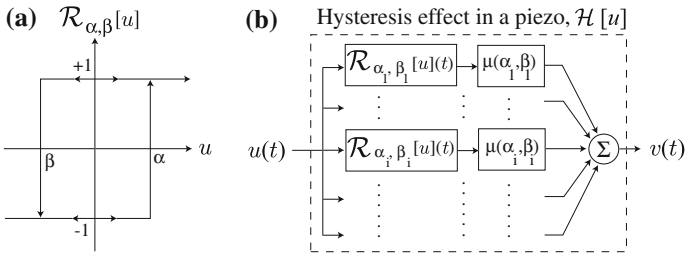


Fig. 11.3 **a** The elementary relay (*Preisach hysteron*) and **b** the output $v(t)$ is the sum of weighted hysterons

$$\mathcal{R}_{\alpha,\beta}[u](t) = \begin{cases} +1 & u(t) > \alpha, \\ -1 & u(t) < \beta, \\ \text{unchanged} & \beta \le u(t) \le \alpha, \end{cases} \tag{11.8}$$

where $u(t)$ is the input. The pair (α, β) in Eq. (11.8), such that $\alpha \ge \beta$, represents the "up" and "down" switching values of the relay, respectively. Figure 11.3a shows an example of an elementary relay (also called Preisach hysteron).

The output $v(t)$, which is an infinite sum of hysterons (Fig. 11.3b), is written as

$$v(t) = \mathcal{H}[u](t) = \iint_{\alpha \ge \beta} \mu(\alpha,\ \beta)\mathcal{R}_{\alpha,\beta}[u](t)\ \mathrm{d}\alpha\ \mathrm{d}\beta, \tag{11.9}$$

where $\mu(\alpha, \beta)$ is called the Preisach weighting function, and (α, β) belongs to the Preisach plane **P**, defined as,

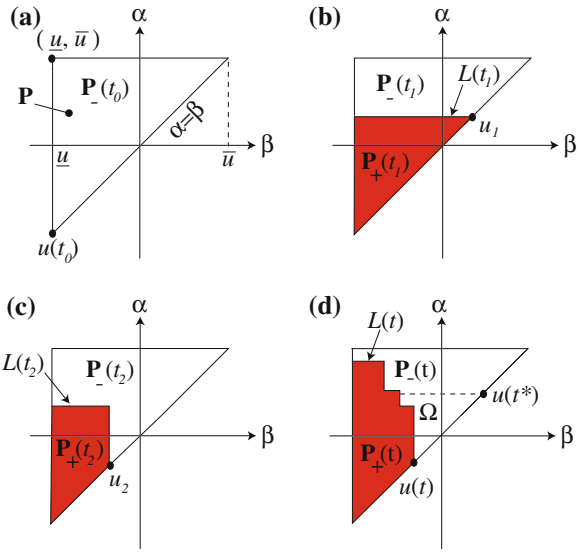


Fig. 11.4 Behavior of the Preisach boundary

$$\mathbf{P} \triangleq \{(\alpha, \beta) | \alpha \geq \beta; \underline{u} \leq \alpha, \beta \leq \bar{u}\}, \tag{11.10}$$

which happens to be the limiting right-triangle region shown in Fig. 11.4a. Equation (11.10) implies that only relays enclosed in the right-triangle region are affected by the input u (cf. Figure 11.4a).

Depending on which relays have been switched to $+1$ or -1, at time t the Preisach plane $\mathbf{P}$ is divided into two regions, *i.e.*,

$$\mathbf{P}_+(t) \triangleq \{(\alpha, \beta) \in \mathbf{P} \ : \ \text{output } \mathcal{R}_{\alpha,\beta}[u](t) = +1\}, \tag{11.11}$$

$$\mathbf{P}_-(t) \triangleq \{(\alpha, \beta) \in \mathbf{P} \ : \ \text{output } \mathcal{R}_{\alpha,\beta}[u](t) = -1\}, \tag{11.12}$$

with $\mathbf{P} = \mathbf{P}_+(t) \cup \mathbf{P}_-(t)$.

To better understand the Preisach model, consider its geometric interpretation. Assume at some time t_0 the input $u(t_0) = \underline{u}$ as shown in Fig. 11.4a. Then, from Eq. (11.8) the output of $\mathcal{R}_{\alpha,\beta}$, $\forall(\alpha, \beta) \in \mathbf{P}$, is -1. As a result, $\mathbf{P}_-(t_0) = \mathbf{P}$ and $\mathbf{P}_+(t_0) = \varnothing$, otherwise known as the state of "negative saturation." Next, assume that the input increases monotonically to an arbitrary maximum value $u_1(t_1)$ at time t_1. All relays with $\alpha < u_1(t_1)$ switch to the $+1$ state, and at time t_1 the boundary separating regions $\mathbf{P}_-(t_1)$ and $\mathbf{P}_+(t_1)$ is a horizontal line as shown in Fig. 11.4b. We denote this boundary by $L(t_1)$. Suppose the input decreases monotonically to an arbitrary value $u_2(t_2) > \underline{u}$ at time t_2. As a result relays $\mathcal{R}_{\alpha,\beta}$ (Eq. (11.8)), with $\beta > u_2(t_2)$, switch to the -1 state, and a vertical line segment $\beta = u_2(t_2)$ is generated as a part of the boundary separating $\mathbf{P}_-(t_2)$ and $\mathbf{P}_+(t_2)$ as shown in Fig. 11.4c.

Further input reversals generate additional horizontal and vertical links, and in general, at time t the boundary $L(t)$ separating regions $\mathbf{P}_-(t)$ and $\mathbf{P}_+(t)$ is a nonincreasing staircase function of β as shown in Fig. 11.4d (Mayergoyz 1991). The last link of the boundary $L(t)$ intersects the line $\alpha = \beta$ at point $\big(u(t), u(t)\big)$. This boundary is referred to as the Preisach *memory curve* as it stores the effect of past input. In fact, it captures the effect of past input extremum (Mayergoyz 1991).

Using the fact that $\mathbf{P} = \mathbf{P}_+(t) \cup \mathbf{P}_-(t)$, the output can be expressed in the following form (Gorbet et al. 1998):

$$v(t) = 2 \iint\limits_{\mathbf{P}_+(t)} \mu(\alpha,\ \beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta - \iint\limits_{\mathbf{P}} \mu(\alpha,\ \beta)\, \mathrm{d}\alpha\, \mathrm{d}\beta. \tag{11.13}$$

Equation (11.13) implies that the output at time t can be uniquely determined by knowing the $\mathbf{P}_+(t)$ region, or equivalently, the boundary $L(t)$ separating the regions $\mathbf{P}_+(t)$ and $\mathbf{P}_-(t)$ (cf. Figure 11.4d). In fact, the vertices of the boundary $L(t)$ captures the current input value, the past input extrema, and the order in which they occur (Mayergoyz 1991).

The Preisach hysteresis model can be obtained experimentally from measured output data, for instance, by applying an appropriate input voltage and measuring the piezoactuator's displacement response. In this case, based on the work of Banks et al. (1997), a discrete form of the output Eq. (11.9) can be approximated by

$$v(t) \approx \sum_{i=1}^{N} \bar{\mathcal{R}} \mu_r A_r, \tag{11.14}$$

where A_i represents the area associated with the ith node, μ_i is the average value of the weighting surface over area A_i and $\bar{\mathcal{R}}$ takes on value $+1$ or -1 depending on the state of the node, or relay at the node.

Several approaches are available for estimating the Preisach weighting surface $\mu(\cdot, \cdot)$ from the data (Majima et al. 2001; Tan et al. 2001) (see example Preisach weighting surface in Fig. 11.5a). One approach is to generate a collection of first-order descending (FOD) curves, compile the curves into a FOD surface, and then differentiate the FOD surface to find an estimate of the Preisach weighting surface $\mu(\cdot, \cdot)$ (Mayergoyz 1991). Although the method is straightforward, the differentiation process can amplify noise in the measured data, thus creating significant error. An alternative, and more favorable approach, is to find $\mu(\cdot, \cdot)$ by discretizing the Preisach plane and using a least-squares technique to determine the values of μ at a finite number of locations in the Preisach plane $\mathbf{P}$, defined by Eq. (11.10), where $\underline{u}$ and $\bar{u}$ are the minimal and maximal input values, respectively (Tan et al. 2001; Galinaitis and Rogers 1998).

The Preisach weighting function μ for the x-axis of a piezo tube scanner is shown in Fig. 11.5b. The weighting function μ for the y-axis is similar. The details of the modeling can be found in Leang (2004); Leang and Devasia (2006). The model output

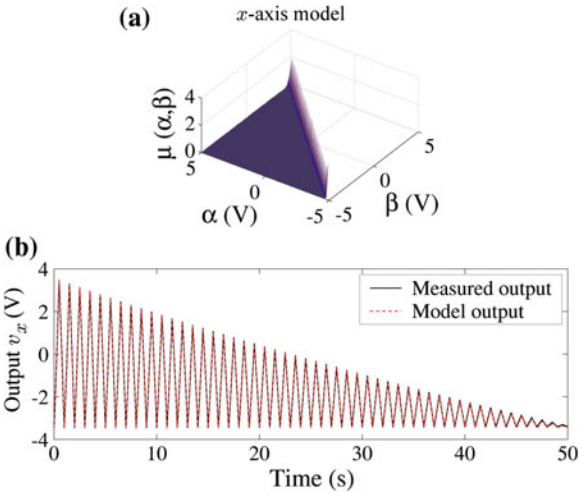


Fig. 11.5 **a** Estimated Preisach weighting surface in the x-axis of a piezo tube scanner. **b** Comparison of model output and measured output versus time

and the measured output data for the x-axis are compared in Fig. 11.5c, which shows very close agreement, e.g., the maximum error is 1.19 % of the total displacement range.

11.2.5 Classical Prandlt-Ishlinksii Model

Another operator-type hysteresis model which has recently been investigated for piezoactuators is the Prandlt-Ishlinskii model (Brokate and Sprekels 1996; Kuhnen 2003; Janaideh et al. 2008). In this model, the output is characterized by the play or stop operator shown in Fig. 11.6 (Brokate and Sprekels 1996). Let the input u be continuous and monotone over the interval $t_i \in T_i \triangleq [t_i, t_{i+n}]$, for $n = 1, 2, \ldots, N$, then the play operator $\mathcal{P}_r$ is defined as

$$\mathcal{P}_r[u](0) = p_r(u(0), 0) = 0, \tag{11.15}$$

$$\mathcal{P}_r[u](t) = p_r(u(t), p_r[u](t)), \tag{11.16}$$

where $p_r(u(t), p_r[u](t_i)) = \max(u - r, \min(u + r, p_r[u](t_{i-1})))$. The play operator's threshold is denoted by r, and also indicated in Fig. 11.6. The square bracket '[]' notation indicates an operation on a function. The output $y(t)$ is a weighted sum of play operators,

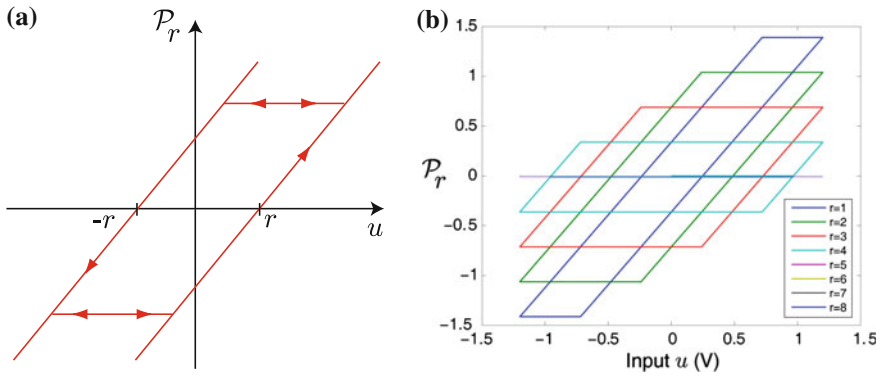


Fig. 11.6 **a** The play or stop operator. **b** A collection of play operators used to model the hysteresis in a PZT bimorph actuator

$$y(t) = kf(t) + \int_0^R v(r)\mathcal{P}_r[u](t)dr, \tag{11.17}$$

where k is a positive constant, $f(t)$ is a function, and $v(r)$ is density function that affects the shape and size of the hysteresis curve.

The hysteresis in a PZT bimorph actuator was modeled by selecting $f(t) = a_0u(t) + a_1$, where a_0 and a_1 are constants, a density function $v(r) = \lambda e^{-\delta r}$, and $r = \rho j$, for $j = 1, 2, \ldots, 8$. The coefficients were determined using input-output data and a nonlinear least-square optimization algorithm. A comparison of the measured output and the output from the PI hysteresis model is shown in Fig. 11.7. The parameters are $a_0 = 1.4613$, $a_1 = -0.0122$, $\lambda = 0.0211$, $\delta = -2.7036$, $\rho = 0.3507$, and $k = 1$. The maximum and root-mean-squared error are 5.75 and 1.34 %, respectively.

Compared to the Preisach model, the PI model is less computationally demanding to implement and invert for feedforward control. However, the drawback is that the classical PI model is limited to symmetric hysteresis behavior. This is not necessarily a disadvantage considering that hysteresis loops in PZT materials is typically symmetric at low to moderate fields. By incorporating the generalized play operator, the PI model can be adapted to saturated hysteresis loops (Janaideh et al. 2008).

11.3 Feedforward Hysteresis Compensation

11.3.1 Feedforward Control Using the Presiach Model

A feedforward input that compensates for hysteresis is obtained by inverting a hysteresis model, such as the Preisach model. But rather than inverting the Preisach model directly, an inverse-Preisach model can be found from the measured input and

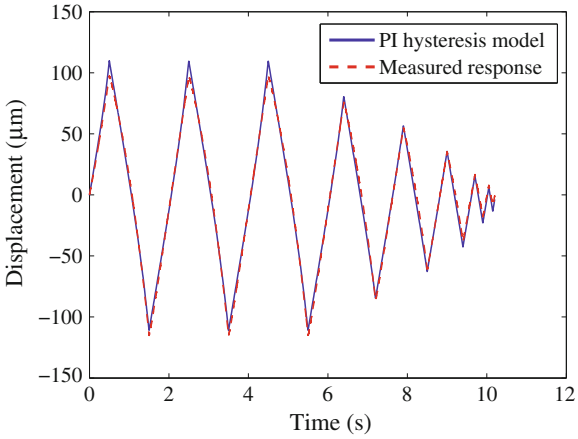


Fig. 11.7 Comparison of PI model output and measured output versus time for a PZT bimorph actuator

output data. The inverse model is found using the same method to find the traditional Preisach model as described above; however, the roles of the input and output are reversed. It is shown in Croft et al. (2001) that when the input $u(t)$ is considered as the output and the output $v(t)$ is considered as the input, the inverse Preisach model takes the form

$$u(t) = \mathcal{H}^{-1}[v](t) \triangleq \iint\limits_{\hat{\alpha} \geq \hat{\beta}} \gamma(\hat{\alpha}, \hat{\beta}) \mathcal{R}_{\hat{\alpha},\hat{\beta}}[v](t) d\hat{\alpha} d\hat{\beta}, \tag{11.18}$$

where the parameters $\hat{\alpha}$, $\hat{\beta}$, $\gamma(\hat{\alpha}, \hat{\beta})$, and the elementary relay $\mathcal{R}_{\hat{\alpha},\hat{\beta}}$ are associated with the inverse-Preisach model. Like the traditional Preisach model, it is assumed that the nonlinearity operates within closed major loops; therefore, the weighting function $\gamma(\hat{\alpha}, \hat{\beta})$ is zero outside of the upper triangle defined by the boundaries $\hat{\alpha} = \hat{\beta}$, $\hat{\alpha} = \bar{v}$, and $\hat{\beta} = \underline{v}$, where $\bar{v}$ and $\underline{v}$ are the upper and lower bounds on the output, respectively.

An inverse-Preisach model can be obtained by using the measured FOD curves to construct a counterpart inverse FOD curve, where the roles of the input and output variables are reversed. With the inverse model in hand, a desired output trajectory is passed through the inverse model to generate an input that compensates for hysteresis effect. The results of this technique for AFM imaging are presented below.

11.3.1.1 Application to AFM Imaging: Feedforward Control of Hysteresis and Dynamics

When an AFM application calls for large-range and high-speed motion, both hysteresis and dynamics compensation are required for precision output tracking. In this

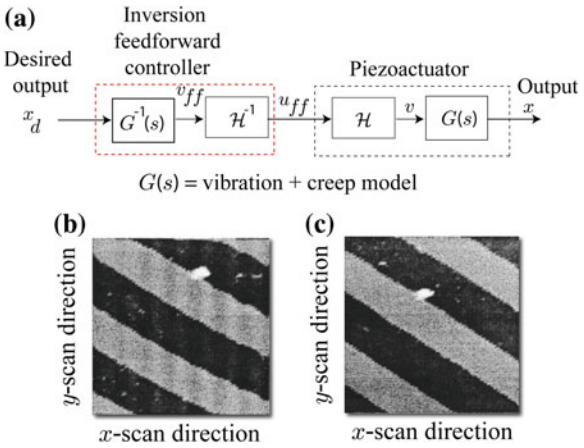


Fig. 11.8 Feedforward control of dynamics $G(s)$ and hysteresis $\mathcal{H}$ for large-range, low- and high-speed positioning. The feedforward control input $u_{ff}(t)$ in (**a**) is obtained by passing the desired output trajectory $x_d(t)$ through the inverse models of hysteresis and dynamics in reverse order. The atomic force microscope images are acquired without feedforward compensation in (**b**) and with feedforward compensation in (**c**). The feedforward input minimizes hysteresis, vibration, and creep. The images are presented with permission from ASME from (Croft et al. 2001)

case, the feedforward control input $u_{ff}(t)$, which accounts for both the dynamic and hysteresis effects, is obtained by passing the desired output trajectory $y_d(t)$ through the inverse models in reverse order as depicted in Fig. 11.8a. This process is performed offline, followed by applying the feedforward input to the piezoactuator. First, the dynamic inverse produces an output $v_{ff}(t)$. The output from this first stage then becomes the input to the inverse-Preisach model, which produces the final feedforward input $u_{ff}(t)$ for hysteresis and dynamics compensation. The image in Fig. 11.8b is acquired without feedforward compensation. The features appear slightly curved because of hysteresis and the ripples show the effect of the dynamics. These distortions are compensated for by applying the feedforward input $u_{ff}(t)$ to the piezoactuator as shown by Fig. 11.8c.

11.3.2 Feedforward Control Using the Prandlt-Ishlinksii Model

Feedforward hysteresis compensation can be accomplished by exploiting the structure of the Prandlt-Ishlinksii model. Particularly, the characteristics of the inverse model is based on the characteristic shape of the inverse hysteresis curve, that is, the input versus output curve shown in Fig. 11.9c (u vs. v plot). It is noted that as the output v increases, the input u increases but traverses onto an upper branch of the inverse-hysteresis curve. In contrast, this behavior is opposite to that observed in the hysteresis curve (v vs. u plot) where the output climbs up on a lower

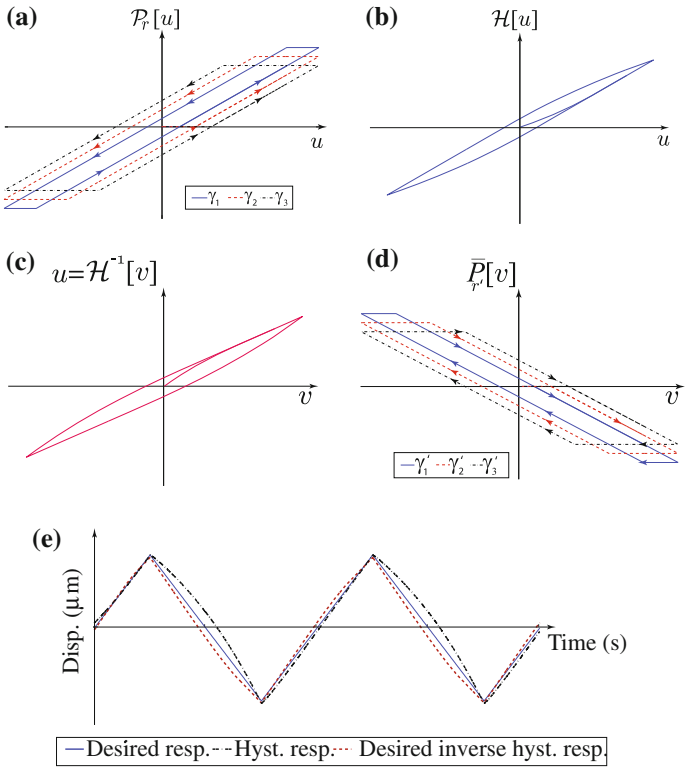


Fig. 11.9 P-I hysteresis model: **a** The play operator with threshold γ_i. **b** An example output versus input plot for the Prandtl-Ishlinskii hysteresis model for a piezoactuator. **c** Inverse hysteresis curve: input versus output plot. **d** A play-type operator for the inverse model with threshold γ_i'. **e** Time response comparing the desired response (*solid line*), hysteresis response (*dash-dot line*), and inverse hysteresis output (*dash line*)

branch as shown in Fig. 11.9a. Therefore, a candidate play-type operator for the inverse-hysteresis model is shown in Fig. 11.9d. Figure 11.9e compares the time responses between the desired output (solid line), output from a hysteretic system (dash-dot line), and the output from the proposed inverse-hysteresis model (dash line). Using this operator offers the advantage that the structure of the forward model can be used directly to map the desired output to the hysteresis-compensating feed-forward input. In other words, the P-I output Eq. (11.17) becomes the inverse map simply by setting the output equal to the input and vice versa.

It is noted that the input-output response for the inverse model shown in Fig. 11.9c is a reflection of Fig. 11.9a about the axis $u = v$. Therefore, the inverse operator shown in Fig. 11.9d is defined as

$$\overline{P}_{r'}[v](0) = \overline{p}_{r'}(h(0), 0) = 0,$$
$$\overline{P}_{r'}[v](t) = \overline{p}_{r'}(h(t), \overline{P}_{r'}[h](t)), \tag{11.19}$$

where

$$\overline{p}_{r'}(h(t_i), \overline{P}_{r'}[h](t_i)) = \max(-h(t_i) - \gamma', \min(-h(t_i) + \gamma', \overline{P}_{r'}[h](t_{i-1}))),$$

$h(t) = g_0' v(t) + g_1'$ with constants g_0' and g_1', and $v(t)$ is the output of the hysteresis behavior. The term γ' denotes the threshold of the new inverse play operator. Using this inverse play-type operator, the output of the inverse-hysteresis model is given by

$$\mathcal{H}^{-1}[v](t) \triangleq h(t) + \int_0^R d_{inv}(\gamma')\overline{P}_{r'}[v](t)d\gamma', \tag{11.20}$$

where $d_{inv}(\gamma')$ is the density function of the inverse P-I model. The performance of the inverse P-I hysteresis compensator is validated in simulations and experiments on a custom-designed high-speed nanopositioning stage described below.

11.3.2.1 Experimental Results

Experiments were performed on a custom-made, three-axis, flexure-guided serial-kinematic nanopositioning system. The experimental system is shown in Fig. 11.10 and the design of a similar stage is described in Chap. 4 and Kenton and Leang (2012) for the interested reader. The serial-kinematic configuration is specifically created for scanning-type applications. For scanning-type applications, such as the rastering movements in AFM imaging, one lateral axis moves much faster (>100-times) than the other axis. Because of this, one axis is designed to have a significantly higher mechanical resonance (Ando et al. 2008; Leang and Fleming 2009). Compliant double-hinged flexures are used to guide the lateral stages in their corresponding actuation directions while limiting out-of-plane (parasitic) motion and dynamic cross coupling. The high-speed (HS) x-stage and the low-speed (LS) y-stage use stiff plate-stack piezoactuators ($5\times5\times10$ mm Noliac SCMAP07) configured serially to provide lateral displacement. Not shown are the details of the z-stage, in which a piezo-stack is embedded into the x-stage body and the free ends secured with plate flexures, where a similar design is described in (Kenton et al. 2011). The stage is outfitted with inductive sensors (Kaman SMU9000-15N) to measure the displacement in the lateral directions. The lateral (x/y) range of motion is determined to be approximately 10×10 μm. The first mechanical resonance for the high-speed x- and low-speed y-stage is 11.10 and 4.68 kHz, respectively. The measured frequency responses of the stage are shown in Fig. 11.11. The results show that the dominant resonances are second-order in nature, and they are actuation modes.

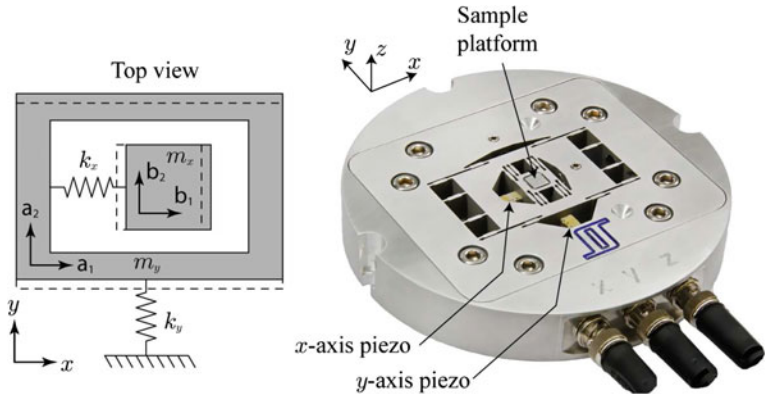


Fig. 11.10 The serial-kinematic three-axis nanopositioning system

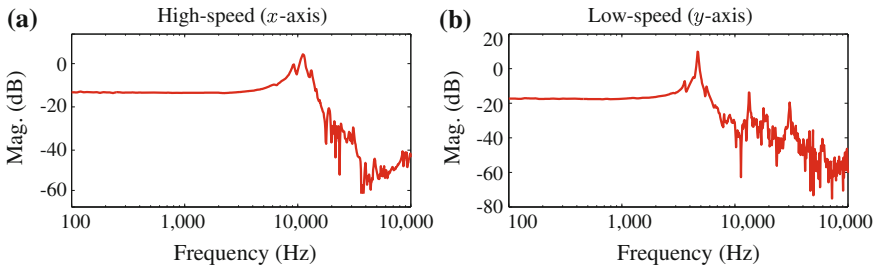


Fig. 11.11 Measured frequency response of nanopositioning stage: **a** high-speed axis and **b** low-speed axis

In the first experiment, the cascade model structure was validated where the P-I model was used to model the hysteresis response. The hysteresis model was obtained by actuating the piezoactuator with a triangle input signal at 10 Hz, full range. The 10 Hz frequency is chosen to avoid the creep effect and minimize the dynamics. Then the input voltage $u(t)$ and the response $x(t)$ were collected and imported to Matlab to a custom-designed least-square optimization program to calculate the P-I parameters. The parameters include g_0 and g_1 for $f(t) = g_0 u(t) + g_1$; and λ, δ and ρ for density function $d(\gamma) = \lambda e^{-\delta\gamma}$, where $\gamma = \rho j$ for $j = 1, 2, \cdots, n$. The parameters of the P-I model were identified as: $g_0 = 0.6081$, $g_1 = 0.0039$, $\lambda = 4.7649$, $\delta = 3.434$ and $\rho = 0.0769$ with eight play operators $j = 8$. Finally, the hysteresis model was validated by comparing the model's response with the measured response as presented in Fig. 11.12. The maximum modeling error is less than 1.87 % and the root-mean-square of the error is 1.39 %.

The hysteresis and dynamics model for the piezoactuator is created by cascading the P-I model $\mathcal{H}[\cdot]$ with the linear dynamics model $G(s)$. The open-loop response of the model is compared to the measured open-loop response of the piezoactuator. The responses are generated by applying a triangle input signal (100 Hz and 1 kHz)

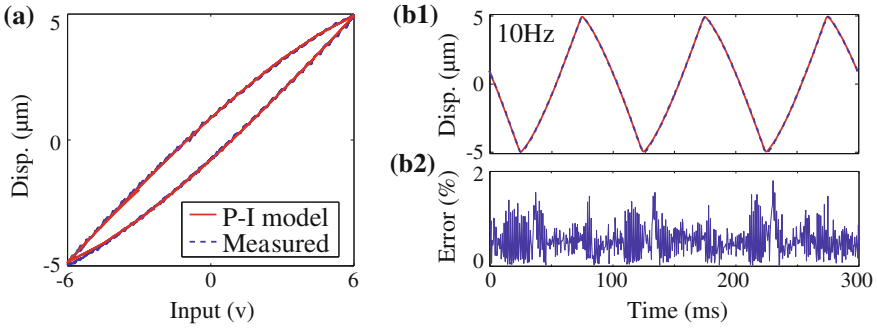


Fig. 11.12 Comparison of measured hysteresis behavior to the output of the P-I hysteresis model at 10 Hz. **a** The hysteresis curves. (b1) and (b2) The displacement and the error between measured and model output versus time at 10 Hz

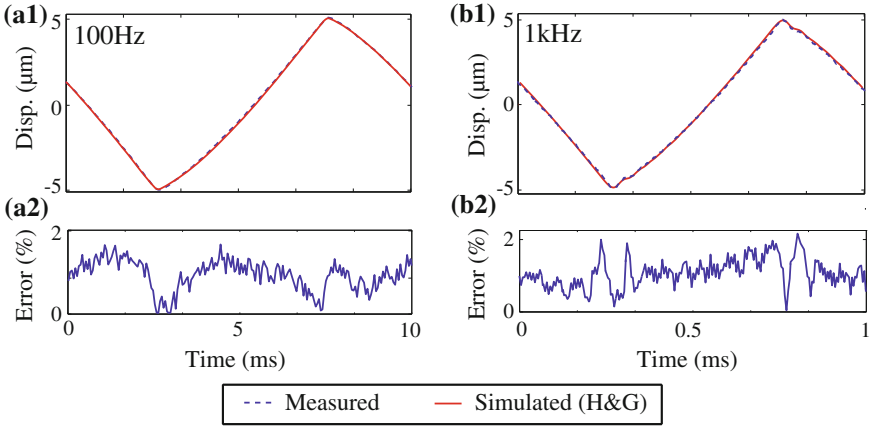


Fig. 11.13 Experimental validation of the system model, $\mathcal{H}[\cdot]$ cascaded with $G(z)$, of the piezoactuator. The displacement and error between the measured and model output versus time at **a1** and **a2** 100 Hz and **b1** and **b2** 1 kHz scanning

to the model and the experimental system such that the displacement is $\pm 5\ \mu$m. The results are shown in Fig. 11.13 for the 100 Hz and 1 kHz responses. It is noted that the maximum error between the model and measured response is less than 2 % up to a scanning frequency of 1 kHz. Therefore, the cascade model structure based on the P-I model is relatively accurate for modeling the combined hysteresis and dynamic effects in the piezoactuator.

The inverse hysteresis model for compensating hysteresis is given by Eq. (11.20). The density function is chosen as $d_{inv}(\gamma') = \lambda' e^{-\delta'\gamma'}$, and the threshold $\gamma' = \rho' j$ for $j = 1, 2, \ldots, 8$. The parameters g_0', g_1',λ', δ', and ρ' were determined using the measured input-output data from the forward hysteresis model described above. The process follows the same steps used to calculate the parameters for the (forward) P-I

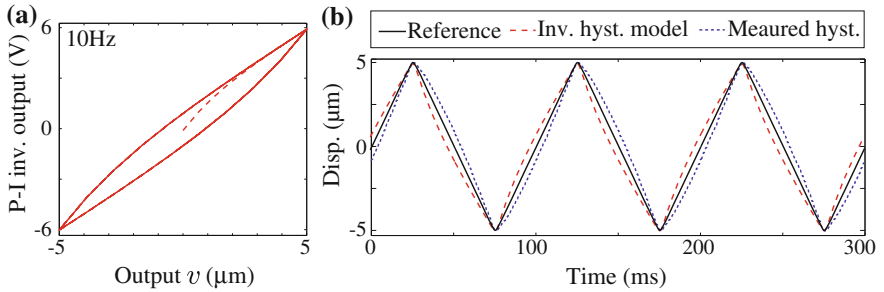


Fig. 11.14 **a** The inverse hysteresis curve. **b** The inverse hysteresis model (*dash line*) compared to the measured hysteresis response (*dash-dot line*) and desired response (*solid line*)

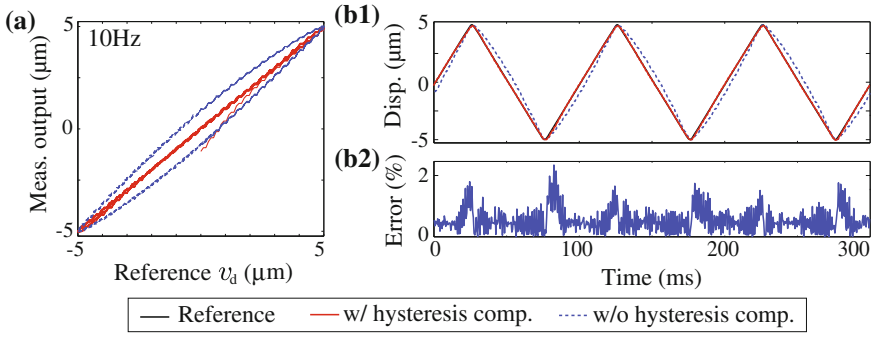


Fig. 11.15 The performance of the inverse hysteresis compensator (*scanning at* 10 Hz). **a** The hysteresis curves for the piezoactuator with (*solid line*) and without (*dash line*) feedforward hysteresis compensation. **b1** The comparison between the reference and the measured (*with hysteresis compensation*) output. **b2** The measured tracking error

hysteresis model, where a nonlinear least-square optimization program was applied to determine the parameters as $g_0' = 1.4583, g_1' = -0.0181, \lambda' = 1.4505, \delta' = 2.5001$, and $\rho' = 0.1611$.

The inverse-hysteresis curve is shown in Fig. 11.14a and its time response is compared to the measured hysteresis response in Fig. 11.14b. The $\mathcal{H}^{-1}$ model is applied to compensate for hysteresis in the x-axis actuator of the experimental system at different frequencies to investigate its effectiveness. Figure 11.15a and b show the performance of $\mathcal{H}^{-1}$ for tracking a desired triangle trajectory at 10 Hz. The $\mathcal{H}^{-1}$ compensates for the hysteresis effect, and subsequently linearizes the system and makes the system's output track the reference trajectory, where the maximum tracking error is 2.1 % at 10 Hz.

The performance of $\mathcal{H}^{-1}$ is further compared to the output response of the dynamic model $G(z)$ in simulation for tracking triangle trajectories at 100 Hz and 1 kHz, since, by compensating for the hysteresis, the output response is dominated by the dynamics effect $G(z)$. Figure 11.16 shows the measured and simulated output versus input

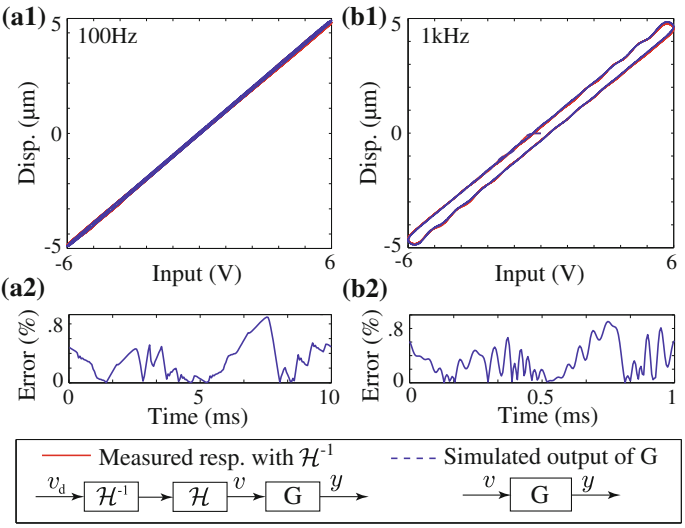


Fig. 11.16 The comparison between the measured output with hysteresis compensation and the simulated output. **a1** and **a2** 100 Hz triangular trajectory. **b1** and **b2** 1 kHz triangle trajectory

plots, where the maximum error is less than 0.92 % at 1 kHz. The results show that the hysteresis effect can be effectively compensated for using the proposed inverse model. In fact, the inverse P-I model can be implemented online and combined with feedback controllers such as repetitive controllers for high-speed nanopositioning (Shan and Leang 2012).

11.4 Chapter Summary

In this chapter, hysteresis modeling and compensation were introduced. Models such as the Preisach and the Prandlt-Ishlinksii models were discussed in detail, including experimental results that demonstrate the use of the models. Several model-based feedforward hysteresis compensation approaches were discussed.

References

Adriaens HJMTA, de Koning WL, Banning R (2000) Modeling piezoelectric actuators. In: IEEE/ASME transactions on mechatronics, vol 5, no 4, pp 331–341, Dec 2000

Ando T, Uchihashi T, Fukuma T (2008) High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes. Prog Surf Sci 83(7–9):337–437

Banks HT, Kurdila AJ, Webb G (1997) Identification of hysteretic confluence operators representing smart actuators: convergent approximations. North Carolina State University CRSC, Tech. Rep., April 1997

Barrett RC, Quate CF (1991) Optical scan-correction system applied to atomic force microscopy. Rev Sci Instrum 62(6):1393–1399

Bertotti G, Mayergoyz I (2006a) The science of hysteresis, vol 1. Elsevier, New York

Bertotti G, Mayergoyz I (2006b) The science of hysteresis, vol 2. Elsevier, New York

Bertotti G, Mayergoyz I (2006c) The science of hysteresis, vol 3. Elsevier, New York

Brokate M, Sprekels J (1996) Hysteresis and phase transitions. Springer, New York

Cao H, Evans AG (1993) Nonlinear deformation of ferroelectric ceramics. J Amer Ceram Soc 76:890–896

Coleman BD, Hodgdon ML (1986) A constitutive relation for rate-independent hysteresis in ferromagnetically soft materials. Int J Engng Sci 24(6):897–919

Croft D, Shed G, Devasia S (2001) Creep, hysteresis, and vibration compensation for piezoactuators: atomic force microscopy application. the ASME. J Dyn Syst Meas Contr 123:35–43

Cross R (1988) Unemployment, hysteresis, and the natural rate hypothesis. Basil Blackwell Ltd., New York

Galinaitis WS, Rogers RC (1998) "Control of a hysteretic actuator using inverse hysteresis compensation", in SPIE Conf. Math Control Smart Struct 3323:267–277

Ge P, Jouaneh M (1995) Modeling hysteresis in piezoceramic actuators. Precis Eng 17(3):211–221

Goldfarb M, Celanovic N (1997) Modeling piezoelectric stack actuators for control of micromanipulation. IEEE Cont Syst Mag 17(3):69–79

Gorbet RB, Wang DWL, Morris KA (1998) Preisach model identification of a two-wire sma actuator. In: Proceedings IEEE International Conference on Robotics and Automation, pp 2161–2167

Hu M, Du H, Ling S-F, Zhou Z, Li Y (2004) Motion control of an electrostrictive actuator. Mechatronics 14(2):153–161

Janaideh MA, Rkaheja S, Su C-Y (2008) Compensation of hysteresis nonlinearties in smart actuators. In: ASME Conference on Smart Materials, Adaptive Structures and Intelligent Systems, pp SMASIS2008–486

Janaideh MA, Su C-Y, Rakheja S (2008) Development of the rate-dependent Prandtl-Ishlinskii model for smart actuators. Smart Mater Struct 17:035026 (11pp)

Jiles DC, Atherton DL (1986) Theory of ferromagnetic hysteresis. J Magn Magn Mater 61:48–60

Kenton BJ, Fleming AJ, Leang KK (2011) A compact ultra-fast vertical nanopositioner for improving SPM scan speed. Rev Sci Instr 82:123703

Kenton BJ, Leang KK (2012) Design and control of a three-axis serial-kinematic high-bandwidth nanopositioner. IEEE/ASME Trans Mechatron 17(2):356–369

Kuhnen K (2003) Modeling, identification and compensation of complex hysteretic nonlinearities: a modified prandtl-ishlinskii approach. Eur J Control 9(4):407–418

Leang KK (2004) Iterative learning control of hysteresis in piezo-based nanopositioners: theory and application in atomic force microscopes, Ph.D. dissertation, Mechanical Engineering

Leang KK, Devasia S (2006) Design of hysteresis-compensating iterative learning control for piezo positioners: application to atomic force microscopes. Mechatronics 16(3–4):141–158

Leang KK, Fleming AJ (2009) High-speed serial-kinematic AFM scanner: design and drive considerations. Asian J Control Spec issue Adv Control Meth Scan Probe Microsc Res Tech 11(2): 144–153

Majima S, Kodama K, Hasegawa T (2001) Modeling of shape memory alloy actuator and tracking control system with the model. IEEE Trans Cont Syst Tech 9(1):54–59

Mayergoyz ID (1991) Mathematical models of hysteresis. Springer, New York

Preisach F (1935) Uber die magnetische nachwirkung. Zeitschrift fur Physik 94:277–302

Shan Y, Leang KK (2012) Dual-stage repetitive control with Prandtl-Ishlinskii hysteresis inversion for piezo-based nanopositioning. Mechatronics 22:271–281

Tan X, Venkataraman R, Krishnaprasad PS (2001) "Control of hysteresis: theory and experimental results", in SPIE Modeling. Signal Process Control Smart Struct 4326:101–112

Chapter 12
Charge Drives

Due to the hysteresis exhibited by piezoelectric actuators, many nanopositioning devices require sensor-based closed-loop control. Although closed-loop control can be effective at eliminating nonlinearity at low speeds, the bandwidth compared to open-loop is severely reduced. In addition, sensor-induced noise can significantly degrade the achievable resolution.

In this chapter, charge drives are introduced as a simple alternative when feedback control cannot be applied or provides inadequate performance. These situations arise in high-speed imaging and positioning applications where wide-bandwidth sensor noise is intolerable or where no feedback sensors are present.

12.1 Introduction

Due to their high stiffness, compact size and effectively infinite resolution piezoelectric actuators are universally employed in nanopositioning systems. However, as discussed in Chap. 2 a major disadvantage of piezoelectric actuators is the hysteresis exhibited at high electric fields. To avoid positioning errors, nanopositioning systems require some form of compensation for piezoelectric nonlinearity. Techniques to accomplish this including feedback and feedforward control were reviewed in Chap. 1.

Since the late 80s, it has been known that driving piezoelectric transducers with current or charge rather than voltage significantly reduces hysteresis (Newcomb and Flinn 1982). Simply by regulating the current or charge, a 5-fold reduction in the hysteresis can be achieved (Ge and Jouaneh 1996; Fleming 2010). Although the circuit topology of a charge or current drive is much the same as a simple voltage amplifier, the uncontrolled nature of the output voltage typically results in the load capacitor being linearly charged. Recent developments have eliminated low-frequency drift and permitted grounded loads, which are necessary in nanopositioning systems (Fleming and Moheimani 2006; Fleming and Leang 2008; Fleming 2013).

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_12,

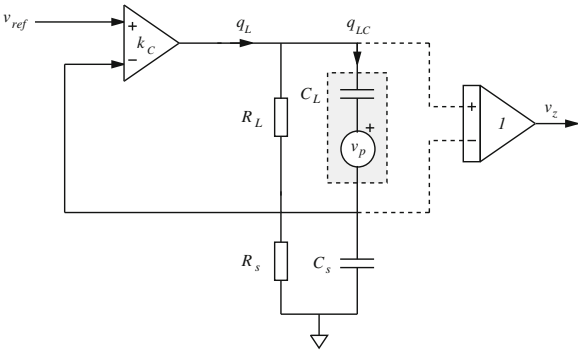


Fig. 12.1 Simplified diagram of a generic charge source

In the following section, the design of charge drives is discussed. These are then applied to both stack actuators and piezoelectric tube actuators in Sects. 12.3 and 12.4, respectively. Section 12.5 contains information specific to the implementation of charge drives for multielectrode piezoelectric tube nanopositioners, which are commonly used in microscopy applications. A summary of the advantages and drawbacks of charge drives then follows in Sects. 12.6 and 12.7.

12.2 Charge Drives

The simplified schematic of a charge drive circuit is shown in Fig. 12.1. The piezoelectric load, modeled as a capacitor and voltage source v_p, is shown in gray. The high gain feedback loop (k_C) works to equate the applied reference voltage v_{ref}, to the voltage across a sensing capacitor C_s. Neglecting the resistances R_L and R_s, at frequencies well within the bandwidth of the control loop, the load charge q_L is equal to

$$q_L = V_{\text{ref}} C_s, \tag{12.1}$$

i.e., we have a charge amplifier with a gain of C_s Coulombs/V.

The foremost difficulties associated with the charge drive in Fig. 12.1 are due to the resistances R_L and R_s. These resistances model the parasitic leakage resulting from the input terminals of the feedback opamps, capacitor dielectric leakage, and v_z measurement. In practice, this parasitic resistance is often swamped with additional physical resistances required to manage the voltage drift associated with the input bias current of the feedback network and instrumentation.

If there exists a parallel load resistance R_L, the actual charge $q_{LC}(s)$ flowing through the load transducer becomes

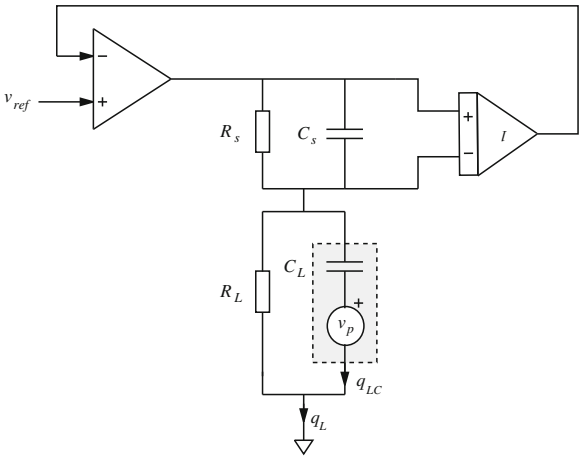


Fig. 12.2 DC accurate charge source for grounded capacitive loads (Fleming and Moheimani 2006). The piezoelectric load, modeled as a capacitor and voltage source v_p, is shown in gray

$$q_{LC}(s) = q_L(s)\frac{s}{s + \frac{1}{R_L C_L}}. \tag{12.2}$$

The amplifier now contains a high-pass filter with cutoff $\omega_c = \frac{1}{R_L C_L}$. That is,

$$\frac{q_{LC}(s)}{V_{\text{ref}}(s)} = C_s \frac{s}{s + \frac{1}{R_L C_L}}. \tag{12.3}$$

In a typical piezoelectric tube drive scenario, with C_L=10 nF, a 1 μA output offset current requires a 10 MΩ parallel resistance to limit the DC voltage offset to 10 V. Phase lead exceeds 5° below 18 Hz. Such poor low-frequency performance precludes the use of charge drives in applications requiring accurate low-frequency tracking, e.g., Atomic Force Microscopy.

A solution for the problem of voltage drift was first presented in Fleming and Moheimani (2004). An auxiliary voltage feedback loop was included to correct low-frequency behavior and allow for constant charge offsets. The circuit implementation required the design of separate voltage and charge feedback controllers. A simplified design relying on the intrinsic voltage control offered by the parasitic resistances was later presented in Yi and Veillette (2005). Neither of these circuits were capable of driving grounded loads. As piezoelectric tubes have multiple external electrodes and a common (often grounded) internal electrode, the requirement for a grounded-load is a necessity.

A charge-driven circuit designed for nanopositioning systems with grounded loads was presented in Fleming and Moheimani (2006). This circuit is shown in Fig. 12.2. The piezoelectric load, modeled as a capacitor and voltage source v_p, is shown

in gray. The amplifier uses a high-voltage differential buffer to equate the voltage measured across the sensing capacitor C_s to the reference voltage v_{ref}.

Neglecting the resistances R_L and R_s, at frequencies well within the bandwidth of the control loop, the load charge q_L is equal to

$$q_L = V_{\text{ref}} C_s. \tag{12.4}$$

That is, the gain is C_s Coulombs/V. When connected to a capacitive load, the equivalent voltage gain is C_s/C_L.

To understand the operation of the amplifier at low frequencies, the transfer function from the applied reference voltage v_{ref} to the load charge q_{LC} must be studied. This can be obtained by first considering the transfer function between the applied reference voltage v_{ref} and the charge q_L,

$$\frac{q_L(s)}{v_{\text{ref}}(s)} = C_s \frac{s + \frac{1}{C_s R_s}}{s}. \tag{12.5}$$

The transfer function from the reference voltage to load charge can then be found by combining Eqs. (12.5) and (12.2)

$$\begin{aligned} \frac{q_{LC}(s)}{v_{\text{ref}}(s)} &= \frac{q_L(s)}{v_{\text{ref}}(s)} \frac{q_{LC}(s)}{q_L(s)} \\ &= C_s \frac{s + \frac{1}{C_s R_s}}{s} \frac{s}{s + \frac{1}{R_L C_L}} \end{aligned} \tag{12.6}$$

That is, the transfer function contains a pole due to the load resistance R_L and a zero due to the sensing resistance R_s. These dynamics can be eliminated by setting $C_L R_L = C_s R_s$, i.e.,

$$\frac{R_L}{R_s} = \frac{C_s}{C_L}. \tag{12.7}$$

Now the amplifier has no low-frequency dynamics and a constant gain of C_s Coulombs/Volt. Effectively the two resistances R_L and R_s form a voltage amplifier at low frequencies that has the same gain as the charge drive at higher frequencies.

As the amplifier can be viewed as the concatenation of a voltage and charge amplifier, it is important to identify the frequency range where each mode of operation is dominant. Consider the schematic shown in Fig. 12.3. If v_{ref} is set to zero, during perfect charge operation i.e., when q_{LC} is correctly regulated to zero, the voltage v_z will be equal to v_p. During voltage dominant behavior, v_z will be regulated to zero. Such characteristics can easily be measured experimentally.

When $v_{\text{ref}} = 0$, which implies $q_L = 0$ the transfer function from v_p to v_z reveals the voltage or charge dominance of the amplifier. At frequencies where $v_z \approx v_p$, the amplifier is charge dominant, and voltage dominant when $v_z \approx 0$. For the hybrid amplifier shown in Fig. 12.2, when $v_{\text{ref}} = 0$,

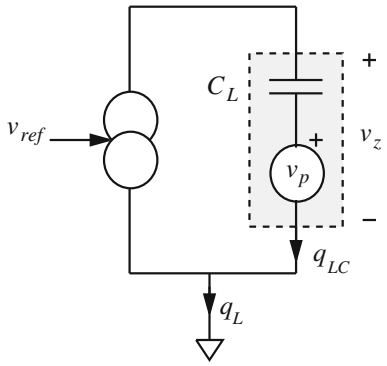


Fig. 12.3 Test for voltage/charge dominance

$$\frac{v_z(s)}{v_p(s)} = \frac{s}{s + \frac{1}{R_L C_L}}. \tag{12.8}$$

That is, at frequencies above $\frac{1}{R_L C_L}$ s^{-1} the amplifier is charge dominant, and voltage dominant below. Obviously, given Eq. (12.8), the objective will be to select a load resistor R_L as large as possible. This may be limited by other factors such as opamp current noise attenuation, bias-current induced offset voltages, and the common-mode and differential leakage of the opamp. In practice $\frac{v_z(s)}{v_p(s)}$ is best measured by simply applying a voltage to another electrode and using that as a reference, as the frequencies under consideration are well below the tube's first mechanical resonance, the applied voltage will be related by a constant. Such experiments are described in Sect. 6.4.2

Alike a typical voltage amplifier, the hybrid amplifier offers little or no hysteresis reduction over the frequency range of voltage dominance. For the same reason, no improvement in creep can be expected. Creep time-constants are usually greater than 10 min, which in this discussion, is effectively DC. At these frequencies, the amplifier behaves analogously to a standard voltage amplifier.

The high frequency bandwidth of a charge drive is limited by the same factors as a voltage amplifier. Bandwidth is limited by a secondary pole in the feedback loop formed by the output impedance and load capacitance. Due to additional phase lag contributed by this secondary pole, the amplifiers bandwidth is restricted to around one-tenth the pole's frequency if large stability margins are to be retained.

In addition to the secondary pole discussed above, charge drives are also limited by the bandwidth of the differential amplifier in the charge measuring circuit. If this is near or less than the frequency of the secondary pole, it will degrade phase margin and necessitate a reduction in bandwidth. Although high voltage differential amplifiers such as the AD629 are available for a few dollars, discrete designs can achieve much higher bandwidths, but with increased complexity. If closed-loop bandwidths of greater than a 100 Hz are required, a high-performance differential amplifier is mandatory.

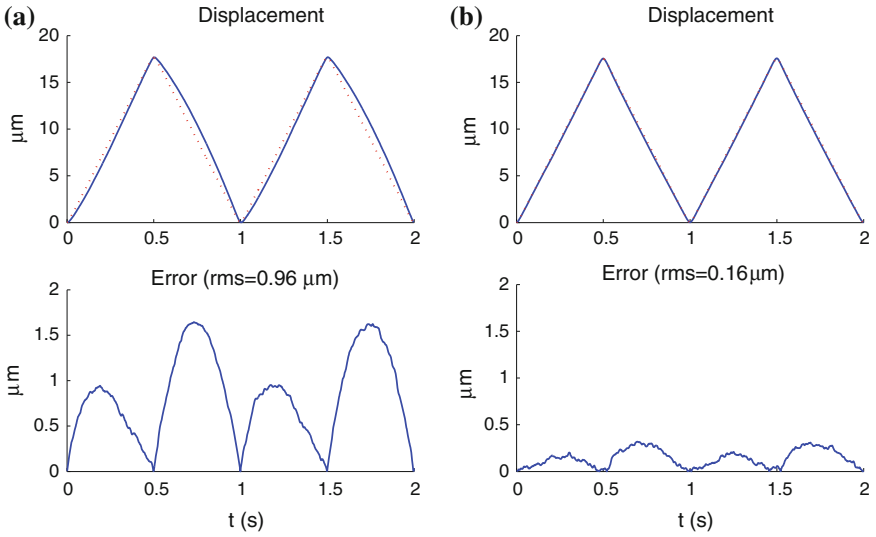


Fig. 12.4 The displacement of the P733 nanopositioner driven by a voltage amplifier (**a**) and charge drive (**b**). The *dotted line* in the displacement plot is the input signal scaled to act as a reference

12.3 Application to Piezoelectric Stack Nanopositioners

In this section, the positioning performance of a charge drive is compared to a voltage amplifier when driving the Physik Instrumente P733 nanopositioner described in Sect. 3.2.2. This device has a specified range of $30 \times 30 \times 10\,\mu$m in the X, Y, and Z axis.

In this experiment, the charge drive is connected to the Z-axis actuator, which has a capacitance of 3.2 μF. To provide a voltage gain 20, equal to that of the voltage amplifier, the charge gain is set to 64 μC/V.

In Fig. 12.4, the full-range displacement of the nanopositioner is plotted in response to a 1 Hz triangle wave with both voltage and charge actuation. With voltage drive, the maximum absolute positioning error is 1.6 μm, or 9.3 % of the range. In Fig. 12.4b, the use of a charge drive reduces the maximum positioning error to only 300 nm, or 1.8 % of the range, which may be a tolerable error in many applications.

The hysteresis exhibited by the actuator is most clearly observed by plotting the reference command against displacement. This is performed for both voltage and charge actuation in Fig. 12.5. Clearly, the charge drive significantly reduces the maximum deviation from linear. When plotting hysteresis it is important to ensure that no other sources of phase delay are present in the data. This includes linear phase lag due to mechanical dynamics, amplifiers, sensors, and other instruments in the signal chain. As such effects can be erroneously neglected, linear phase lag can be mistaken for hysteresis since it results in a similar waveform. The most common source of phase lag is from driving amplifiers, which are typically low in bandwidth

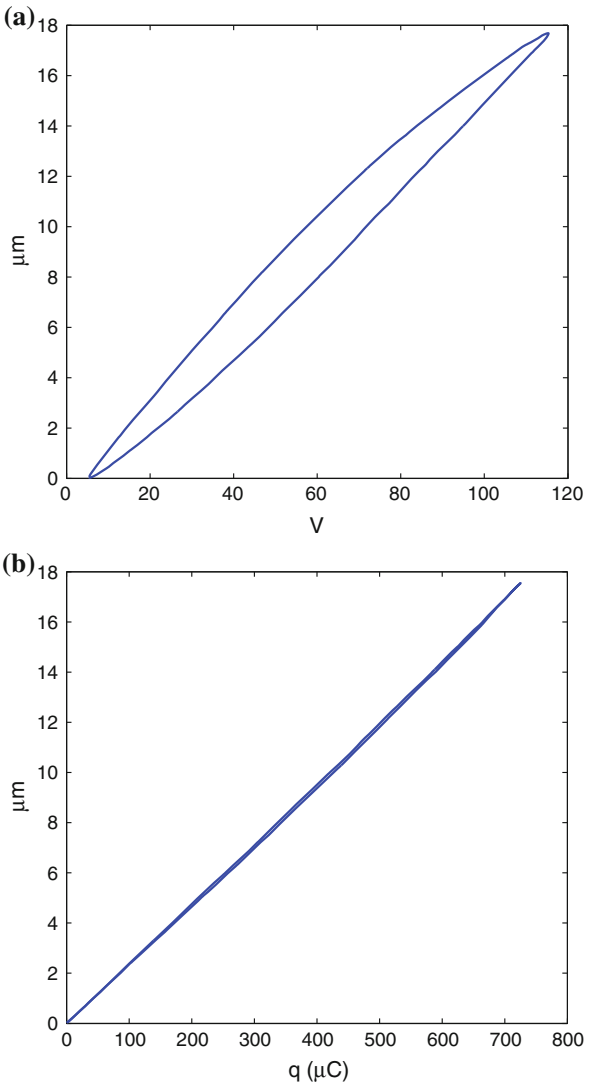


Fig. 12.5 The displacement of the P-733 nanopositioner as a function of voltage (**a**) and charge (**b**). The input was a 1 Hz triangle wave

when driving large capacitive loads. In these experiments, the driving frequency of 1 Hz is at least two decades lower than the bandwidth of amplifiers, sensors, and mechanical dynamics, thus additional sources of phase lag are negligible.

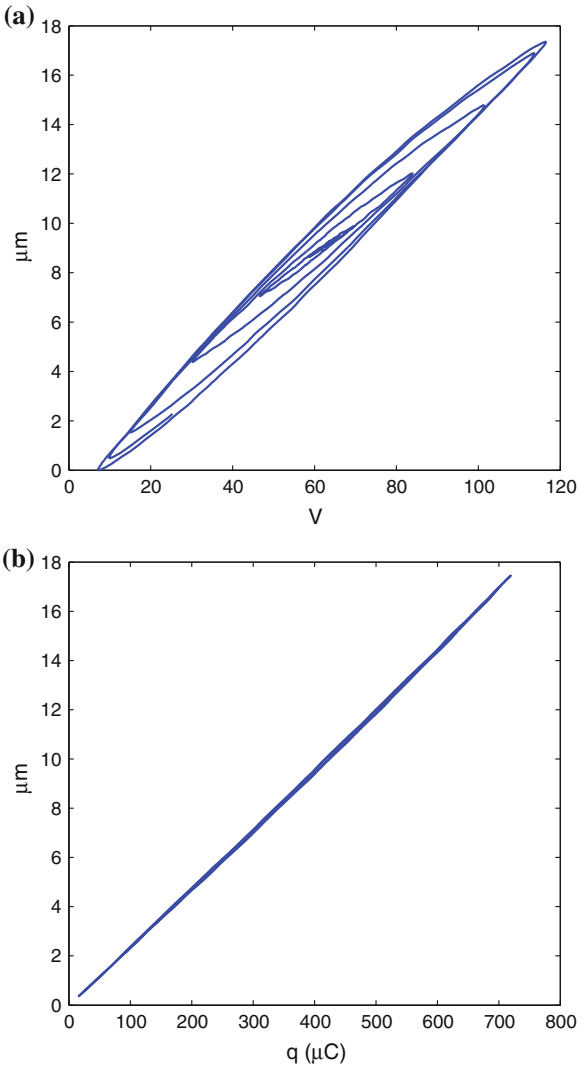


Fig. 12.6 The displacement of the P733 nanopositioner as a function of voltage (**a**) and charge (**b**). The input was a 1 Hz triangle wave, ramped in amplitude over 5 s

In addition to the worst-case, or full range hysteresis, it is also useful to observe the dependence on driving amplitude. In Fig. 12.6, the displacement of the nanopositioner is plotted in response to a 1 Hz triangle wave that is increase in amplitude over five periods. While the voltage-driven positioning nonlinearity markedly increases with signal amplitude, the charge-driven nonlinearity remains low in all cases.

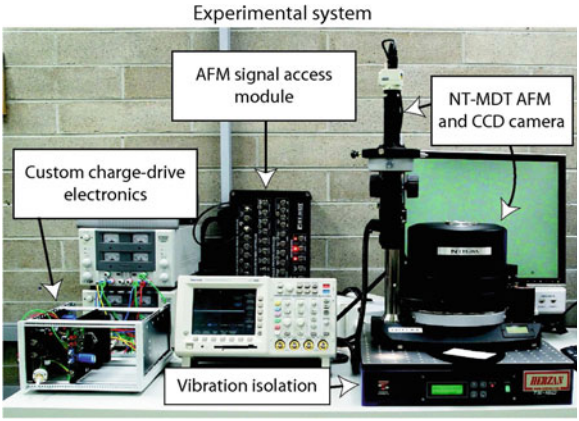


Fig. 12.7 A photograph of the experimental SPM system with charge drive electronics

12.4 Application to Piezoelectric Tube Nanopositioners

A key component of Scanning Probe Microscopes (SPM's) (Meyer et al. 2004) is the nanopositioning system required to manoeuvre the probe or sample. To avoid imaging artefacts, SPMs require some form of compensation for the positioning nonlinearity. Techniques to accomplish this, including feedback, feedforward and image-based compensation are reviewed in Abramovitch et al. (2007) and Clayton et al. (2009).

The most popular technique for compensation in commercial scanning probe microscopes is sensor-based feedback control using integral or Proportional-Integral (PI) control. Such controllers are simple, robust to modeling error, and due to high loop-gain at low-frequencies, effectively reduce piezoelectric nonlinearity. However, the disadvantages of closed-loop control include: cost, additional complexity, limited bandwidth, and sensor-induced noise.

In this section, charge control is applied to linearize an SPM positioning stage. The aim is to provide a simple alternative to feedback control where such techniques cannot be applied or provide inadequate performance. For example, in high-speed imaging (Ando et al. 2005; Humphris et al. 2005; Rost et al. 2005; Fantner et al. 2006; Picco et al. 2007; Fleming 2009), it is difficult or impossible to achieve a satisfactory controller bandwidth. Sensor noise is another major issue when atomic resolution is required, particularly if the controller bandwidth is greater than a few Hertz. Also, in many 'home-made' and application specific microscopes, feedback sensors are not present and the only control option is open-loop, which is the case for all of the scanners reported in Ando et al. (2005), Humphris et al. (2005), Rost et al. (2005), Fantner et al. (2006), Picco et al. (2007) and Fleming (2009).

Pictured in Fig. 12.7, an NT-MDT Ntegra SPM was retrofitted with a charge drive on the fast scanning x-axis. A signal access module allowed direct access to the

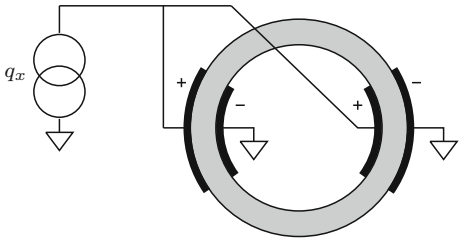


Fig. 12.8 *Top view* of the tube scanner. The x-axis electrodes are quartered on the inside and outside and driven in parallel by the charge source

scanner electrodes and reference signal. The charge gain was set to provide an equivalent voltage gain equal to the standard internal controller gain of 15. Accordingly, no modifications to the scan-controller or software interface were required.

The scanner is an NT-MDT Z50309cl piezoelectric tube scanner with 100 μm range. As shown in Fig. 12.8, the tube has quartered internal and external electrodes that allow the scanner to be driven in a bridged configuration. That is, where the internal and external electrodes are driven with equal but opposite voltages. The naming arises from the way in which the electrodes 'bridge' the two driving sources together, effectively doubling the differential voltage experienced by the actuator. Compared to the more popular grounded internal electrode configuration, the bridged configuration requires half the driving voltage to achieve full range. In these experiments, one pair of electrodes are grounded to allow an analogy with stack-based positioners that are driven with this configuration. Further discussion specific to piezoelectric tube scanners, including the application of charge drives to bridged electrodes, is contained in Sect. 12.5.

During imaging, the AFM was operated in constant height, contact mode, using a cantilever with spring constant 0.2 N/m. The lateral deflection of the piezo actuator was measured using capacitive sensors incorporated into the scanner assembly. A 1 Hz triangle wave was applied to develop scans of 5, 20, and 50 μm, corresponding to 5, 20, and 50 % of the maximum scan range. The scanner trajectories and tracking errors are plotted in Fig. 12.9. Maximum absolute error for voltage and charge drive is compared in Table 12.1.

The displacement nonlinearity was only 2 % in the 5 μm voltage-driven scan; this was reduced to 0.86 % using charge actuation. In the 20 and 50 μm scans, voltage-driven nonlinearity was more significant, 4.9 and 7.2 %, respectively. This was reduced to 0.36 and 0.78 % using charge, a reduction of 93 and 89 %.

AFM images of a 20 nm feature-height parallel calibration grating (3 μm pitch) are pictured in Fig. 12.10. Images were recorded by linearizing the y-axis with a capacitive sensor and driving the x-axis with voltage, then charge. For the 5 μm scan in Fig. 12.10a, the 2 % voltage nonlinearity is not discernable. However, for the 20 and 50 μm scans in Fig. 12.10c, e, the 4.9 and 7.2 % nonlinearity clearly distorts the image. In all three charge-driven scans, Fig. 12.10b–f, the nonlinearity is less than 1 % and image distortion is imperceptible. Reference lines in Fig. 12.10 are superimposed on each image for comparision.

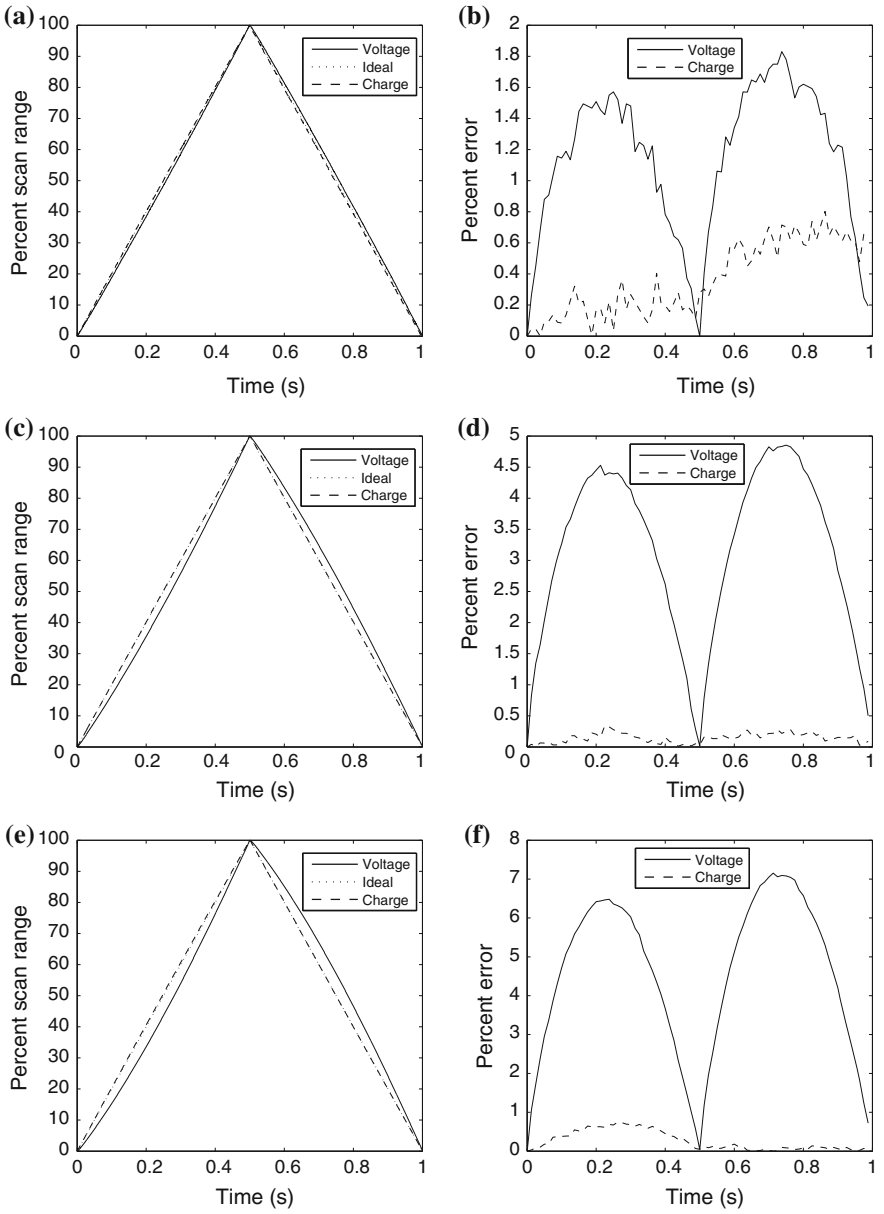


Fig. 12.9 The measured scanner deflection and percentage error for 5, 20, and 50 μm scans. The input was a 1 Hz triangle wave. **a** 5 μm scan. **b** 5 μm scan error. **c** 20 μm scan. **d** 20 μm scan error. **e** 50 μm scan. **f** 50 μm scan error

Table 12.1 Open-loop scan error with voltage and charge actuation

Scan range (μm)	Absolute Scan Error		Reduction (%)
	Voltage (%)	Charge (%)	
5	2.0	0.86	54
20	4.9	0.36	93
50	7.2	0.78	89

12.5 Alternative Electrode Configurations

Commercial scanning probe microscopes contain piezoelectric tube nanopositioners that utilize one of two possible electrode configurations: the grounded internal electrode configuration, or quartered internal electrode configuration. The application of charge drives to each of these scenarios is discussed below.

The techniques discussed in this section are not relevant to piezoelectric stack-based scanners. These actuators are unipolar and require only a single voltage or charge source with one grounded electrode. This configuration is used in the previous sections.

12.5.1 Grounded Internal Electrode

The most common electrode configuration on piezoelectric tube scanners is a single-grounded internal electrode with quartered external electrodes. Electrodes on opposite sides are driven with equal but opposite voltages to induce deflection in that axis. Although the tubes themselves are simple to fabricate, this configuration requires two bipolar voltage amplifiers for each electrode, four in total to achieve x and y lateral motion.

As charge amplifiers are more complicated than voltage amplifiers it is undesirable to require four of them. However, the drive requirements can be simplified if the two electrodes are mechanically and electrically identical. If so, the voltage induced on the charge driven electrode can simply be negated and applied to the opposite electrode as shown in Fig. 12.11a. For an explanation, consider the electrical equivalent circuit in Fig. 12.11b. The piezoelectric elements under each left- and right-hand electrode are modeled as the capacitances c_{p1} and c_{p2} in series with the piezoelectric strain voltages v_{p1} and v_{p2}. As the electrodes are on opposite sides of the tube, and equal but opposite voltages are applied to both electrodes, the piezoelectric strain voltages v_{p1} and v_{p2} will also be equal but opposite. Under this assumption, if the voltage v_1 is applied oppositely to the right-hand electrode, i.e., if $v_2 = -v_1$, the charge q_2 will be equal but opposite to q_1, and the tube will behave linearly as if two independent charge amplifiers were used.

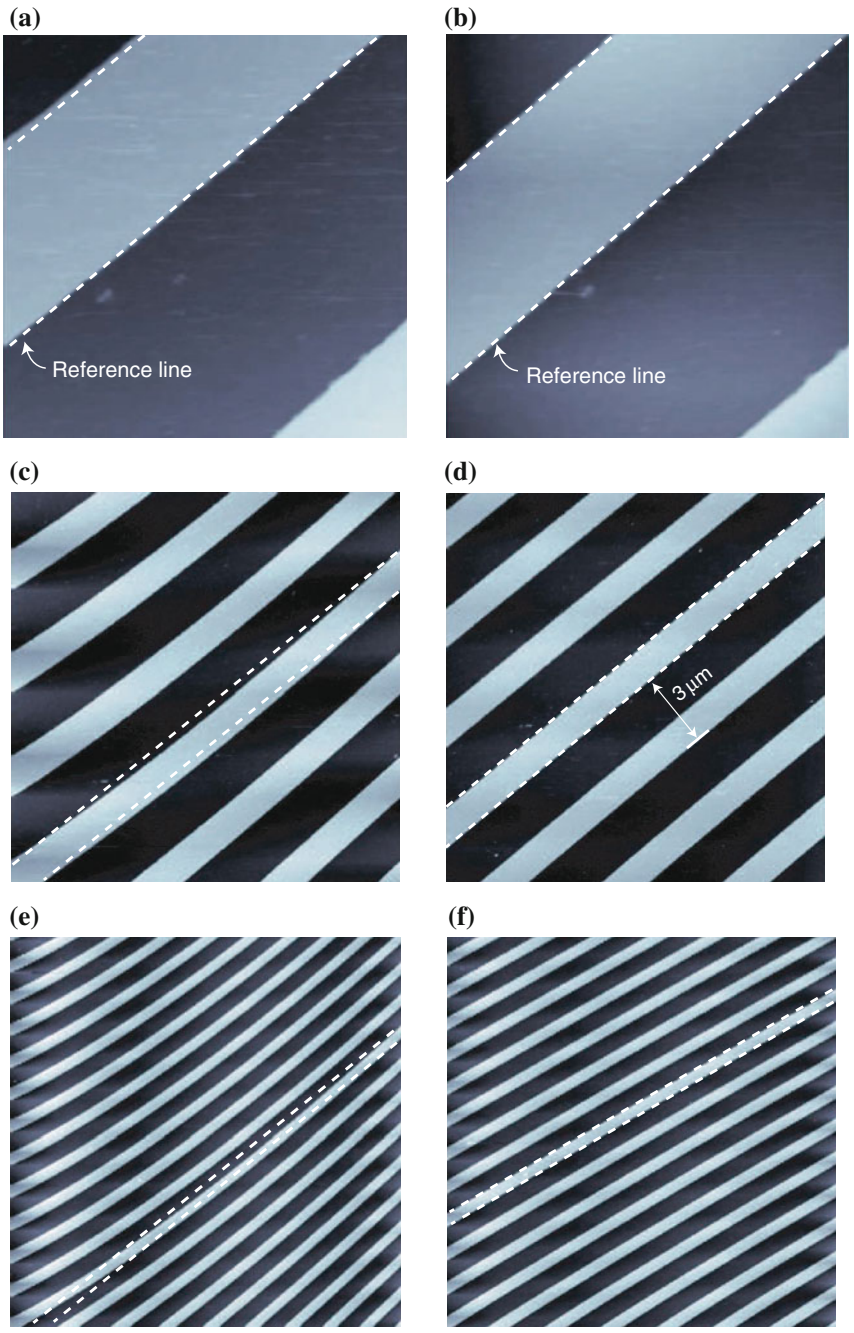


Fig. 12.10 A comparison of images recorded using voltage and charge actuation. The sample is a periodic calibration grating with 20 nm feature height. **a** Voltage drive 5 × 5 µm. **b** Charge drive 5 × 5 µm. **c** Voltage drive 20 × 20 µm. **d** Charge drive 20 × 20 µm. **e** Voltage drive 50 × 50 µm. **f** Charge drive 50 × 50 µm.

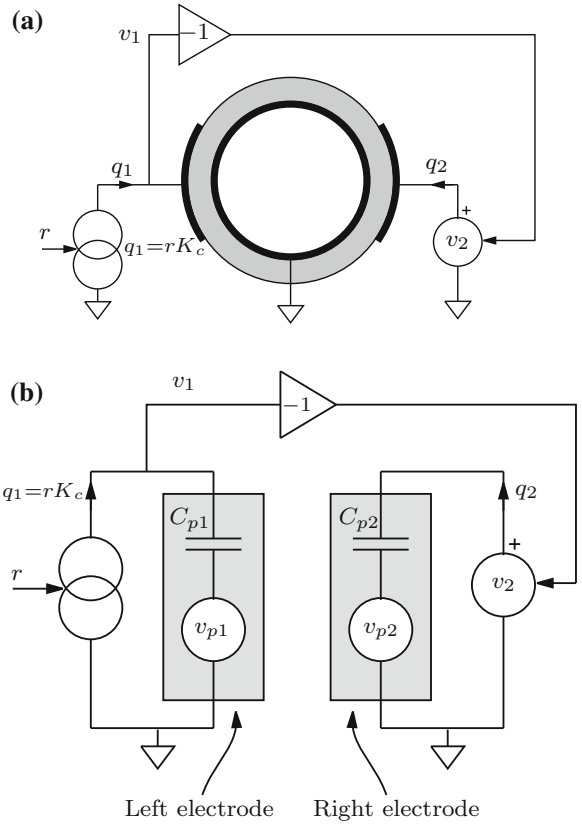


Fig. 12.11 The grounded internal electrode configuration (*top*) and eqivalent electrical circuit (*bottom*)

12.5.2 Quartered Internal Electrode

As illustrated in Fig. 12.8 and discussed in Sect. 12.4, the quartered internal electrode configuration, although more difficult to fabricate, requires half the voltage of the previous technique to achieve the same deflection. This is a major advantage as high-voltage amplifiers are costly and two independent amplifiers are required for each axis.

The application of a charge drive to bridged electrodes is somewhat different from the standard voltage-driven configuration. Usually opposite voltages are applied to the inner and outer electrode while the left- and right-hand electrode pairs are connected in parallel. As the bridged electrodes connect the two sources in series, two charge drives would not form a stable circuit. This is analogous to connecting two voltage sources in parallel.

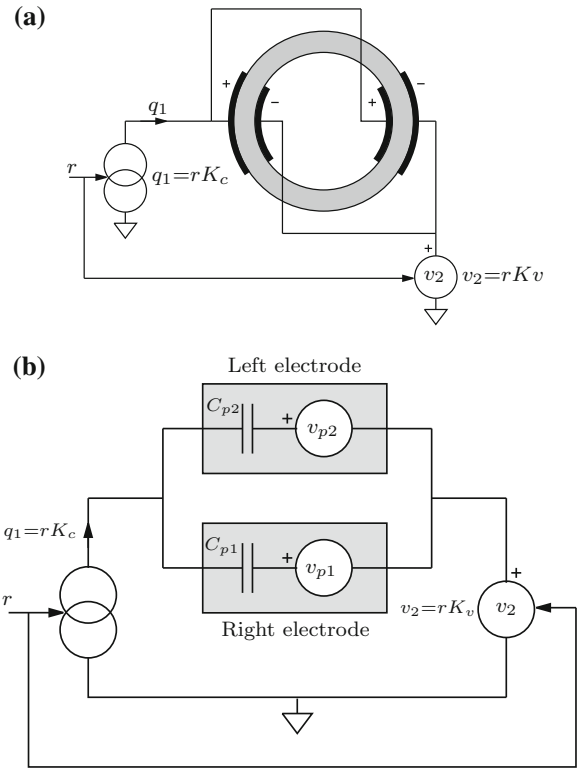


Fig. 12.12 The quartered, or bridged internal electrode configuration (*top*) and eqivalent electrical circuit (*bottom*)

A suitable electrical connection that requires only a single charge drive is shown in Fig. 12.12. Interestingly, varying the voltage on the electrodes marked negative does not alter the amount of deposited charge or corresponding displacement. However, by setting the voltage on the negative electrode approximately equal but opposite to the voltage developed by the charge drive, twice as much charge can be deposited with the same voltage. So far as the charge drive is concerned, driving the negative electrodes with an opposite voltage results in a doubling of the load capacitance. Thus, twice as much charge can be deposited with the same voltage.

The electrical equivalent circuit of a charge-driven tube with internal electrodes is contained in Fig. 12.12b. If a reference signal r is applied to a charge amplifier with gain K_c Coulombs/Volt, the load voltage will be approximately

$$v_1 = rK_c/C_p \tag{12.9}$$

(neglecting v_{p1} and v_{p2} that are much lesser than $v_1 - v_2$), where C_p is the parallel combination of C_{p1} and C_{p2}. Thus, if the voltage gain K_v is set to $K_v = -K_c/C_p$,

the voltage v_2 will be approximately $-v_1$ and the charge drive will result in an approximately balanced voltage across the load. Another option is to adopt a similar approach to the previous section, however this requires additional circuitry to buffer and measure the voltage developed by the charge drive (v_1).

The configuration in Fig. 12.12a was implemented on the experimental setup discussed in Sect. 12.4. The bridged load allowed a 200 V charge drive to obtain the full 400 V differential required for maximum deflection. An experimental 100 μm scan comparing both voltage and charge actuation is plotted in Fig. 12.13. At full range, the maximum scan error using voltage is 9.7 %, compared to 2.0 % using charge.

It is interesting to note the assymetry of nonlinearity in Figs. 12.9 and 12.13. The decreasing part of the charge-driven scan has less nonlinearity in all cases. The maximum charge-driven scan error, even at full range with bridged electrodes is only 0.5 % compared to 9.7 % using voltage.

12.6 Charge Versus Voltage

In this section, the advantages and drawbacks of charge drives are discussed for open-loop positioning applications.

12.6.1 Advantages

There are two motivating factors for the use of charge drives in nanopositioning systems: reduction of hysteresis; and vibration compensation.

In Sect. 12.4, the nonlinearity of a tube scanner driven to half its full-scale range was measured at 7.2 %. Subsequent images demonstrate that this magnitude of error is intolerable. Conversely, when driven with charge, scan error remains below 1 % and is imperceptible in the images. Thus, while closed-loop control of voltage-driven nanopositioners is mandatory in imaging applications, the use of charge drives can provide satisfactory linearity with no feedback. Follow-on benefits include zero sensor-induced noise, no controller imposed bandwidth limitations, simpler scanner design (due to the absence of sensors) and lower cost.

In high-speed nanopositioning and microscopy applications (Ando et al. 2005; Humphris et al. 2005; Rost et al. 2005; Fantner et al. 2006; Picco et al. 2007; Fleming 2009) where feedback control is not feasible, the use of charge drives has the potential to significantly increase imaging performance. Feedback control is not an option due to bandwidth and noise considerations.

In addition to hysteresis reduction, damping of resonant modes can also be accomplished without the need for feedback. In Chap. 6, the first mechanical scanner resonance is attenuated by shunting the actuator electrodes with a parallel passive

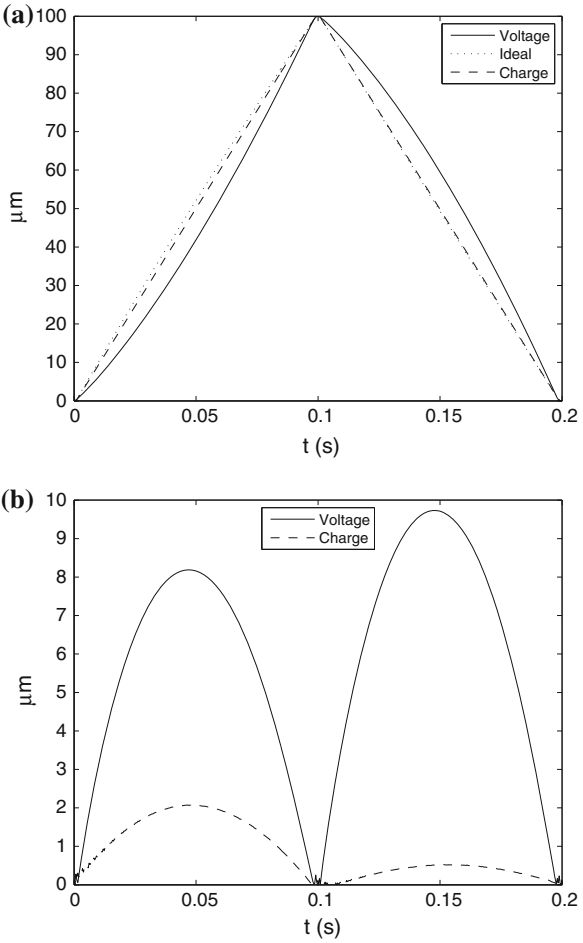


Fig. 12.13 The bridged voltage and charge-driven deflection in response to a 5 Hz triangle wave (**a**); and scan error (**b**)

impedance. The impedance is tuned to resonate with the transducers capacitance at the frequency of problematic modes. Greater than 20 dB attenuation of the first lateral mode is demonstrated.

12.6.2 Disadvantages

The disadvantages of charge drives are the increased circuit complexity, voltage range reduction and necessity for gain tuning.

Although floating-load charge drives are similar to standard inverting voltage amplifiers, the grounded-load configuration in Fig. 12.2 requires a high-performance differential buffer. The differential buffer requires high-input impedance, common-mode-range equal to the high-voltage supply and common-mode-rejection-ratio greater than 80 dB over the bandwidth of the amplifier. These specifications are not met by available integrated devices but can be achieved with discrete designs, with increased circuit complexity. However, if the application does not require operation beyond 100 Hz, the differential buffer can be constructed easily with off-the-shelf parts, for example the AD629.

The differential buffer present in the grounded-load configuration contributes some additional noise, which is likely to be greater than the thermal noise of resistors in a voltage feedback amplifier. Thus, a grounded-load charge drive will generate more noise than a voltage amplifier of the same gain. The situation is different for a floating-load charge drive. This does not require a differential buffer and can provide less noise than a comparable voltage amplifier as the feedback network does not contribute thermal noise.

In addition to amplifier noise, electromagnetic interference can contribute strongly to circuits with high-impedance nodes. In this regard, the grounded-load configuration is superior to the floating-load configuration as it is more easily shielded.

Another consideration is the reduction in voltage range due to the drop across the sensing capacitor C_s. The output voltage range is limited by the maximum amplifier voltage minus the feedback voltage. This requires a slightly higher supply voltage to develop the same transducer displacement. For high-voltage devices greater than 100 V, the maximum 10 V drop across C_s is not significant. However, in lower voltage applications, this reduction may become significant as standard ICs are limited to between 36 and 50 V. Simply increasing C_s and decreasing V_{ref} is an option for improving voltage range.

Aside from issues with the actual circuitry, the only significant difference between voltage and charge actuation is the need to adjust charge gain. At DC and low-frequencies, the voltage gain is fixed by the ratio of resistances R_L and R_s—these are easily interchanged or adjusted. To achieve the same gain at higher frequencies, C_s would need to be adjusted accordingly. This is impossible as variable capacitors of sufficient capacitance are not available. A better option is to select C_s larger than necessary, then add a gain α to the differential buffer, this allows a reduction of charge gain to that desired. After the charge gain is set, the resistance ratio R_L/R_s needs to be adjusted to $\alpha C_s/C_L$.

12.7 Impact on Closed-Loop Control

At normal imaging speeds of less than 10 Hz scan-rate, simple integral controllers with either damping controllers or notch filters for resonance compensation provide sufficient performance and are widely applied (Leang and Devasia 2007). Over the frequency range where loop-gain is greater than 1, typically from DC to tens of Hz, the

scanner displacement tracks additive sensor noise. Even with low-noise capacitive sensors (noise density 20 pm/$\sqrt{\text{Hz}}$), a controller bandwidth of 100 Hz results in greater than 1 nm peak-peak noise. This precludes standard closed-loop scanners from achieving atomic resolution. The situation can be improved by dropping the controller bandwidth to 10 Hz. Although this provides the possibility for atomic resolution, the limited bandwidth restricts usage to extremely slow scanning only.

With charge control, sensors are not required for linearization. Thus, no sensor-induced noise is present. However, to eliminate creep and thermal drift in the scanner, a slow feedback loop can be added. In this case, sensor noise is negligible as the bandwidth of such a control-loop would be less than 1 Hz.

Charge drives are also suited to systems containing feedforward controllers. Many linear feedforward controllers have been proposed that significantly improve the speed and accuracy of positioning stages with little added complexity, a review of such techniques can be found in Abramovitch et al. (2007), Leang et al. (2009) and Clayton et al. (2009). In the past, a major drawback of feedforward control has been the difficulty of eliminating hysteresis over a wide variety of operating conditions. When using charge drives, hysteresis is heavily reduced and feedforward control can be effectively applied, even at high scan ranges Clayton et al. (2008).

12.8 Chapter Summary

In this chapter, charge drives were introduced as an open-loop technique for reducing the hysteresis exhibited by piezoelectric actuators. Experimental results demonstrated an improvement in linearity of greater than 90 % for both piezoelectric tube and stack actuated nanopositioning systems. The advantages are:

- Reduction of hysteresis to less than 1 % of the scan range.
- Straightforward replacement for voltage amplifiers.
- Compatible with sensor-less vibration control.

Disadvantages include:

- Greater circuit complexity.
- Requires tuning to set the gain.
- Low frequency performance is limited by the transducer capacitance.

References

Abramovitch DY, Andersson SB, Pao LY, Schitter G (2007) A tutorial on the mechanisms, dynamics, and control of atomic force microscopes. In: Proceedings of American Control Conference, New York City, NY, Jul 2007, pp 3488–3502

Ando T, Kodera N, Uchihashi T, Miyagi A, Nakakita R, Yamashita H, Matada K (2005) High-speed atomic force microscopy for capturing dynamic behavior of protein molecules at work.e-J Surf Sci Nanotechnol 3:384–392

Clayton GM, Tien S, Devasia S, Fleming AJ, Moheimani SOR (2008) Inverse-feedforward of charge controlled piezopositioners. Mechatronics 18:273–281

Clayton GM, Tien S, Leang KK, Zou Q, Devasia S (2009) A review of feedforward control approaches in nanopositioning for high-speed SPM. J Dyn Syst Meas Contr 131(1–19):61–101

Fantner GE, Schitter G, Kindt JH, Ivanov T, Ivanova K, Patel R, Holten-Andersen N, Adams J, Thurner PJ, Rangelow IW, Hansma PK (2006) Components for high speed atomic force microscopy. Ultramicroscopy 106(2–3):881–887

Fleming AJ (2009) High-speed vertical positioning for contact-mode atomic force microscopy. In: Proceedings of IEEE/ASME international conference on advanced intelligent mechatronics, Singapore, July 2009, pp. 522–527

Fleming AJ (2010) Quantitative SPM topographies by charge linearization of the vertical actuator. Rev Sci Instrum 81(10):103701–103705

Fleming AJ (2013) Charge drive with active DC stabilization for linearization of piezoelectric hysteresis. IEEE Trans Ultrason Ferroelectr Freq Control 60(8):1630–1637 (published: 01)

Fleming AJ, Leang KK (2008) Charge drives for scanning probe microscope positioning stages. Ultramicroscopy 108(12):1551–1557

Fleming AJ, Moheimani SOR (2004) Hybrid DC accurate charge amplifier for linear piezoelectric positioning. In: Proceedings 3rd IFAC symposium on mechatronic systems, Sydney, Australia, Sept 2004

Fleming AJ, Moheimani SOR (2006) Sensorless vibration suppression and scan compensation for piezoelectric tube nanopositioners. IEEE Trans Control Syst Technol 14(1):33–44

Ge P, Jouaneh M (1996) Tracking control of a piezoceramic actuator. IEEE Trans Control Syst Technol 4(3):209–216

Humphris ADL, Miles MJ, Hobbs JK (2005) A mechanical microscope: high-speed atomic force microscopy. Appl Phys Lett 86:034106-1–034106-3

Leang KK, Devasia S (2007) Feedback-linearized inverse feedforward for creep, hysteresis, and vibration compensation in afm piezoactuators. IEEE Trans Control Syst Technol 15(5):927–935

Leang KK, Zou Q, Devasia S (2009) Feedforward control of piezoactuators in atomic force microscope systems. Control Syst Mag 29(1):70–82

Meyer E, Hug HJ, Bennewitz R (2004) Scanning probe microscopy. The lab on a tip. Springer, Heidelberg

Newcomb CV, Flinn I (1982) Improving the linearity of piezoelectric ceramic actuators. IEE Electron Lett 18(11):442–443

Picco LM, Bozec L, Ulcinas A, Engledew DJ, Antognozzi M, Horton M, Miles MJ (2007) Breaking the speed limit with atomic force microscopy. Nanotechnology 18(4):0440301–0440304

Rost MJ, Crama L, Schakel P, van Tol E, van Velzen-Williams GBEM, Overgauw CF, ter Horst H, Dekker H, Okhuijsen B, Seynen M, Vijftigschild A, Han P, Katan AJ, Schoots K, Schumm R, van Loo W, Oosterkamp TH, Frenken JWM (2005) Scanning probe microscopes go video rate and beyond. Rev Sci Instrum 76(5):053710-1–053710-9

Yi KA, Veillette RJ (2005) A charge controller for linear operation of a piezoelectric stack actuator. IEEE Trans Control Syst Technol 13(4):517–526

Chapter 13
Noise in Nanopositioning Systems

Mechanical and electrical noise in nanopositioning systems is unavoidable and dictates the maximum positioning resolution. The major sources of noise include sensor noise, amplifier noise, and external disturbances. In this chapter, these noise sources are discussed and their influence on positioning resolution is evaluated.

13.1 Introduction

A key performance specification of a nanopositioner, or indeed many other controlled systems, is the resolution. The resolution is essentially the amount of random variation that remains at the output, even when the system is at rest. The resolution is critical for defining the smallest possible dimensions in a manufacturing processes or the smallest measurable features in an imaging application.

At this point, it is important to distinguish between resolution and trueness. While the resolution is a measure of noise and random variation, the trueness defines the position accuracy which includes errors such as sensor nonlinearity, abbe error, and cosine error. A discussion of nanopositioner accuracy and trueness is contained in Chap. 4 and Hicks et al. (1997).

Although resolution is a critical performance criteria, there is unfortunately no strict definition available in the literature. There is also no published industrial standards for the measurement or reporting of nanopositioner resolution. Predictably, this has led to a wide variety of fragmented techniques used throughout both academia and industry. As a result, it is extremely difficult to compare the performance of different control strategies or commercial products.

The most reliable method for the measurement of nanopositioner resolution is to utilize an auxiliary sensor that is not involved in the feedback loop. However, this requires a sensor with less additive noise and greater bandwidth than the displacement to be measured. Due to these strict requirements, the direct measurement approach is often impractical or impossible. Instead, the closed-loop positioning noise is usually predicted from measurements of the noise sources, such as sensor noise.

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_13,

In industrial and commercial applications, the methods used to measure and report closed-loop resolution are widely varied. Unfortunately, many of these techniques do not provide complete information and may even be misleading. For example, the RMS noise and resolution is commonly reported without mentioning closed-loop or measurement bandwidth, which is essentially meaningless. Other published techniques are misleading and may grossly underestimate the true positioning noise. One arguable technique is the use of time-domain responses with an undisclosed low bandwidth or filtered sensor. The true noise and resolution is effectively hidden by the low-pass dynamics of the sensor or filter. Another arguable practice is the prediction of resolution directly from a spectral density without integration over the frequency range. As a result of these varied reporting standards, care must be taken when interpreting the specifications of some commercial nanopositioner manufacturers.

In the academic literature, the practices for reporting noise and resolution also vary. The most common approach is to predict the closed-loop noise from measurements of the sensor noise (Sebastian et al. 2008; Fleming 2010). However, this approach can underestimate the true noise since the influence of the high-voltage amplifier is neglected. In the hard drive industry, the standard performance metric for resolution is the track pitch and the standard deviation of the measurement (Al Mamun and Ge 2005; Abramovitch and Franklin 2002). However, the main sources of error in a disk drive arise from aeroelastic effects and track eccentricities which are not present in a nanopositioning system.

In this chapter, the background theory required for a strict definition of resolution is presented in a tutorial fashion. The resolution is then defined in Sect. 13.3 followed by a description of typical system noise sources. With this background, it is then possible to quantify the closed-loop positioning noise and resolution in Sect. 13.5. Some illustrative numerical examples follow in Sect. 13.6.

In practice, the experimental characterization of random noise can be challenging. To ensure a statistically valid estimate, experimental guidelines are discussed in Sect. 13.7. These techniques are then applied to evaluate the closed-loop resolution of a piezoelectric tube nanopositioner in Sect. 13.8.

In complement to the frequency domain approach, resolution and positioning noise can also be evaluated with time-domain data. The background theory and experimental procedures for this approach are discussed in Sect. 13.9. An experimental example is also presented with similar results to the frequency domain approach.

The final topic in this chapter proposes a new and simple method for the measurement of closed-loop resolution. The "applied voltage" method requires only a single recording of the actuator voltage but is demonstrated to produce results identical to the more involved methods. This technique is widely applicable for resolution measurement in both academic and industrial nanopositioning systems.

13.2 Review of Random Processes

This section provides a tutorial introduction to the basic statistical tools necessary for the understanding of noise processes in nanopositioning applications.

13.2.1 Probability Distributions

A random variable can either be discrete or continuous. Discrete random variables have a finite number of possible values with associated probabilities of occurrence. A six-sided dice is an example of a discrete random variable, it has only six possible values. The noise processes that occur in nanopositioning applications are described by continuous random variables. An example of a continuous random variable is electronic voltage noise that has an effectively infinite number of possible values.

Since there are an infinite number of possible values within a fixed range, for example 0 and 1, the probability that a continuous random variable is *exactly* equal to any real value is infinitesimal. Thus, a continuous random process must be described by the probability that an outcome will lie within a certain range, not be equal a certain value. The probability distribution function $F_{\mathcal{X}}(x)$ is commonly used for this purpose. It is defined as the probability $P(\mathcal{X} \leq x)$ that a particular realization of a random variable $\mathcal{X}$ will be less than or equal to x.

An example of a probability distribution function is shown in Fig. 13.1a. This distribution represents a continuous random variable that is equally likely to occur between 0 and 1.

From the definition of the probability distribution function, it follows that

1. $F_{\mathcal{X}}(x) \to 0$ as $x \to -\infty$,
2. $F_{\mathcal{X}}(x) \to 1$ as $x \to \infty$, and
3. $F_{\mathcal{X}}(x)$ is a non-decreasing function of x.

Another useful method for describing the distribution of a random variable is the probability density function $f_{\mathcal{X}}(x)$. This function is the derivative of the probability distribution function, that is

$$f_{\mathcal{X}}(x) = \frac{\mathrm{d}}{\mathrm{d}x} F_{\mathcal{X}}(x). \tag{13.1}$$

The area under the density function between x_1 and x_2 represents the probability that a realization of $\mathcal{X}$ will occur between those values. That is,

$$\int_{x_1}^{x_2} f_{\mathcal{X}}(x)\, \mathrm{d}x = P(x_1 \leq \mathcal{X} \leq x_2). \tag{13.2}$$

13.2.2 Expected Value, Moments, Variance, and RMS

The mean value of a random process is the statistical average over a number of experiments or realizations. If the probability distribution function is known, the statistical mean can be computed exactly and is known as the expected value. The expected value of a continuous random variable $\mathcal{X}$ is

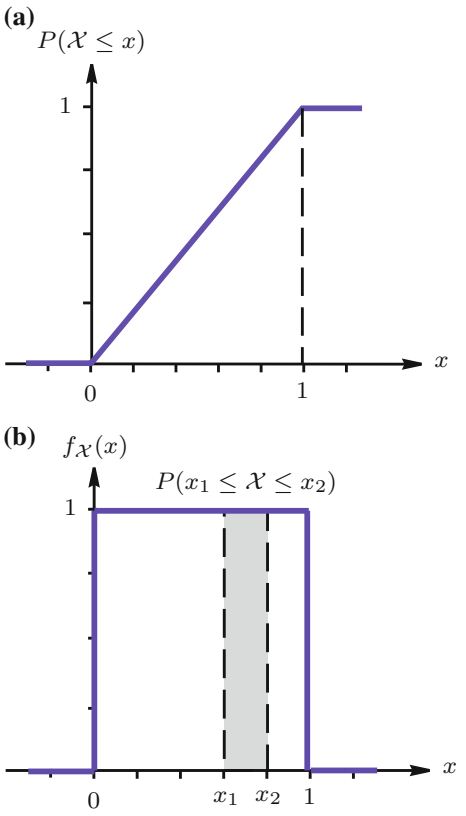


Fig. 13.1 Example of a probability distribution and density function. The random variable $\mathcal{X}$ is equally likely to occur between 0 and 1. **a** Probability distribution $F_{\mathcal{X}}(x)$. **b** Probability density function $f_{\mathcal{X}}(x)$

$$E[\mathcal{X}] = \int_{-\infty}^{\infty} x f_{\mathcal{X}}(x)\, \mathrm{d}x, \tag{13.3}$$

where x is a realization of $\mathcal{X}$

The expected value of a function of $\mathcal{X}$ can also be found,

$$E[g(\mathcal{X})] = \int_{-\infty}^{\infty} g(x) f_{\mathcal{X}}(x)\, \mathrm{d}x, \tag{13.4}$$

An example function of $\mathcal{X}$ is $g(\mathcal{X}) = \mathcal{X}^n$. The expected value of $\mathcal{X}^n$ is known as the nth *moment* of $\mathcal{X}$. Of particular interest are the first and second moments of $\mathcal{X}$. While the first moment is simply the expected value or mean value of $\mathcal{X}$, the second moment is given by

$$E\left[\mathcal{X}^2\right] = \int_{-\infty}^{\infty} x^2 f_{\mathcal{X}}(x)\, \mathrm{d}x. \tag{13.5}$$

The second moment of $\mathcal{X}$ is the expected value of $\mathcal{X}^2$ and indicates the amplitude of $\mathcal{X}$. For random variables that are not zero mean, it is preferable to remove the mean before calculating the second moment, this quantity is referred to as the *variance*. The variance of $\mathcal{X}$ is a measure of the dispersion of $\mathcal{X}$ about is mean, and is proportional to the power in the varying part of a random variable. Variance is defined by:

$$\mathrm{Var}\mathcal{X} = E\left[(\mathcal{X} - E[\mathcal{X}])^2\right]. \tag{13.6}$$

An alternative expression for the variance can be derived from Eq. (13.6) that can be easier to compute,

$$\mathrm{Var}\mathcal{X} = E\left[\mathcal{X}^2\right] - E[\mathcal{X}]^2. \tag{13.7}$$

Another term used to quantify the dispersion of a random process is the *standard deviation*. This is defined as the square root of variance and is usually given the symbol σ,

$$\sigma = \text{standard deviation of } \mathcal{X} = \sqrt{\mathrm{Var}\mathcal{X}}. \tag{13.8}$$

The standard deviation σ is equivalent to the Root-Mean-Square (RMS) value of a random process.

If a random process is the sum of two or more random processes, the total variance is the sum of each variance, that is,

$$\mathrm{Var}(\mathcal{X} + \mathcal{Y}) = \mathrm{Var}\mathcal{X} + \mathrm{Var}\mathcal{Y}. \tag{13.9}$$

Other properties of interest can be found in van Etten (2005), Brown and Hwang (1997).

13.2.3 Gaussian Random Variables

The Gaussian, or *normal* distribution is a family of distributions that closely approximate many physical noise sources. These include thermal noise and electronic noise. The probability density function of a Gaussian random variable is

$$f_{\mathcal{X}}(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}(x - m_{\mathcal{X}})\right], \tag{13.10}$$

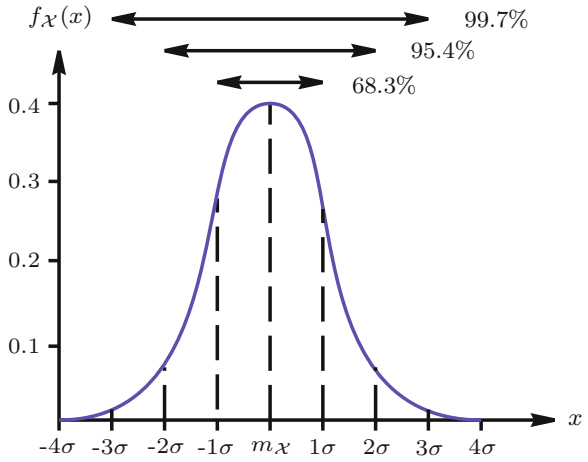


Fig. 13.2 The probability density function of a Gaussian or normally distributed process

where x is a realization of the random variable $\mathcal{X}$, and the only parameters are the mean $m_{\mathcal{X}}$, and variance σ^2. The Gaussian probability density function is plotted in Fig. 13.2.

Since a Gaussian distribution has only two parameters, the mean $m_{\mathcal{X}}$ and variance σ^2, the following shortened notation is often used,

$$\mathcal{X} \sim N(m_{\mathcal{X}}, \sigma^2). \tag{13.11}$$

Although no closed-form solution is available for the probability distribution function of a Gaussian random variable, numerical values are readily available from calculators and tables. However, for the purposes here, there is no need to find the probability for arbitrary values of x. Instead, we are interested in only three specific ranges of x, these ranges and the associated probabilities are known as the *empirical rule* and are listed below and shown in Fig. 13.2.

$P(-\sigma \leq \mathcal{X} \leq \sigma)$	68.3 %
$P(-2\sigma \leq \mathcal{X} \leq 2\sigma)$	95.4 %
$P(-3\sigma \leq \mathcal{X} \leq 3\sigma)$	99.7 %

The empirical rule quantifies the dispersion of a random variable. Of particular interest is the probability associated with the $\pm 3\sigma$ range, this states that 99.7 % of realizations will occur within a range of 6σ. Thus, the 6σ range is often used as an approximation of the peak-to-peak amplitude of a random variable.

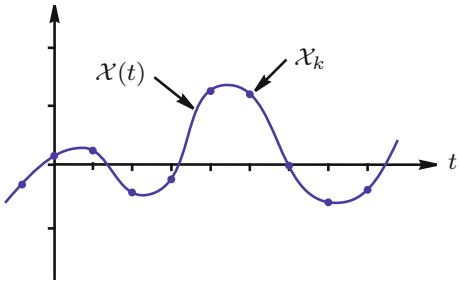


Fig. 13.3 Time evolution of a typical noise process $\mathcal{X}(t)$ and its samples $\mathcal{X}_k$

13.2.4 Continuous Random Processes

Up until now we have considered continuous random variables, i.e., realizations of a random variable with a continuous distribution of amplitude. This description does not convey any information on how a particular noise source changes with time. For example, consider the typical noise waveform plotted in Fig. 13.3. Such waveforms are referred to as *continuous random processes*, these processes have a continuous distribution of amplitude and are also continuous in time. To describe the time evolution of a random process, the concepts of stationarity, correlation, and higher order density functions must first be introduced. With these tools, we can describe a Gaussian random process, which will be used to approximate the random noise sources found in nanopositioning applications.

13.2.5 Joint Density Functions and Stationarity

The probability distribution function of a random process $\mathcal{X}(t)$ is defined at a fixed instance of time t_1,

$$F_{\mathcal{X}}(x_1; t_1) = P\left(\mathcal{X}(t_1) \leq x_1\right). \tag{13.12}$$

This definition can be extended directly to two random variables,

$$F_{\mathcal{X}}(x_1, x_2; t_1, t_2) = P\left(\mathcal{X}(t_1) \leq x_1, \mathcal{X}(t_2) \leq x_2\right), \tag{13.13}$$

where $F_{\mathcal{X}}(x_1, x_2; t_1, t_2)$ is referred to as the *joint probability distribution function.* Using a similar extension, the Nth order joint distribution function is,

$$F_{\mathcal{X}}(x_1, \cdots, x_N; t_1, \ldots, t_N) = P\left(\mathcal{X}(t_1) \leq x_1, \ldots, \mathcal{X}(t_N) \leq x_N\right). \tag{13.14}$$

The associated Nth order *joint probability density function* can be found by differentiation,

$$f_{\mathcal{X}}(x_1, \ldots, x_N; t_1, \ldots, t_N) = \frac{\partial^N F_{\mathcal{X}}(x_1, \ldots, x_N; t_1, \ldots, t_N)}{\partial x_1, \ldots, \partial x_N}. \tag{13.15}$$

If the joint probability density function of a random process does not change with time, the process is said to be *stationary*.

13.2.6 Correlation Functions

The *autocorrelation* function of a random process describes the relationship between adjacent samples of a noise source. For a stationary process, autocorrelation is defined by:

$$R_{\mathcal{X}}(\tau) = E\left[\mathcal{X}(t)\mathcal{X}(t+\tau)\right]. \tag{13.16}$$

Note that $R_{\mathcal{X}}(0)$ is simply the variance of a process. If the autocorrelation function decays quickly with τ, the amplitude of a process shows very little correlation between adjacent instances of time and thus varies extremely quickly. Likewise, slowly varying processes have "wide" autocorrelation functions.

The *cross-correlation* function of two random processes describes how similar the processes are. For stationary processes, cross-correlation is defined by:

$$R_{\mathcal{X}\mathcal{Y}}(\tau) = E\left[\mathcal{X}(t)\mathcal{Y}(t+\tau)\right]. \tag{13.17}$$

13.2.7 Gaussian Random Processes

Many noise sources in engineering and scientific applications can be adequately approximated by a Gaussian random process (van Etten 2005; Brown and Hwang 1997). A Gaussian random process is a continuous random process with a Gaussian distribution. The joint density function is defined by:

$$f_{\mathcal{X}}(x_1, \ldots, x_N; t_1, \ldots, t_N) = \frac{\sqrt{\left|\mathbf{C}_{\mathcal{X}}^{-1}\right|}}{(2\pi^{N/2})} \exp\left[-\frac{(\mathbf{x}-\mathbf{m}_{\mathcal{X}})^T \mathbf{C}_{\mathcal{X}}^{-1} (\mathbf{x}-\mathbf{m}_{\mathcal{X}})^T}{2}\right], \tag{13.18}$$

where

$$(\mathbf{x}-\mathbf{m}_{\mathcal{X}}) = \begin{bmatrix} x_1 - m_{\mathcal{X}} \\ \vdots \\ x_N - m_{\mathcal{X}} \end{bmatrix}, \tag{13.19}$$

and $\mathbf{C}_{\mathcal{X}}$ is the covariance matrix whose elements C_{ij} are dictated by the autocorrelation function $R_{\mathcal{X}}(t_i, t_j)$ where $i, j = 1 \ldots N$, that is

$$\mathbf{C}_{\mathcal{X}} = \begin{bmatrix} R_{\mathcal{X}}(t_1, t_1) & \cdots & R_{\mathcal{X}}(t_1, t_N) \\ \vdots & \ddots & \vdots \\ R_{\mathcal{X}}(t_N, t_1) & \cdots & R_{\mathcal{X}}(t_N, t_N) \end{bmatrix} \tag{13.20}$$

$$= \begin{bmatrix} \sigma^2 & \cdots & R_{\mathcal{X}}(t_1, t_N) \\ \vdots & \ddots & \vdots \\ R_{\mathcal{X}}(t_N, t_1) & \cdots & \sigma^2 \end{bmatrix}. \tag{13.21}$$

The key properties of a Gaussian process are:

- The first and higher order joint density functions of a Gaussian process are parameterized only by the process mean $m_{\mathcal{X}}$ and autocorrelation function $R_{\mathcal{X}}$.
- If a Gaussian process is filtered by a linear system, the result is another Gaussian process.

13.2.8 Power Spectral Density

The power spectral density $S_{\mathcal{X}}(f)$ of a random process represents the way in which the power or variance of the process is distributed across frequency f. For example, if the random process under consideration was measured in Volts V, the power spectral density would have the units of V^2/Hz.

The power spectral density can be found by either the averaged periodogram technique, or from the autocorrelation function. The periodogram technique involves averaging a large number of Fourier transforms of a random process,

$$2 \times E\left[\frac{1}{T}\left|\mathscr{F}\left\{\mathcal{X}_T(t)\right\}\right|^2\right] \Rightarrow S_{\mathcal{X}}(f) \text{ as } T \Rightarrow \infty. \tag{13.22}$$

This approximation becomes more accurate as T becomes larger or more records are used to compute the expectation. In practice, $S_{\mathcal{X}}(f)$ is best measured using a Spectrum or Network Analyzer, these devices compute the approximation progressively so that large time records are not required. Practical techniques for the measurement of power spectral density are discussed contained in Sect. 13.7.

The power spectral density can also be computed from the autocorrelation function. The relationships between autocorrelation and power spectral density are known as the Wiener–Khinchin relations, given by:

$$S_{\mathcal{X}}(f) = 2\mathscr{F}\left\{R_{\mathcal{X}}(\tau)\right\} = 2\int_{-\infty}^{\infty} R_{\mathcal{X}}(\tau)e^{-j2\pi f\tau}\,\mathrm{d}\tau, \text{ and} \tag{13.23}$$

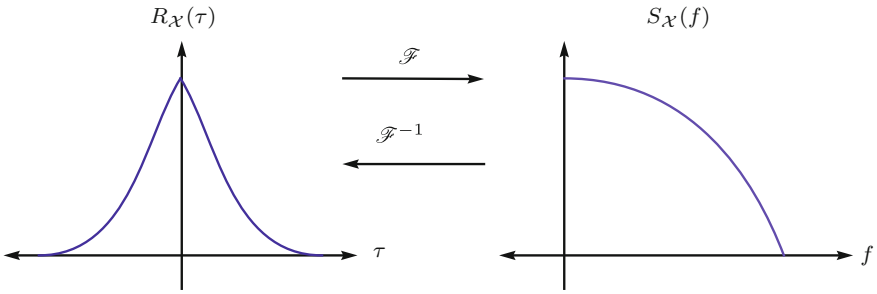


Fig. 13.4 The autocorrelation function $R_{\mathcal{X}}(\tau)$ and power spectral density $S_{\mathcal{X}}(f)$ are Fourier transform pairs

$$R_{\mathcal{X}}(\tau) = \frac{1}{2}\mathscr{F}^{-1}\left\{S_{\mathcal{X}}(f)\right\} = \frac{1}{2}\int_{-\infty}^{\infty} S_{\mathcal{X}}(f)e^{j2\pi f\tau}\,\mathrm{d}f, \tag{13.24}$$

or equivalently, if frequency is expressed in rad/s,

$$S_{\mathcal{X}}(\omega) = 2\mathscr{F}\left\{R_{\mathcal{X}}(\tau)\right\} = 2\int_{-\infty}^{\infty} R_{\mathcal{X}}(\tau)e^{-j\omega\tau}\,\mathrm{d}\tau, \text{ and} \tag{13.25}$$

$$R_{\mathcal{X}}(\tau) = \frac{1}{2}\mathscr{F}^{-1}\left\{S_{\mathcal{X}}(\omega)\right\} = \frac{1}{2}\frac{1}{2\pi}\int_{-\infty}^{\infty} S_{\mathcal{X}}(\omega)e^{j\omega\tau}\,\mathrm{d}\omega. \tag{13.26}$$

These relations are shown graphically in Fig. 13.4.

If the power spectral density is known, the variance of the generating process can be found from the area under the curve, i.e.,

$$\sigma^2 = E\left[\mathcal{X}^2(t)\right] = R_{\mathcal{X}}(0) = \int_0^{\infty} S_{\mathcal{X}}(f)\,\mathrm{d}f, \tag{13.27}$$

or equivalently, if the power spectral density is in rad/s,

$$\sigma^2 = E\left[\mathcal{X}^2(t)\right] = R_{\mathcal{X}}(0) = \frac{1}{2\pi}\int_0^{\infty} S_{\mathcal{X}}(\omega)\,\mathrm{d}\omega \tag{13.28}$$

If a random process is the sum of two or more random processes, the total power spectral density is the sum of the constituent spectral densities, that is, if $\mathcal{Z} = \mathcal{X} + \mathcal{Y}$,

$$S_Z(f) = S_X(f) + S_Y(f). \tag{13.29}$$

Other properties of interest can be found in van Etten (2005), Brown and Hwang (1997).

13.2.9 Filtered Random Processes

If a single-input single-output linear system $G(s)$ is excited by a stationary random process, labeled $u(t)$, the steady-state power spectral density of the output is

$$S_y(f) = |G(j2\pi f)|^2 \, S_u(f), \tag{13.30}$$

or equivalently,

$$S_y(\omega) = |G(j\omega)|^2 \, S_u(\omega), \tag{13.31}$$

where $S_u(f)$ and $S_y(f)$ are the power spectral densities of the input and output signals $u(t)$ and $y(t)$.

From Eqs. (13.30) and (13.31), if a constant gain k is applied to a random signal $u(t)$ with power spectral density $S_u(f)$, the resulting power spectral density is $k^2 S_u(f)$. The variance is also scaled by k^2.

Using the Weiner–Khinchin relations in Eqs. (13.25) and (13.25), the autocorrelation function, power spectral density, and variance of the output signal can be found from either the autocorrelation or power spectral density of the input. For example, using the Weiner–Khinchin relations and Eqs. (13.31) and (13.28), the variance of the output signal is

$$E\left[y^2(t)\right] = R_y(0) = \int_0^\infty S_y(f) \, \mathrm{d}f \tag{13.32}$$

$$= \int_0^\infty |G(j2\pi f)|^2 \, S_u(f) \, \mathrm{d}f, \tag{13.33}$$

or equivalently, with frequency in radian units,

$$E\left[y^2(t)\right] = \frac{1}{2\pi} \int_0^\infty |G(j\omega)|^2 \, S_u(\omega) \mathrm{d}\omega.$$

In general, the distribution function of the output signal $y(t)$ will not have the same form as the input, however, if the excitation is a Gaussian random process, the output is also a Gaussian random process, defined by the mean and autocorrelation

$$m_y = m_x G(0) \quad \text{and} \quad R_y(\tau) = \frac{1}{2}\mathscr{F}^{-1}\left\{|G(j2\pi f)|^2 S_u(f)\right\}. \tag{13.34}$$

13.2.10 White Noise

In situations where detailed statistical information about a noise process is not available, the process is often assumed to be filtered or band-limited *white noise*. White noise itself, is a stationary random process with a constant spectral density A, and an impulsive autocorrelation function, i.e.,

$$S_{\text{wn}}(f) = A \quad \text{and} \quad R_{\text{wn}}(\tau) = \frac{A}{2}\delta(\tau). \tag{13.35}$$

As the autocorrelation is impulsive, the variance and power of a white noise process are infinite. However this is not problematic as white noise processes are not used independently, but rather as driving processes for low-pass or bandpass systems such as position sensors, mechanical systems, and electronic circuits. If the system dynamics are not known, it is common to approximate a low-pass filtered white noise process as *band-limited white noise.* This process has a perfectly band-limited power spectral density, i.e.,

$$S_{\text{blwn}}(f) = \begin{cases} A, & |f| \leq f_c \\ 0, & |f| > f_c \end{cases}, \tag{13.36}$$

where f_c is the bandwidth in Hz. The corresponding autocorrelation, variance, and RMS value are

$$R_{\text{blwn}}(\tau) = f_0 A \frac{\sin(2\pi f_c \tau)}{2\pi f_c \tau}, \quad \sigma^2 = A f_c, \quad \text{and} \quad \sigma = \sqrt{A}\sqrt{f_c}. \tag{13.37}$$

In most cases, the spectrum of a noise signal will not be perfectly band limited; rather, the noise process will be filtered by a low-pass filter. This is a complication as the spectral density is no longer a constant, rather it rolls off slowly after the cut-off frequency of the filter. To remove this complication, it is possible to determine the *equivalent noise bandwidth* of the spectrum (van Etten 2005; Brown and Hwang 1997). This is the bandwidth of a perfectly band-limited white noise process with the same power as the filtered process. That is, the variance of a low-pass filtered white noise process can be represented as $\sigma^2 = A f_e$, where f_e is the equivalent noise bandwidth. The equivalent noise bandwidth can be readily determined for any system (van Etten 2005; Brown and Hwang 1997), including the commonly used filters listed in Table 13.1.

Table 13.1 The equivalent noise bandwidth f_e of commonly used filters with cut-off frequency f_c

Filter order	f_e
1	$1.57 \times f_c$
2	$1.11 \times f_c$
3	$1.05 \times f_c$
4	$1.025 \times f_c$

13.2.11 Spectral Density in V/$\sqrt{Hz}$

Rather than plotting the frequency distribution of power or variance σ^2, it is often convenient to plot the frequency distribution of standard deviation σ or Root-Mean-Square (RMS) value. This distribution will be referred to as the spectral density. It is related to the standard power spectral density function simply by a square root, i.e.,

$$\text{Spectral Density} = \sqrt{S_{\mathcal{X}}(f)}. \tag{13.38}$$

The units of $\sqrt{S_{\mathcal{X}}(f)}$ are units/$\sqrt{\text{Hz}}$ rather than units^2/Hz. The spectral density is preferred in the electronics literature as the RMS value of a noise process can be determined directly from the noise density and effective bandwidth. If the noise density is a constant c $V/\sqrt{\text{Hz}}$ and the process is perfectly band limited to $\pm f_c$, the RMS value or standard deviation of the resulting signal is $c\sqrt{f_c}$. Note that c is the noise density, which is equal to the square root of the power spectral density, i.e., $c = \sqrt{A}$. To distinguish between power spectral density and noise density, A will be used for power spectral density and $\sqrt{A}$ will be used for noise density.

An advantage of the spectral density is that a gain k applied to a signal $u(t)$ also scales the spectral density by k. This differs from the standard power spectral density function that must be scaled by k^2. It is important to note that in this work, power spectral density functions are assumed to be expressed in units^2/Hz unless otherwise stated.

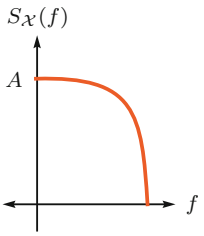
While the variance and power spectral density of multiple random processes can be summed, the standard deviation and spectral density must be square-summed, i.e.,

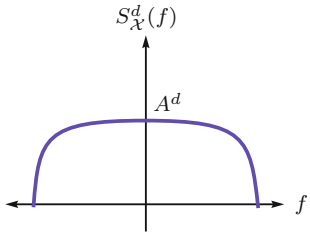
$$\sigma_{\mathcal{X}+\mathcal{Y}} = \sqrt{\sigma_{\mathcal{X}}^2 + \sigma_{\mathcal{Y}}^2}. \tag{13.39}$$

Other properties of interest can be found in van Etten (2005), Brown and Hwang (1997).

13.2.12 Single- and Double-Sided Spectra

Unfortunately for the user, there is an alternative definition to the power spectral density described in Sect. 13.2.8. The alternative definition uses the full spectrum





	Single Sided Spectrum	Double Sided Spectrum
Definition	$S_{\mathcal{X}}(f) = 2\int_{-\infty}^{\infty} R_{\mathcal{X}}(\tau)e^{-j2\pi f\tau}\,d\tau$	$S^d_{\mathcal{X}}(f) = \int_{-\infty}^{\infty} R_{\mathcal{X}}(\tau)e^{-j2\pi f\tau}\,d\tau$
Variance	$\sigma^2 = \int_0^{\infty} S_{\mathcal{X}}(f)\,df$	$\sigma^2 = \int_{-\infty}^{\infty} S^d_{\mathcal{X}}(f)\,df$
RMS Value	$\sigma = \int_0^{\infty} \sqrt{S_{\mathcal{X}}(f)}\,df$	$\sigma = \int_{-\infty}^{\infty} \sqrt{S^d_{\mathcal{X}}(f)}\,df$
White Noise	$\sigma^2 = A \times f_e$ $\sigma = \sqrt{A}\sqrt{f_e}$	$\sigma^2 = A^d \times 2f_e$ $\sigma = \sqrt{A^d}\sqrt{2f_e}$
Conversion	$S_{\mathcal{X}}(f) = 2S^d_{\mathcal{X}}(f)$ $\sqrt{S_{\mathcal{X}}(f)} = \sqrt{2}\sqrt{S^d_{\mathcal{X}}(f)}$	$S^d_{\mathcal{X}}(f) = \frac{1}{2}S_{\mathcal{X}}(f)$ $\sqrt{S^d_{\mathcal{X}}(f)} = \frac{1}{\sqrt{2}}\sqrt{S_{\mathcal{X}}(f)}$

Fig. 13.5 A comparison between the single-sided and double-sided power spectral density. f_e is the equivalent noise bandwidth discussed in Sect. 13.2.10

for calculation of the variance, hence it will be referred to as the double-sided power spectral density $S^{\mathrm{d}}_{\mathcal{X}}(f)$. The existence of these two definitions has resulted in much confusion so a detailed explanation is provided here.

The definitions of the single- and double-sided power spectral densities are compared in Fig. 13.5. The only difference is a factor of 2 in the single-sided definition which accounts for the missing negative frequencies in the equation for variance.

As the autocorrelation is an even function, and the Fourier transform of an even function is also even, the power spectral density is symmetrical around 0 Hz. Hence, only half of the power spectrum is required to calculate the variance or RMS value; however, a factor of 2 must be included.

The single-sided spectrum is more commonly used for noise calculations, for example, in electronics (Wulff 2006); however, the double-sided spectrum is also used in some signal processing applications. Although it is straightforward to convert between either definition, some care must be taken when converting between double-sided and single-sided spectral densities as this involves a square root of the factor 2. The conversion factors are listed in Fig. 13.5.

Note that when calculating the variance or RMS value of a double-sided spectrum, the noise bandwidth is equal to twice the cut-off frequency since the spectrum is defined from $-f_e$ to f_e. However, since the single-sided spectrum already has a factor of 2 built into the definition, the noise bandwidth is equal to the cut-off frequency.

Historically, the single-sided power spectral density was first proposed by Einstein in his 1914 article "Method for the determination of statistical values of observations regarding quantities subject to irregular fluctuations" Einstein (1914) (translated in Einstein (1987)). The later definition arose from the more natural relationship to the Fourier transform.

When measuring the power spectral density using a dynamic signal analyzer, it is important to note whether the data is a double-sided or single-sided spectrum, and perform a conversion if necessary.

13.3 Resolution and Noise

When a nanopositioner has settled to a commanded location, a small amount of random motion remains. This random motion dictates the *resolution* of the nanopositioning system. Although there is no standard definition for resolution, the ISO 5725 standard defines *precision* as the standard deviation (RMS Value) of a measurement.

In this work, the resolution will be defined as the the maximum steady-state peak-to-peak variation in actual position. This is equivalent to the minimum distance between two nonoverlapping points. The resolution of the x and y axis of a lateral nanopositioner, δ_x and δ_y, is illustrated graphically in Fig. 13.3. Since the peak-to-peak variation is closely related to the standard deviation (or RMS value), the above definition of resolution is proportional to the ISO definition for precision.

Referring to Fig. 13.3, observe that the amplitude of random motion occasionally exceeds the limits specified by the resolution. As the actual position is a random process with a large range of possible values, it is not practical to define a resolution based on the maximum possible variation. Instead, the resolution is specified together with a probability that the actual position will be within the resolution limit.

If the random position variation is assumed to be Gaussian distributed, the resolution can be quantified by the variance of the process. In Fig. 13.6b the probability density functions of the x axis position at a single point and the neighboring points are plotted. In this example, the resolution has been defined as δ_x, $\delta_y = 4\sigma$, which has an associated probability of 95.4 %. Restated, there is a 4.6 % chance that the position will exceed the resolution limit and stray into a neighboring area. This probability is shaded gray in Fig. 13.6b.

For nanopositioning applications, a 99.7 % probability that the position falls within $\delta_x = 6\sigma_x$ is an appropriate definition for the resolution. To be precise, this definition should be referred to as the 6σ-resolution and specifies the minimum spacing between two adjacent points that do not overlap 99.7 % of the time. Although there is no international standard for the measurement or reporting of resolution in a positioning system, the ISO 5725 Standard on Accuracy (Trueness and Precision) of

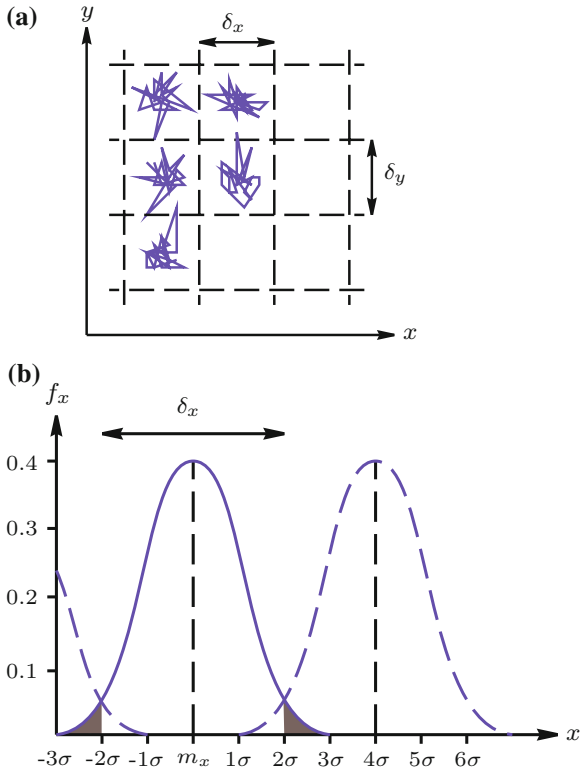


Fig. 13.6 **a** Random motion of a two-axis nanopositioner. δ_x and δ_y are the x- and y-axis resolutions. **b** The Gaussian distributed density function of the x-axis position, where σ and m_x are the standard deviation and mean, respectively. The shaded areas represent the probability of the position being outside the range specified by the resolution δ_x

Measurement Methods and Results (ISO 1994) defines precision as the standard deviation (RMS Value) of a measurement. Thus, the 6σ-resolution is equivalent to six times the ISO definition for precision.

13.4 Sources of Nanopositioning Noise

In the following subsections, the three major sources of noise in nanopositioning systems are discussed. These sources are the sensor noise, external noise, and the amplifier output voltage noise. The power spectral density of each source will be derived to allow estimation of the closed-loop position noise in Sect. 13.5.

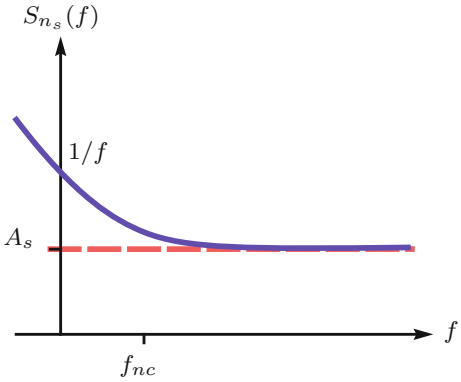


Fig. 13.7 Power spectral density of a baseband sensor (*solid line*) and a modulated sensor (*dashed line*). A_s is the noise density and f_{nc} is the $1/f$ noise corner frequency

13.4.1 Sensor Noise

The noise characteristics of a position sensor depend mainly on the physical method used for detection. These methods were discussed in detail in Chap. 5. Although there is a vast range of sensing techniques available, for the purpose of noise analysis, these can be grouped into two categories:

- **Baseband Sensors**. These sensors involve a direct measurement of position from a physical variable that is sensitive to displacement. Examples include resistive strain sensors, piezoelectric strain sensors, and optical triangulation sensors. The power spectral density of a baseband sensor is typically described by the sum of white noise and $1/f$ noise, where $1/f$ noise has a spectral density that is inversely proportional to frequency (van Etten 2005; Brown and Hwang 1997). $1/f$ noise is used to approximate the power spectrum of physical processes such as the flicker noise in resistors and current noise in transistor junctions. The power spectral density of a baseband sensor $S_{n_s}(f)$ can be written

$$S_{n_s}(f) = A_s \frac{f_{\text{nc}}}{|f|} + A_s, \tag{13.40}$$

  where A_s is the midband density, expressed in units^2/Hz and f_{nc} is the $1/f$ corner frequency. This function is plotted in Fig. 13.7.

- **Modulated Sensors**. In contrast to baseband sensors, modulated sensors use high-frequency excitation to detect position. Examples include capacitive sensors, inductive sensors and Linear Variable Displacement Transformers (LVDTs). Although these sensors require a demodulation process that inevitably adds noise, this disadvantage is far outweighed by the reduction of $1/f$ noise. The power spectral density $S_{n_s}(f)$ of a modulated sensor can generally be approximated by:

$$S_{n_s}(f) = A_s, \tag{13.41}$$

where A_s is the noise density, expressed in units^2/Hz. The power spectral density of a modulated sensor is compared to a baseband sensor in Fig. 13.7.

In the above discussion, the power spectral densities (13.40) and (13.41) were assumed to resemble white noise at high frequencies. In practice, however, all sensors contain low-pass dynamics that roll-off the power spectral density at high frequencies. In closed-loop nanopositioners, the position noise is determined mainly by the closed-loop bandwidth which is significantly lower in frequency than the sensor dynamics. Thus, in the following sections, sensor dynamics will be neglected.

In nanopositioning applications, modulated sensors are far preferable to baseband sensors as exhibit less $1/f$ noise. The inherent $1/f$ noise of a baseband sensor causes the measurement to drift around at low frequencies. Alike a Brownian process, the output of a baseband sensor does not vary about a mean and has a large variance.

As nanopositioners are required to perform well in both static and dynamic positioning applications, particular attention must be paid to the low-frequency sensor noise. Thus, the focus in following sections is on modulated sensors with an approximately constant noise spectral density.

13.4.2 External Noise

The external force noise exerted on a nanopositioner is highly dependent on the ambient environmental conditions and can not be generalized. Typically, the power spectral density will consist of broad spectrum background vibration with a number of narrow band spikes at harmonic frequencies of the mains power source and any local rotating machinery. Although the external force noise must be measured in-situ, for the purposes of simulation, it is useful to assume a white power spectral density A_w, i.e.,

$$S_w(f) = A_w. \tag{13.42}$$

Clearly, a white power spectral density will not provide an accurate estimate of externally induced position noise. However, it does illustrate the response of the control system to noise from this source. That is, it reveals whether the control system attenuates or amplifies external noise and over what frequency regions. A constant power spectral density of A_w will be used for this purpose the in following sections.

13.4.3 Amplifier Noise

The high-voltage amplifier is a key component of any piezoelectric actuated system. It amplifies the control signal from a few volts up to the hundreds of volts required to obtain full stroke from the actuator.

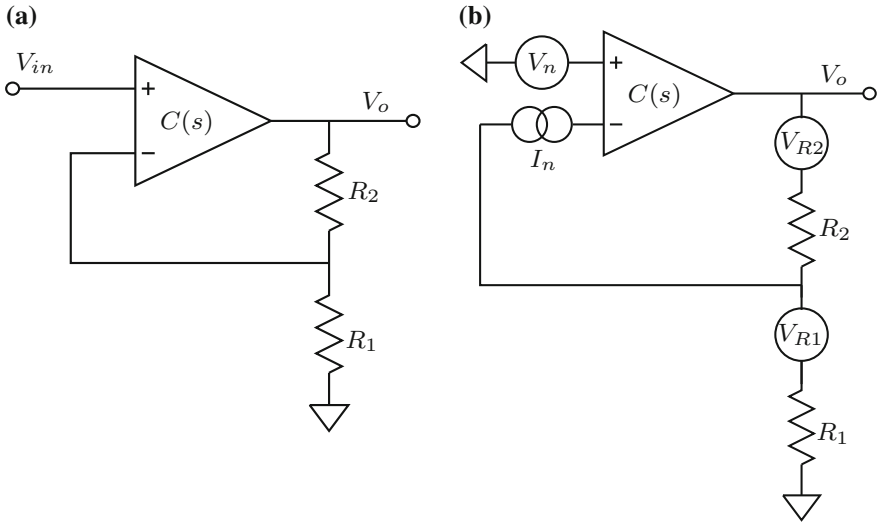


Fig. 13.8 Basic schematic of a voltage amplifier (**a**) and equivalent noise circuit (**b**). The noise sources V_n and I_n represent the equivalent input voltage noise and current noise of the amplifier. V_{R1} and V_{R2} are the thermal noise of the feedback resistors. **a** Voltage amplifier **b** Equivalent noise circuit

For the purpose of noise analysis, the simplified schematic diagram of a noninverting amplifier in contained in Fig. 13.8a. This model is sufficient to represent the characteristics of interest. The opamp represents the differential gain stage and output stage of the amplifier. As high-voltage amplifiers are often stabilized by a dominant pole, the open-loop dynamics can be approximated by a high-gain integrator, i.e., $C(s) = \alpha_{\text{ol}}/s$, where α_{ol} is the open-loop DC gain. With this approximation, the closed-loop output voltage is

$$V_{\text{o}} = \frac{\alpha_{\text{ol}}}{s}\left(V_{\text{in}} - V_{\text{o}}\frac{R_1}{R_2 + R_1}\right). \tag{13.43}$$

The closed-loop amplifier transfer function can then be derived as:

$$\frac{V_{\text{o}}}{V_{\text{in}}} = \frac{1}{\beta}\frac{\alpha_{\text{ol}}\beta}{s + \alpha_{\text{ol}}\beta}, \tag{13.44}$$

where β is the feedback gain $\frac{R_1}{R_2+R_1}$. The closed-loop DC gain and bandwidth are:

$$\begin{aligned} \text{DC Gain} &= \frac{1}{\beta} = \frac{R_2 + R_1}{R_1}, \\ \text{Bandwidth} &= \alpha_{\text{ol}}\beta = \alpha_{\text{ol}}\frac{R_1}{R_2 + R_1} \ \text{ rad/s}. \end{aligned} \tag{13.45}$$

Table 13.2 Example noise and resistance parameters for the amplifier shown Fig. 13.8b

	BJT circuit	JFET circuit
V_n	$10\,\text{nV}/\sqrt{\text{Hz}}$	$50\,\text{nV}/\sqrt{\text{Hz}}$
I_n	$10\,\text{pA}/\sqrt{\text{Hz}}$	$0.1\,\text{pA}/\sqrt{\text{Hz}}$
R_1	$10.5\,\text{k}\Omega$	$10.5\,\text{k}\Omega$
R_2	$200\,\text{k}\Omega$	$200\,\text{k}\Omega$

Two cases are considered, one where the differential input stage is constructed from Bipolar Junction Transistors (BJTs) and another where Junction Field Effect Transistors (JFETs) are used

Table 13.3 The output voltage noise contributions of the high-voltage amplifier circuit in Fig. 13.8, where k is Boltzmann's constant (1.38×10^{-23} j/K) and T is the temperature in Kelvin

Source	$V_\text{o} =$	BJT circuit (nV/$\sqrt{\text{Hz}}$)	JFET circuit (nV/$\sqrt{\text{Hz}}$)
Voltage noise V_n	$V_n \frac{R_2+R_1}{R_1}$	201	1,002
Current noise I_n	$I_n R_2$	2,000	20
R_1 noise = $\sqrt{4kTR_1}$	$\sqrt{4kTR_1}\frac{R_2}{R_1}$	166	229.3
R_2 noise = $\sqrt{4kTR_2}$	$\sqrt{4kTR_2}$	37	52.5
Total		2,024	1,030

The random noise of a high-voltage amplifier is dominated by the thermal noise of the feedback resistors and the noise generated by the amplifier circuit that precedes the most gain, i.e., the differential input stage. These noise processes are illustrated in Fig. 13.8b and are assumed to be Gaussian distributed white noise with spectral density expressed in nV or pA per $\sqrt{\text{Hz}}$. Typical values for the resistances and noise sources are shown in Table 13.2.

To find the spectral density of the output voltage, the contribution from each source must be computed then square-summed. The equations relating each noise source to the output voltage (Horowitz and Hill 1989) are collated in Table 13.3. Also included in Table 13.3 are the simulated noise values for the example parameters listed in Table 13.2. Both circuits have a gain of 20 achieved with a 200 and 10.5 kΩ feedback resistor network.

The difference between the two circuits is the choice of transistors in the input differential gain stage of the amplifier. One uses Bipolar Junction Transistors (BJTs) while the other uses Junction Field Effect Transistors (JFETs). While BJTs have a lower noise voltage than JFETs, they also exhibit significant current noise which renders them unsuitable in applications involving large source impedances. As the feedback resistor R_2 in a high-voltage amplifier is typically in the hundreds of kΩ or MΩ, the dominant noise process in a BJT circuit is always the current noise I_n. This is observed in the BJT example in Table 13.3. JFETs are not often used in low-noise applications as they exhibit higher voltage noise than BJT circuits. However, due to the extremely low-current noise of JFETs and the importance of current noise in this application, JFETs are preferable.

In Table 13.3 the output power spectral densities of the two example circuits are listed. To find the total RMS and peak-to-peak noise voltage, the equivalent noise bandwidth of the amplifier can be determined using Table 13.1. However, there is

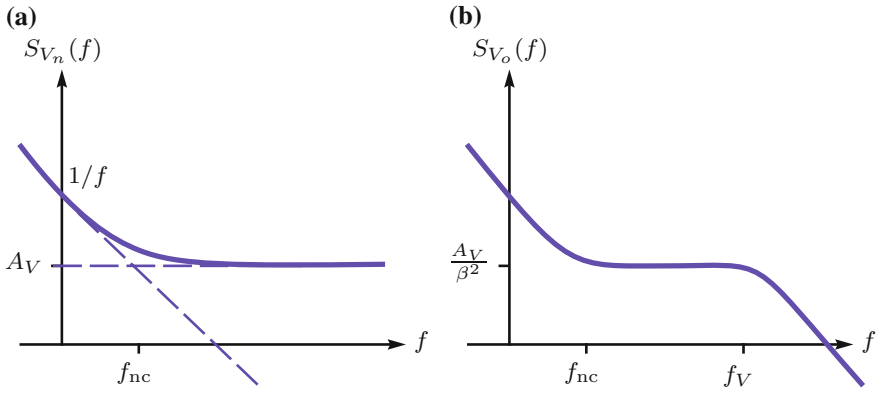


Fig. 13.9 Power spectral density of the input and output voltage noise of a high-voltage amplifier. **a** Input voltage noise V_n. **b** Output voltage noise V_o

an additional characteristic of voltage amplifiers that has not yet been taken into account. The noise power spectral densities of V_n and I_n are not constant, in fact they increase at low frequencies, this is referred to as $1/f$ noise or flicker noise. Taking the JFET example where current noise is negligible, the total amplifier noise is equal to the sum of a white noise process with density A_V and a noise process with spectral density inversely proportional to frequency. The frequency at which the two density curves intersect is known as the noise corner frequency f_{nc}. The corresponding power spectral density of V_n is plotted in Fig. 13.9a.

The noise density in Fig. 13.9a can be described as the sum of a white noise process and $1/f$ noise, i.e., the power spectrum can be written

$$S_{V_n}(f) = A_V \frac{f_{\mathrm{nc}}}{|f|} + A_V. \tag{13.46}$$

where f_{nc} is the noise corner frequency and A_V is the midband density, expressed in V^2/Hz.

Since the voltage noise V_n strongly dominates the output noise of a typical JFET amplifier, the other sources can be readily neglected. The power spectral density of the amplifier output voltage is then approximately

$$S_{V_\mathrm{o}}(f) = \left(A_V \frac{f_{\mathrm{nc}}}{|f|} + A_V\right) \frac{1}{\beta^2} \left| \frac{\alpha_{\mathrm{ol}}\beta}{j2\pi f + \alpha_{\mathrm{ol}}\beta} \right|^2, \tag{13.47}$$

$$= \frac{A_V}{\beta^2} \left(\frac{f_{\mathrm{nc}}}{|f|} + 1\right) \frac{f_V^2}{f^2 + f_V^2}, \tag{13.48}$$

where $f_V = \alpha_{\mathrm{ol}}\beta/2\pi$ is the closed-loop bandwidth of the amplifier (in Hz) and $1/\beta$ is the DC gain. The power spectral density of the output voltage noise V_o is plotted in Fig. 13.9b.

In addition to the power spectral density, the time-domain variance of the output voltage noise V_o is also of interest. This can be determined directly from the power spectral density,

$$E\left[V_o^2\right] = \frac{A_V}{\beta^2}\int_0^\infty \left(\frac{f_{nc}}{|f|}+1\right)\frac{f_V^2}{f^2+f_V^2} \tag{13.49}$$

$$= \frac{A_V}{\beta^2}\left(\int_0^\infty \frac{f_{nc}}{|f|}\frac{f_V^2}{f^2+f_V^2}\,\mathrm{d}f + \int_0^\infty \frac{f_V^2}{f^2+f_V^2}\,\mathrm{d}f\right) \tag{13.50}$$

In this expression there are two integral terms. The second integral term represents the variance of a first-order filter driven with white noise and can be evaluated using Table 13.1. The first integral can be evaluated with the following integral pair obtained from Poularikas (1999, 45.3.6.14)

$$\int \frac{1}{f}\frac{1}{bf^2+a}\mathrm{d}f = \frac{1}{2a}\log\frac{f^2}{bf^2+a}. \tag{13.51}$$

Rearranging (13.50) and substituting the result for the second-term yields

$$E\left[V_o^2\right] = \frac{A_V}{\beta^2}\left(f_{nc}f_V^2\int_0^\infty \frac{1}{f}\frac{1}{f^2+f_V^2}\,\mathrm{d}f + 1.57 f_V\right), \tag{13.52}$$

which can be solved with the integral pair (13.51) where $a = f_V^2$ and $b = 1$. The result is

$$E\left[V_o^2\right] = \frac{A_V}{\beta^2}\left(\frac{f_{nc}}{2}\log\left[\frac{f^2}{f^2+f_V^2}\right]_0^\infty + 1.57 f_V\right). \tag{13.53}$$

The first term in this equation is problematic as it represents a process of infinite variance. The reason for this is the low-frequency drift associated with $1/f$ noise. Alike a Brownian process, it drifts around at low frequencies and does not vary around a mean. In the analysis of devices that exhibit $1/f$ noise, i.e., opamps, it is preferable to make a distinction between drift and noise. Noise is defined as the varying part of a signal with frequency components above f_L Hz, while drift is defined as random motion below f_L Hz. In nanopositiong applications, a suitable choice for f_L is 0.1 Hz.

The expression for variance can be modified to include only frequencies above f_L,

$$E\left[V_o^2\right] = \frac{A_V}{\beta^2}\left(\frac{f_{nc}}{2}\log\frac{f_L^2+f_V^2}{f_L^2} + 1.57 f_V\right). \tag{13.54}$$

From this equation, two important properties can be observed:

1. The variance is not strongly dependent on f_L so the choice of this parameter is not critical; and
2. The variance is proportional to the noise corner frequency f_{nc}, so this parameter should be minimized at all costs.

For an example of the importance of $1/f$ noise, consider a standard voltage amplifier with a gain of 20, a bandwidth of 2 kHz, an input voltage noise density of 10,000 nV^2/Hz (or $100\,\text{nV}/\sqrt{\text{Hz}}$), and a noise corner frequency of 100 Hz. The total variance of the output voltage noise is $0.0165\,\text{mV}^2$, which is equivalent to an RMS value of 0.13 mV and a peak-to-peak amplitude of 0.77 mV. The $1/f$ noise accounts for 24 % of the variance.

If the noise corner frequency is increased by a factor of 10, the peak-to-peak noise approximately doubles to 1.4 mV and the $1/f$ noise now accounts for 76 % of the variance.

13.5 Closed-Loop Position Noise

In the previous section, it was concluded that the foremost sources of noise in a nanopositioning application are amplifier noise, sensor noise, and external noise. The spectral densities of these sources were summarized in Table 13.4. In this section, the closed-loop position noise due to each source is derived.

13.5.1 Noise Sensitivity Functions

To derive the closed-loop position noise, the response of the closed-loop system to each noise source must be considered. In particular, we need to specify the location where each source enters the feedback loop. The amplifier noise V_{o} appears at the plant input. In contrast, the external noise w acts at the plant output and the sensor noise n_s disturbs the measurement.

A single axis feedback loop with additive noise sources is illustrated in Fig. 13.10. For the sake of simplicity, the voltage amplifier is considered to be part of the controller. The transfer function from the amplifier voltage noise V_{o} to the position d is the input sensitivity function,

$$\frac{\text{d}(s)}{V_{\text{o}}(s)} = \frac{P(s)}{1 + C(s)P(s)}. \tag{13.55}$$

Likewise, the transfer function from the external noise w to the position d is the sensitivity function,

$$\frac{\text{d}(s)}{w(s)} = \frac{1}{1 + C(s)P(s)}. \tag{13.56}$$

Table 13.4 Summary of the foremost noise sources in a nanopositioning system

Noise source	Symbol	Power spectral density
Amplifier voltage noise	$S_{V_o}(f)$	$\frac{A_V}{\beta^2}\left(\frac{f_{nc}}{\lvert f\rvert}+1\right)\frac{f_V^2}{f^2+f_V^2}$
Sensor noise	$S_{n_s}(f)$	A_s
External noise	$S_w(f)$	A_w

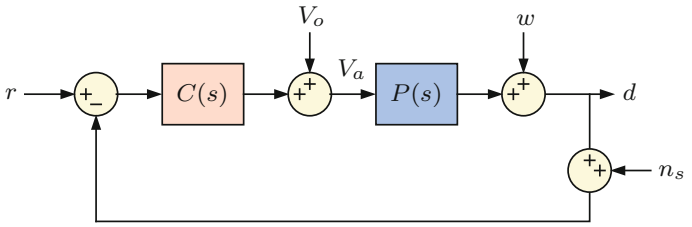


Fig. 13.10 A single axis feedback control loop with a plant P and controller C. The amplifier voltage noise V_o acts at the plant input, the external noise w effects the actual position, and the sensor noise n_s disturbs the measurement

Finally, the transfer function from the sensor noise n_s to the position d is the negated complementary sensitivity function,

$$\frac{\mathrm{d}(s)}{n_s(s)} = \frac{-C(s)P(s)}{1+C(s)P(s)}. \tag{13.57}$$

13.5.2 Closed-Loop Position Noise Spectral Density

With the knowledge of the sensitivity functions and the noise spectral densities, the spectral density of the position noise due to each source can be derived. The position noise spectral density due to the amplifier output voltage noise $S_{\mathrm{d}V_o}(f)$ is

$$S_{\mathrm{d}V_o}(f) = S_{V_o}(f)\left|\frac{\mathrm{d}(j2\pi f)}{V_o(j2\pi f)}\right|^2, \tag{13.58}$$

$$= \frac{A_V}{\beta^2}\left(\frac{f_{nc}}{|f|}+1\right)\frac{f_V^2}{f^2+f_V^2}\left|\frac{\mathrm{d}(j2\pi f)}{V_o(j2\pi f)}\right|^2. \tag{13.59}$$

Similarly, the position noise spectral density due to the external force noise $S_{\mathrm{d}w}(f)$ is

$$S_{\mathrm{d}w}(f) = S_w(f)\left|\frac{\mathrm{d}(j2\pi f)}{w(j2\pi f)}\right|^2, \tag{13.60}$$

$$= A_w \left| \frac{\mathrm{d}(j2\pi f)}{w(j2\pi f)} \right|^2. \tag{13.61}$$

Finally, the position noise spectral density due to the sensor noise $S_{\mathrm{d}n_s}(f)$ is

$$S_{\mathrm{d}n_s}(f) = S_{n_s}(f) \left| \frac{\mathrm{d}(j2\pi f)}{n_s(j2\pi f)} \right|^2, \tag{13.62}$$

$$= A_s. \left| \frac{\mathrm{d}(j2\pi f)}{n_s(j2\pi f)} \right|^2. \tag{13.63}$$

The total position noise spectral density $S_{\mathrm{d}}(f)$ is the sum of the three individual sources,

$$S_{\mathrm{d}}(f) = S_{\mathrm{d}V_{\mathrm{o}}}(f) + S_{\mathrm{d}w}(f) + S_{\mathrm{d}n_s}(f). \tag{13.64}$$

The position noise variance can then be found from Eq. (13.27)

$$E\left[d^2\right] = \int_0^\infty S_{\mathrm{d}}(f)\, \mathrm{d}f. \tag{13.65}$$

In general, this integral is best evaluated numerically as the spectral density and sensitivity functions can be complicated. An alternative method is to find the variance due to each noise source, then add them to find the overall variance. This alternative method can be useful for assessing the relative magnitude of each noise source.

If the noise is Gaussian distributed, the 6σ-resolution from Sect. 13.3 is

$$6\sigma\text{-resolution} = 6\sqrt{E\left[d^2\right]}. \tag{13.66}$$

13.5.3 Closed-Loop Noise Approximations with Integral Control

If a simple integral controller is used, i.e., $C(s) = \alpha/s$, the transfer function from the amplifier and external noise to displacement can be approximated by:

$$\frac{\mathrm{d}(s)}{V_{\mathrm{o}}(s)} = \frac{sP(0)}{s + \alpha P(0)}, \qquad \frac{\mathrm{d}(s)}{w(s)} = \frac{s}{s + \alpha P(0)}, \tag{13.67}$$

where $P(0)$ is the DC Gain of the plant. Likewise, the complimentary sensitivity function can be approximated by:

$$\frac{\mathrm{d}(s)}{n_s(s)} = \frac{\alpha P(0)}{s + \alpha P(0)}. \tag{13.68}$$

With these simplified approximations of the sensitivity functions, we can derive the closed-loop position noise spectral density. From (13.59) and (13.67) the position noise density due to the amplifier voltage noise $S_{dV_o}(f)$ is:

$$S_{dV_o}(f) \approx \frac{A_V P(0)^2}{\beta^2}\left(\frac{f_{nc}}{|f|}+1\right)\frac{f_V^2}{f^2+f_V^2}\frac{f^2}{f^2+f_{cl}^2}, \tag{13.69}$$

where $f_{cl} = \frac{\alpha P(0)}{2\pi}$ is the closed-loop bandwidth. The position noise due to the amplifier has a bandpass characteristic with a midband density of $A_V P(0)^2/\beta^2$.

From (13.61) and (13.68), the position noise density due to the external noise $S_{dw}(f)$ is

$$S_{dw}(f) \approx A_w \frac{f^2}{f^2+f_{cl}^2}, \tag{13.70}$$

which has a high-pass characteristic with a high-frequency density of A_w and a corner frequency equal to the closed-loop bandwidth.

The closed-loop position noise due to the sensor $S_{dn_s}(f)$ can be derived from (13.63) and (13.68), and is

$$S_{dn_s}(f) \approx A_s \frac{f_{cl}^2}{f^2+f_{cl}^2}, \tag{13.71}$$

which has a low-pass characteristic with a density of A_s and a corner frequency equal to the closed-loop bandwidth.

The spectral densities due to each source are plotted in Fig. 13.11. As the controller gain α is increased, the closed-loop bandwidth f_{cl} and sensor noise contribution also increases. However, a greater closed-loop bandwidth also results in attenuation of the amplifier voltage noise and external force noise. Hence, a lesser closed-loop bandwidth does not imply a lesser position noise, particularly if the amplifier or external force noise is significant. An important observation is that the amplifier bandwidth f_V should not be unnecessarily higher than the closed-loop bandwidth f_{cl}. In addition, if the sensor-induced noise is small compared to the amplifier induced noise, the closed-loop bandwidth should preferably be greater than the noise corner frequency of the voltage amplifier.

13.5.4 Closed-Loop Position Noise Variance

Due to the complexity of noise spectral density and sensitivity functions, the expression for variance (13.65) is generally evaluated numerically. However, in some cases it is straightforward and useful to derive analytic expressions. One such case is the position noise variance due to sensor noise ($E\left[d^2\right]$ due to n_s) when integral control is applied. As demonstrated in the forthcoming examples, sensor noise is typically

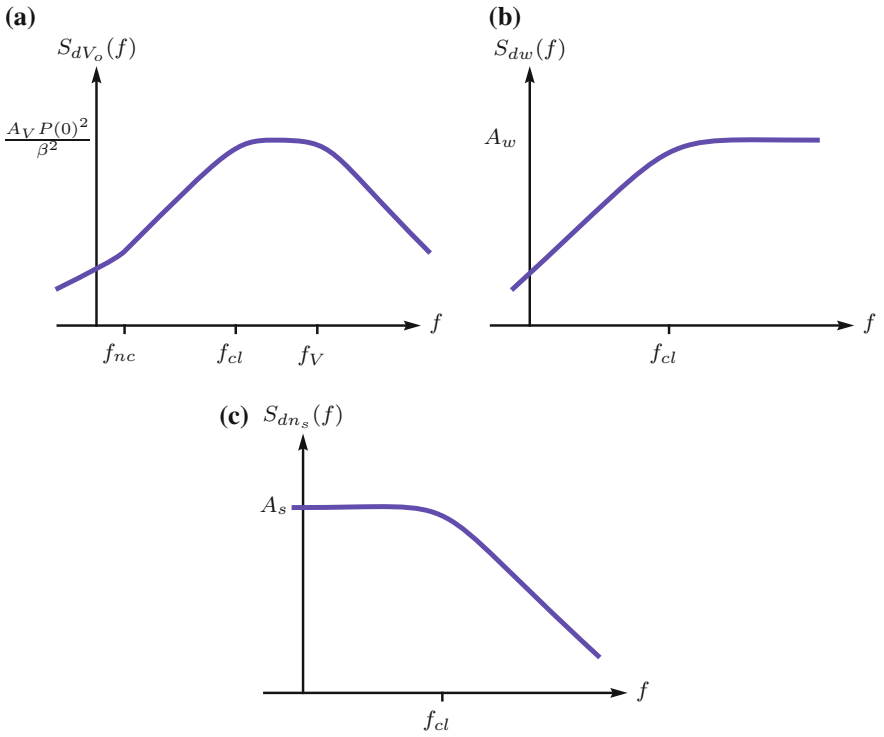


Fig. 13.11 The position noise power spectral density due to the amplifier voltage noise, external disturbance, and sensor noise. **a** The position noise power spectral density due to amplifer voltage noise $S_{\mathrm{dV_o}}(f)$. **b** The position noise power spectral density due to external noise $S_{\mathrm{dw}}(f)$. **c** The position noise power spectral density due to sensor noise $S_{\mathrm{d}n_s}(f)$

the dominant noise process in a feedback controlled nanopositioning system. As a result, other noise sources can sometimes by neglected.

The position noise variance due to the sensor noise can be easily determined from Table 13.1 as the source spectral density is constant and the sensitivity function (13.68) is approximately a first-order low-pass filter. Thus, the position noise variance and RMS value can be determined analytically,

$$E\left[d^2\right] \text{ due to } n_s = A_s \times 1.57 f_{\mathrm{cl}}, \ (A_s \text{ expressed in nm}^2/\mathrm{Hz}) \tag{13.72}$$

$$\sqrt{E\left[d^2\right]} \text{ due to } n_s = \sqrt{A_s}\sqrt{1.57 f_{\mathrm{cl}}}, \ (\sqrt{A_s} \text{ expressed in nm}/\sqrt{\mathrm{Hz}}) \tag{13.73}$$

where $f_{\mathrm{cl}} = \frac{\alpha P(0)}{2\pi}$ is the closed-loop bandwidth in Hz. The corresponding 6σ-resolution is

$$6\sigma\text{-resolution} = 6\sqrt{A_s}\sqrt{1.57 f_{\mathrm{cl}}}. \tag{13.74}$$

This expression can be used to determine the minimum resolution of a nanopositioning system given only the sensor noise density and closed-loop bandwidth. It can also be rearranged to reveal the maximum closed-loop bandwidth achievable given the sensor noise density and required resolution.

$$\text{maximum bandwidth (Hz)} = \frac{(6\sigma\text{-resolution})^2}{56.5 A_s} = \left(\frac{6\sigma\text{-resolution}}{7.51\sqrt{A_s}}\right)^2. \quad (13.75)$$

For example, consider a nanopositioner with integral feedback control and a capacitive sensor with a noise density of 30 pm/$\sqrt{\text{Hz}}$ (900 pm^2/Hz). The maximum bandwidth with a resolution of 1 nm is

$$\begin{aligned}\text{maximum bandwidth} &= \left(\frac{1 \times 10^{-9}}{7.51 \times 30 \times 10^{-12}}\right)^2 \\ &= 11\,\text{Hz}.\end{aligned}$$

13.5.5 A Note on Units

In the discussion thus far it has been assumed that the nanopositioner model $P(s)$ in Fig. 13.10 has an output equal to position, preferably in nanometers. In practice however, this signal will be the output voltage of a displacement sensor with sensitivity, k V/nm or $1/k$ nm/V. Rather than incorporating an additional gain into the equations above, it is preferable to simply perform the analysis with respect to the output voltage, then scale the result accordingly.

For example, if a nanopositioner has an output sensor voltage of 1 mV/nm, the noise analysis can be performed to find the spectral density and variance of the sensor voltage. Once the final power spectral density has been found, it can be scaled to nanometer by multiplying by $1/k^2$, which in this case is $1/(1 \times 10^{-3})^2$. Alternatively, the RMS Value or 6σ-resolution can be found in terms of the sensor voltage then multiplied by $1/k$.

13.6 Simulation Examples

13.6.1 Integral Controller Noise Simulation

In this section, an example nanopositioner is considered with a range of 100 µm at 200 V and a resonance frequency of 1 kHz. The system model is

$$P(s) = 500\,\text{nm/V} \times \frac{\omega_r^2}{s^2 + 2\omega_r\zeta_r s + \omega_r^2}, \quad (13.76)$$

Table 13.5 Specifications of an example nanopositioning system

Parameter	Value	Alternate units
Closed-loop bandwidth f_{cl}	50 Hz	–
Controller gain α	314	–
Amplifier bandwidth f_V	2 kHz	–
Amplifier gain $1/\beta$	50	–
Amplifier input voltage noise A_V	$100\,\text{nV}/\sqrt{\text{Hz}}$	$10{,}000\,\text{nV}^2/\text{Hz}$
Amplifier output voltage noise	$5\,\mu\text{V}/\sqrt{\text{Hz}}$	$25\,\mu\text{V}^2/\text{Hz}$
Amplifier noise corner frequency f_{nc}	100 Hz	–
-Sensor noise A_s	$20\,\text{pm}/\sqrt{\text{Hz}}$	$400\,\text{pm}^2/\text{Hz}$
Position range	100 μm	–
Sensitivity $P(0)$	500 nm/V	–
Resonance frequency ω_r	$2\pi \times 10^3$ r/s	1 kHz
Damping ratio ζ_r	0.05	–

where $\omega_r = 2\pi\ 1{,}000$ and $\zeta_r = 0.05$. The system includes a capacitive position sensor and voltage amplifier with the following specifications.

- The capacitive position sensor has a noise density of $20\,\text{pm}/\sqrt{\text{Hz}}$.
- The voltage amplifier has a gain of 20, a bandwidth of 2 kHz, an input voltage noise density of $100\,\text{nV}/\sqrt{\text{Hz}}$, and a noise corner frequency of 100 Hz.

The feedback controller in this example is a simple integral controller with compensation for the sensitivity of the plant, i.e.,

$$C(s) = \frac{1}{500\,\text{nm/V}}\frac{\alpha}{s}, \tag{13.77}$$

where α is the gain of the controller and also the approximate bandwidth (in rad/s) of the closed-loop system. All of the system parameters are summarized in Table 13.5.

With the noise characteristics and system dynamics defined, the next step is to compute the spectral density of the position noise due to amplifier voltage noise. From Eq. (13.59),

$$S_{\text{d}V_{\text{o}}}(f) = S_{V_{\text{o}}}(f)\left|\frac{\text{d}(j2\pi f)}{V_{\text{o}}(j2\pi f)}\right|^2 \tag{13.78}$$

$$= \frac{A_V}{\beta^2}\left(\frac{f_{\text{nc}}}{|f|}+1\right)\frac{f_V^2}{f^2+f_V^2}\left|\frac{P(j2\pi f)}{1+C(j2\pi f)P(j2\pi f)}\right|^2. \tag{13.79}$$

The power spectral density of position noise due to the sensor noise can also be found from Eq. (13.63)

$$S_{\mathrm{d}n_s}(f) = S_{n_s}(f)\left|\frac{\mathrm{d}(j2\pi f)}{n_s(j2\pi f)}\right|^2 \tag{13.80}$$

$$= A_s\left|\frac{C(j2\pi f)P(j2\pi f)}{1+C(j2\pi f)P(j2\pi f)}\right|^2. \tag{13.81}$$

The total density of the position noise can now be calculated from Eq. (13.64). The spectral density of the position noise due to both the amplifier $S_{\mathrm{d}V_o}(f)$ and sensor $S_{\mathrm{d}n_s}(f)$, together with the total position noise $S_{\mathrm{d}}(f)$ are plotted in Fig. 13.12a. Clearly, the sensor noise is the dominant noise process. This is the case in most nanopositioning systems with closed-loop position feedback.

The variance of the position noise can be determined by solving the integral for variance numerically,

$$\sigma^2 = E\left[d^2\right] = \int_0^\infty S_{\mathrm{d}}(f)\,\mathrm{d}f \tag{13.82}$$

The result is

$$\sigma^2 = 0.24\,\mathrm{nm}^2, \text{ and } \sigma = 0.49\,\mathrm{nm},$$

which implies a 6σ-resolution of 2.9 nm.

In systems with lower closed-loop bandwidth, the $1/f$ noise of the amplifier can become dominant. For example, if the closed-loop bandwidth of the previous example is reduced to 1 Hz, the new power spectral density, plotted Fig. 13.12b, differs significantly. The resulting variance and standard deviation are

$$\sigma^2 = 0.093\,\mathrm{nm}^2, \text{ and } \sigma = 0.30\,\mathrm{nm},$$

which implies a 6σ-resolution of 1.8 nm, not a significant reduction considering that the closed-loop bandwidth has been reduced to 2 % of its previous value. More generally, the resolution can be plotted against a range of closed-loop bandwidths to reveal the trend. In Fig. 13.13, the 6σ-resolution is plotted against a range of closed-loop bandwidths from 100 mHz to 60 Hz. The curve has a well-defined minima of 1.8 nm at 0.4 Hz. Below this frequency amplifier noise is the major contributor, while at higher frequencies sensor noise is more significant.

13.6.2 Noise Simulation with Inverse Model Controller

In the previous example, the integral controller does not permit a closed-loop bandwidth greater than 100 Hz. Many other model-based controllers can achieve much better performance. One simple controller that demonstrates the noise characteristics of a model-based controller is the combination of an integrator and notch filter, or direct inverse controller. The transfer function is simply an integrator combined with an inverse model of the plant,

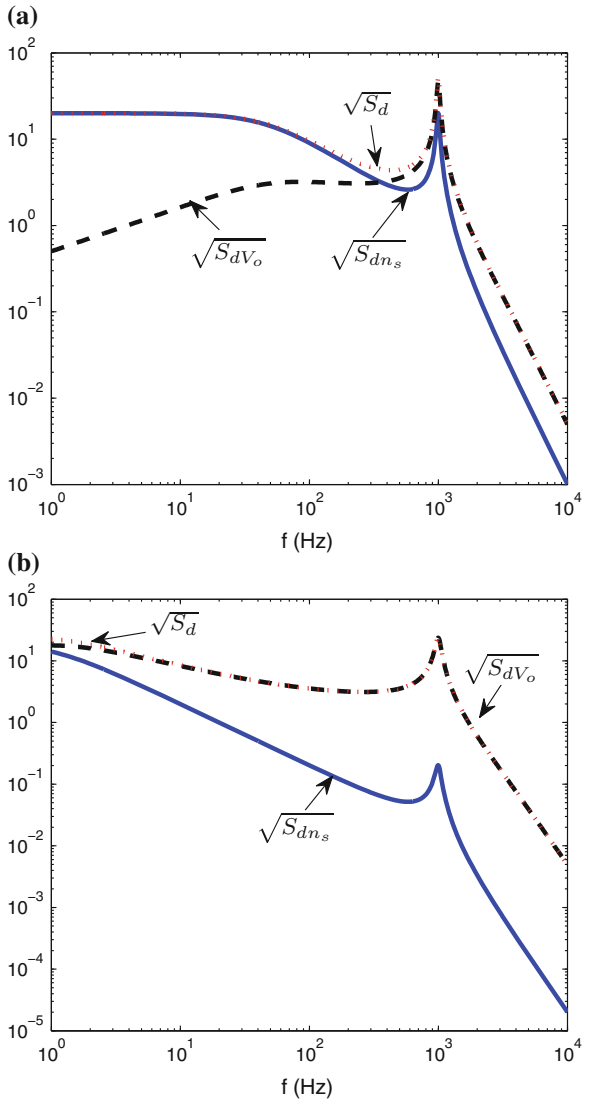


Fig. 13.12 The spectral density of the total position noise $\sqrt{S_{\mathrm{d}}(f)}$ and its two components, the amplifier output voltage noise $\sqrt{S_{\mathrm{dV_o}}(f)}$ and sensor noise $\sqrt{S_{\mathrm{dn_s}}(f)}$ (all in pm/$\sqrt{\mathrm{Hz}}$). With a 50 Hz bandwidth (**a**), the total noise is primarily due to the sensor. However, with a lower bandwidth of 1 Hz (**b**), the noise is dominated by the voltage amplifier. **a** 50 Hz Closed-loop bandwidth **b** 1 Hz Closed-loop bandwidth

$$C(s) = \frac{\alpha}{s}\frac{1}{500\,\mathrm{nm/V}}\frac{s^2 + 2\omega_r\zeta_r s + \omega_r^2}{\omega_r}. \tag{13.83}$$

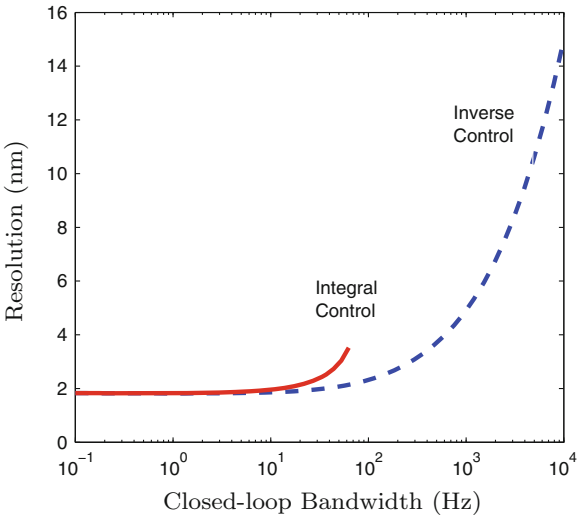


Fig. 13.13 Resolution of the example nanopositioning system with integral control (*solid line*) and inverse control (*dashed*). When the closed-loop bandwidth is less than 10 Hz, the resolution is limited by the amplifier noise. At greater closed-loop bandwidth, the sensor noise becomes dominant. The premature degradation of the integral controller resolution is due to the low gain-margin and resonant closed-loop response

The resulting loop-gain $C(s)P(s)$ is an integrator, so stability is guaranteed and the closed-loop bandwidth is α rad/s. With such a controller, it is now possible to examine the noise performance of feedback systems with wide bandwidth.

The noise spectral densities with a closed-loop bandwidth of 500 Hz are plotted in Fig. 13.14. The major difference from the case without inverse dynamics is the lack of a resonance peak in the sensor-noise spectrum. The resulting variance and standard deviation are:

$$\sigma^2 = 0.37\,\text{nm}^2, \text{ and } \sigma = 0.61\,\text{nm},$$

which implies a 6σ-resolution of 3.7 nm. This is not significantly greater than the 50 Hz controller bandwidth in the previous example, which resulted in a 2.9 nm resolution. When the closed-loop bandwidth of the inverse controller is reduced to 50 Hz, the resolution is 2.1 nm, which is slightly better than the previous example. The difference is due to the absence of the resonance peak in the sensor-induced noise.

The resolution of the inverse controller is plotted for a wide range of bandwidths in Fig. 13.13. The minimum resolution is 1.8 nm at 1 Hz. After approximately 100 Hz, the position noise is predominantly due to the sensor-noise which has a constant density but increasing bandwidth. The latter part of the curve is proportional to the square root of closed-loop bandwidth. This relationship can be explained by considering Eq. (13.74), which is equivalent to

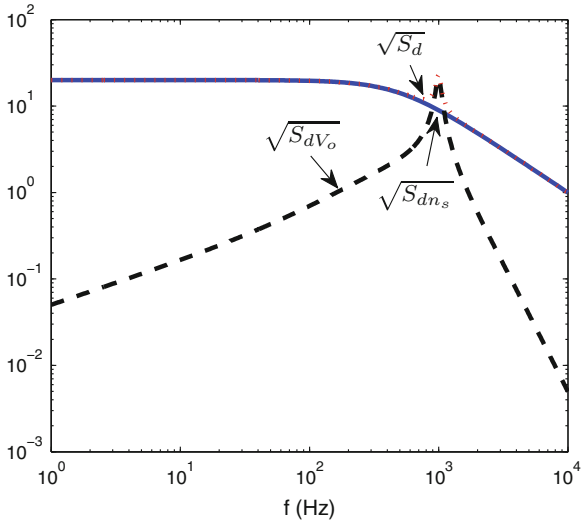


Fig. 13.14 The spectral density, with closed-loop inversion-based control, in pm/$\sqrt{\text{Hz}}$ of the total position noise $\sqrt{S_{\text{d}}(f)}$ and its two components, the amplifier output voltage noise $\sqrt{S_{\text{d}V_{\text{o}}}(f)}$ and sensor noise $\sqrt{S_{dn_s}(f)}$

$$6\sigma\text{-resolution} = 6\sqrt{A_s}\sqrt{1.57 f_{\text{cl}}}, \tag{13.84}$$

where f_{cl} is the closed-loop bandwidth and $\sqrt{A_s}$ is the sensor-noise density in nm/$\sqrt{\text{Hz}}$. Thus, where sensor noise is dominant, the resolution is proportional to the square root of closed-loop bandwidth.

13.6.3 Feedback Versus Feedforward Control

A commonly discussed advantage of feedforward control systems is the absence of sensor-induced noise. However, this view does not take into account the presence of $1/f$ amplifier noise that can result in significant peak-to-peak amplitude. Here we will compare the noise performance of feedback and feedforward control systems.

It is not necessary to derive equations for the noise performance of feedforward systems as this is a special case of the feedback examples already discussed in Sects. 13.6.1 and 13.6.2. The positioning noise of a feedforward control system is equivalent to a feedback control system when $C(s) = 0$ or equivalently, when the closed-loop bandwidth is zero. Thus, the plots of resolution versus bandwidth in Fig. 13.13 are also valid for feedforward control. The feedforward controller resolution is simply the DC resolution of these plots, which in both cases is 2.60 nm.

It is interesting to note that both the integral and inverse controller can achieve slightly less positioning noise than a feedforward control system when the closed-

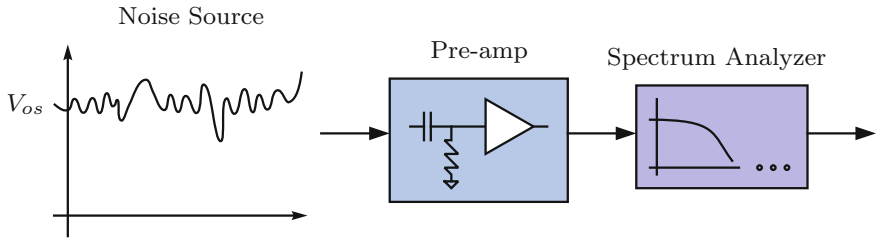


Fig. 13.15 A frequency domain noise measurement with a preamplifier and spectrum analyzer

loop bandwidth is very low. This is due to the amplifier $1/f$ noise that produces greater noise density than the sensor at very low frequencies. In the examples considered, the optimal noise performance could be achieved with a feedback controller of around 1 Hz bandwidth. To increase the positioning bandwidth, a feedforward input is required. The design of combined feedback and feedforward controllers is discussed in Sect. 13.6.3.

13.7 Practical Frequency Domain Noise Measurements

The use of a spectrum analyzer to measure noise directly in the frequency domain has many advantages over time-domain recordings. First, the inputs to a spectrum analyzer are equipped with dynamic signal scaling so that low amplitude signals can easily be dealt with. Second, spectrum analyzers record a very large amount of low-information data, and through averaging and Fourier transformation, create a small amount of high-information data. In addition, representation of noise in the frequency domain provides intuitive information on the nature of the noise and also a better understanding of how it will contribute to the closed-loop position noise of a nanopositioner.

13.7.1 Preamplification

As the amplitude of a typical noise signal is too much small to be applied directly to a spectrum analyzer, it must first be amplified. The signal-path of a noise measurement experiment is illustrated in Fig. 13.15. A low-noise preamplifier is used between the noise signal and spectrum analyzer. Its purpose is to remove offset voltage and to amplify the signal from microvolts or millivolts to around 100 mV RMS or greater.

In addition to the capacity for a large gain, there are two important preamplifier characteristics that must by considered. These are the signal coupling and input noise.

The signal coupling refers to the handling of the input signal before the main gain stage. Typically AC or DC coupling is available. A DC-coupled amplifier applies

the input signal directly to the main gain stage. Due to the large offset voltage usually present in nanopositioning applications, the output voltage will saturate at the required gain. The offset can be manually nulled, but offset drift will likely result in saturation before the experiment is complete.

To remove the offset voltage and low-frequency drift, an AC-coupled preamplifier uses a first-order high-pass filter so that only the varying component of the noise is amplified. However, AC-coupling in some instruments implies a cut-off frequency of up to 20 Hz. This is intolerably high in nanopositioning applications where the cut-off frequency should be less than 0.1 Hz. Noise components with frequency less that 0.1 Hz are usually referred to as drift and are not considered here. Most specialty low-noise preamplifiers have the provision for a low-frequency high-pass filter, for example, the SR560 low-noise amplifier has a high-pass cut-off frequency of 0.03 Hz.

When utilizing low-frequency filters, it is important to allow the transient response of the filter to decay before recording data. When measuring small AC signals with large DC components, it may take in excess of 20 time constants for the transient response to become negligible. With an AC-coupling frequency of 0.03 Hz, the required delay is approximately 100 s. In general, the measurement delay T_D should be at least

$$T_D = \frac{20}{2\pi f_c} \tag{13.85}$$

where f_c is the high-pass filter cut-off.

The voltage noise of the preamplifier is also important. The easiest method is simply to neglect it, which is possible if the RMS spectral density is less than one-tenth of the spectral density generated by the high-voltage amplifier. As noise sources are summed by power, if the spectral density of the instrument noise is one-tenth of noise to be measured, the resulting error will be less than 1 %, which is negligible. A simple technique for testing the noise floor is to connect the preamplifier to the noise source and increase the gain until the desired signal amplitude is reached, for example ± 1 V. Record the signal level, then disconnect the noise source, and short circuit the input to the preamplifier. If the resulting signal level is less than one-tenth of the previous measurement, then the preamplifier noise can be neglected. This simple test is valid when the spectral density of the noise source is relatively constant. A more thorough test involves recording the spectral density with, and without the noise source connected, then comparing results.

An example of an amplifier which is useful in noise measurement experiments is the Stanford Research SRS560. Some of the key specifications are listed in Table 13.6. The provision for a differential input can be extremely important in reducing the impact of voltage differences between the ground potentials of the noise source and preamplifier.

Table 13.6 Properties of the SR560 low-noise amplifier (Stanford Research Systems, Sunnyvale, CA)

Inputs	Single or differential
Noise	$4\,\text{nV}/\sqrt{\text{Hz}}$ (at 1 kHz)
Gain	1–50,000
Bandwidth	300 kHz (gain <1,000)
AC-coupling cut-off	0.03 Hz
Power	Battery or AC line

13.7.2 Spectrum Estimation

With suitable preamplification, the power spectral density of a noise source can be estimated from a recording of time-domain data. The three most commonly used techniques for power spectral estimation are the Bartlett's Method (averaging periodograms), Welch Method (averaging modified periodograms), and the Blackman-Tukey Method (smoothed periodogram) (Proakis and Manolakis 1996). These methods are compared in Proakis and Manolakis (1996) and are comparable in terms of processing requirements and estimation quality.

Bartlett's method, or the averaged periodograms method, involves subdividing an N point time record x into K nonoverlappping segments, where each segment has length M (Proakis and Manolakis 1996). This results in the K data segments

$$x_i(n) = x(n + iM) \qquad \begin{matrix} i = 0, 1, \ldots, K-1 \\ n = 0, 1, \ldots, M-1 \end{matrix}. \tag{13.86}$$

For each section, we compute the periodogram

$$S_x^i(f) = \frac{1}{M}\left|\sum_{n=0}^{M-1} x_i(n) e^{-j2\pi fn}\right|^2 \qquad \begin{matrix} i = 0, 1, \ldots, K-1 \\ f = \frac{F_s}{M}(0, 1, \ldots, M-1) \end{matrix}, \tag{13.87}$$

which is equivalent to

$$S_x^i(f) = \frac{1}{M}\left|\text{DFT}\left(x_i(n)\right)\right|^2 \qquad i = 0, 1, \ldots, K-1. \tag{13.88}$$

The periodograms are then averaged to compute the Bartlett's spectral estimate

$$S_x^{\text{d}}(f) = \frac{1}{K}\sum_{i=0}^{K-1} S_x^i(f). \tag{13.89}$$

Note that the Bartlett's estimate produces a double-sided spectrum as defined in Sect. 13.2.12. To obtain a single-sided spectrum, as used throughout this work, a correction factor must be applied and the frequencies beyond the Nyquist rate discarded, i.e.,

$$S_x(f) = 2S_x^{\mathrm{d}}(f) \qquad f = \frac{F_s}{M}(0, 1, \ldots, M/2)\,. \tag{13.90}$$

When using a spectrum analyzer to record the power spectral density, the instrument collects each segment individually then updates a running average of the estimate. This is convenient as it avoids the need to record a large amount of time-domain data. It also allows the user to assess the variance of the data in real time which is a simple method for deciding how long to run the experiment.

Some points to consider when measuring a noise power spectral density:

- Regardless of whether a window function is used or not, the finite data length of each segment will result in windowing distortion. This distortion is usually most evident around DC where it is convolved with the offset of the signal. The frequency width of windowing distortions can be reduced by increasing the number of samples in each segment. However, this also requires more averaging cycles.
- Low-frequency data points that exhibit windowing artifacts should be removed.
- The Fast Fourier transform is defined at uniformly spaced frequencies, this emphasizes higher frequencies when plotted on a logarithmic scale. When studying the spectra of linear systems, logarithmically spaced frequencies are preferred. To approximate this, a wide bandwidth spectral measurement can be split into a number of one or two decade bands. That is, if a spectral measurement is required from 1 to 10,000 Hz, this can be performed with two recordings, one from 1 to 100 Hz, and another from 100 to 10,000 Hz.
- Typical spectrum analyzers provide a wide range of options for the measurement unit. The units of $\mathrm{V}/\sqrt{\mathrm{Hz}}$ or V^2/Hz are recommended.
- Some spectrum analyzers allow the user to choose between different measurement scales, typically: RMS Voltage (V_{rms}), Peak Voltage (V_{pk}), and Peak-to-Peak Voltage ($V_{\mathrm{p\text{–}p}}$). This scaling factor is only valid for narrow band signals (sine-waves) and is not appropriate for signals of unknown distribution. The RMS Voltage (V_{rms}) should be used for noise measurements.
- When measuring the power spectral density with a dynamic signal analyzer, it is important to note whether the data is a double-sided or single-sided spectrum, and perform a conversion if necessary. See Sect. 13.2.12 for details.

13.7.3 Direct Measurement of Position Noise

Positioning noise and resolution is usually calculated from knowledge of the plant model and system noise sources. Although it is desirable to obtain a direct measurement of positioning noise, this is only practical when an additional sensor is available that exhibits significantly less noise than the feedback sensor.

To explain the difficulty, consider the direct power spectral density measurement of position noise $S_{\mathrm{d}}(f)$ using an additional sensor, this results in

$$S_y(f) = S_{\mathrm{d}}(f) + S_{n_s}(f), \tag{13.91}$$

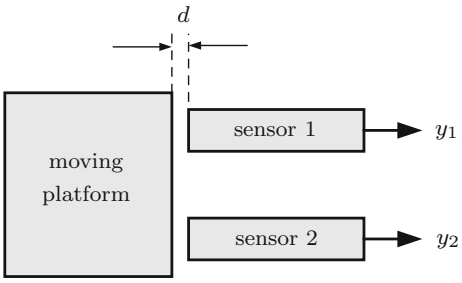


Fig. 13.16 Measurement of position noise d using two sensors to eliminate the influence of sensor noise

where $S_y(f)$ is the measured power spectral density and $S_{n_s}(f)$ is the measurement noise. This equation can be rearranged to reveal $S_{\mathrm{d}}(f)$

$$S_{\mathrm{d}}(f) = S_y(f) - S_{n_s}(f). \tag{13.92}$$

This equation clearly requires the knowledge of $S_{n_s}(f)$ or requires it to be negligible. Although this approach is simple, it is also highly sensitive to uncertainty in $S_{n_s}(f)$, particularly if the positioning and sensor noise are of similar magnitudes or worse.

A better approach is to utilize two identical sensors in the configuration shown in Fig. 13.16. Here, both sensors measure the same displacement but have independent additive noise sources $n_1(t)$ and $n_2(t)$, that is

$$y_1(t) = \mathrm{d}(t) + n_1(t), \tag{13.93}$$

$$y_2(t) = \mathrm{d}(t) + n_2(t). \tag{13.94}$$

The cross-correlation of $y_1(t)$ and $y_2(t)$ is defined by

$$R_{y_1y_2}(\tau) = E\left[y_1(t)y_2(t+\tau)\right] \tag{13.95}$$

which is equal to

$$R_{y_1y_2}(\tau) = E\left[(\mathrm{d}(t) + n_1(t))\,(\mathrm{d}(t+\tau) + n_2(t+\tau))\right] \tag{13.96}$$

$$= E\left[\mathrm{d}(t)\mathrm{d}(t+\tau) + \mathrm{d}(t)n_2(t+\tau) + n_1(t)\mathrm{d}(t+\tau) + n_1(t)n_2(t+\tau)\right]. \tag{13.97}$$

As the displacement noise and sensor noises are generated by different processes, they can be assumed to be independent. We can also make the assumption that each process is stationary, and zero mean as only the varying part is of interest. Under these conditions, the cross-correlation $R_{y_1y_2}(\tau)$ reduces to the autocorrelation of displacement noise $\mathrm{d}(t)$,

$$R_{y_1 y_2}(\tau) = E\left[\mathrm{d}(t)\mathrm{d}(t+\tau)\right] = R_{\mathrm{d}}(\tau). \tag{13.98}$$

As the cross-correlation of y_1 and y_2 is equal to the autocorrelation of d, it follows that the power spectral density of the position noise $S_{\mathrm{d}}(f)$ is equal to the cross power spectral density of y_1 and y_2,

$$S_{y_1 y_2}(f) = S_{\mathrm{d}}(f). \tag{13.99}$$

Thus, by using two independent sensors, the effect of measurement noise can be eliminated. This is a convenient result, as the cross power spectral density or cross-correlation of y_1 and y_2 can be easily acquired with a standard spectrum analyzer.

The actual closed-loop position noise can also be measured using this technique. In this case, three sensors are required, one for the feedback loop, and another two identical sensors for estimating of the cross-correlation or cross power spectral density functions. Further details of this method including expressions for the bias and variance can be found in Fleming (2012).

13.7.4 Measurement of the External Disturbance

To estimate the external disturbances acting on a nanopositioner, i.e., to estimate $S_w(f)$, the actual position noise of the nanopositioner $S_{\mathrm{d}}(f)$ must be measured directly in open loop. This requires the elimination of sensor noise that can be achieved using the technique discussed in the previous section. The amplifier voltage noise must also be eliminated, simply by short-circuiting the actuators. It is also possible to measure the combined contribution of amplifier noise and external disturbance. However, it is preferable to have knowledge of each source individually so that their effect on closed-loop position noise can be known.

13.8 Experimental Demonstration

In this section, an example noise analysis is performed on the piezoelectric tube scanner described in Sect. 13.7.1 whose frequency response is plotted in Fig. 13.17. The goal is to quantify achievable resolution as a function of closed-loop bandwidth.

The voltage amplifier used to drive the tube is a Nanonis HVA4 high-voltage amplifier with a gain of 40. To measure the noise, the input to the amplifier is grounded and the output is amplified by a factor of 1,000 using an SR560 preamplifier as described in Sect. 13.7.1. To remove the signals DC offset, the input of the preamplifier was AC-coupled with a 0.03 Hz cut-off frequency.

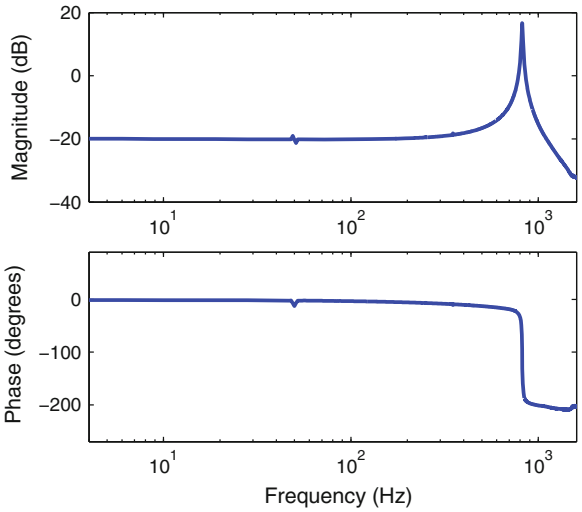


Fig. 13.17 The lateral frequency response of the piezoelectric tube scanner described in Sect. 3.1.1. The response was measured from the applied actuator voltage to the displacement in μm/V

The sensor under consideration is an ADE Tech 4810 Gaging Module with 2804 capacitive sensor. This has a full range of ±100 μm and a sensitivity of 0.1 V/μm. To measure the noise, the sensor is mounted in a aluminum block with a flat-bottomed hole and grub screws to secure the probe and minimize any movement.

The spectral density of each noise source was recorded with an HP 35670A dynamic signal analyzer. Two frequency ranges were used, one from 0 to 12.5 Hz with 400 points to capture low-frequency noise, and another from 0 Hz to 1.6 kHz with 1,600 points. An acceptable measurement variance was achieved with 100 averages for the low-frequency range and 700 averages for the high-frequency range. After exporting the data in $\mathrm{V}/\sqrt{\mathrm{Hz}}$, the two datasets were concatenated in Matlab. Windowing distortions at DC were removed by truncating the first five frequency points of the low-frequency measurement.

The spectral density of the amplifier voltage noise is plotted in Fig. 13.18a. The noise density is approximately 1 $\mu\mathrm{V}/\sqrt{\mathrm{Hz}}$ and has a $1/f$ corner frequency of 3 Hz. Also present are some significant harmonic components, predominantly at the mains frequency of 50 Hz. The resulting open-loop position noise can also be found using Eq. (13.58) and the frequency response plotted in Fig. 13.17. The position noise, shown in Fig. 13.18b, is similar to the voltage noise except for a new peak caused by the system resonance. The constant density is approximately 0.1 $\mathrm{pm}/\sqrt{\mathrm{Hz}}$.

The spectral density of the sensor noise is plotted in Fig. 13.19. Above the $1/f$ corner frequency of 2 Hz, the noise density is 25 $\mathrm{pm}/\sqrt{\mathrm{Hz}}$. This is significantly greater than the position noise due to the voltage amplifier plotted in Fig. 13.18b.

With knowledge of the voltage and sensor noise, the closed-loop positioning noise can now be computed. For the sake of demonstration, an inverse controller similar

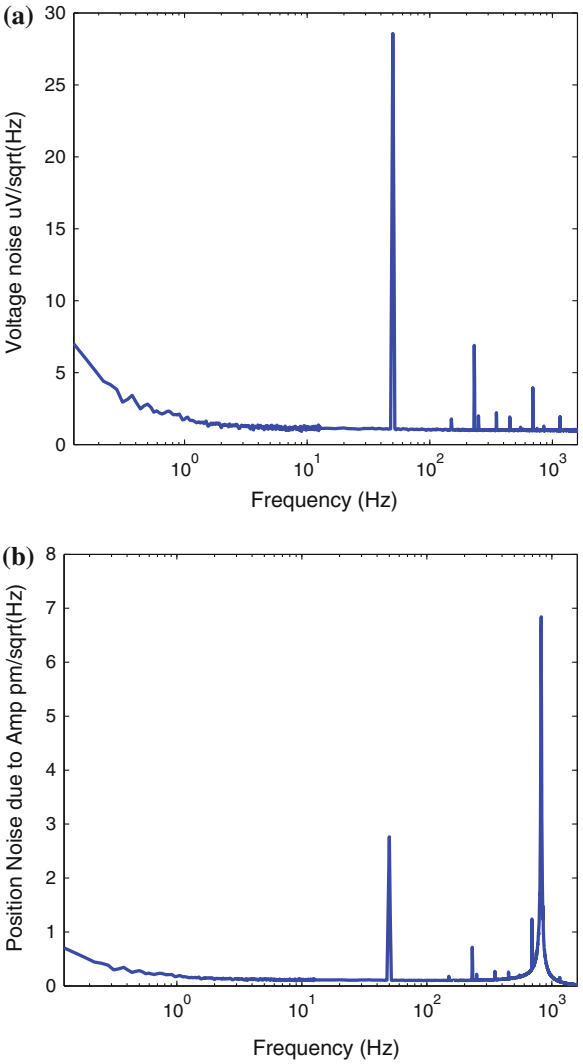


Fig. 13.18 RMS spectral density of the amplifier noise and the resulting displacement. **a** Spectral density of the amplifier voltage noise $\sqrt{S_{V_o}(f)}$ in $\mu\text{V}/\sqrt{\text{Hz}}$. **b** Spectral density of the resulting displacement noise $\sqrt{S_{\text{d}n_s}(f)}$ in $\text{pm}/\sqrt{\text{Hz}}$

to that used in Sect. 13.6.2 will be considered. This is representative of a wide range of model-based controllers. The controller transfer function is,

$$C(s) = \frac{\alpha}{s}\frac{1}{P(s)}, \tag{13.100}$$

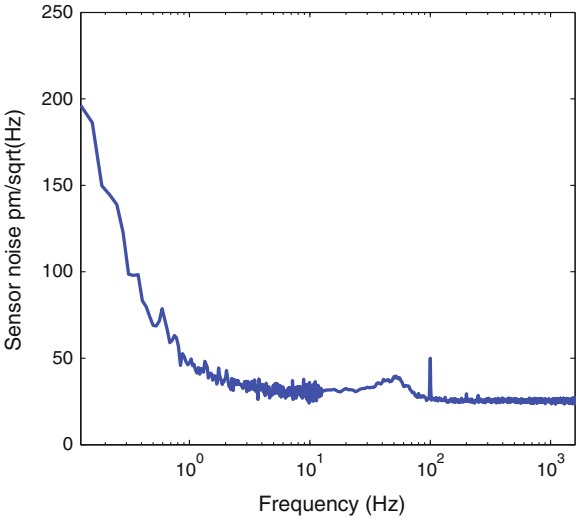
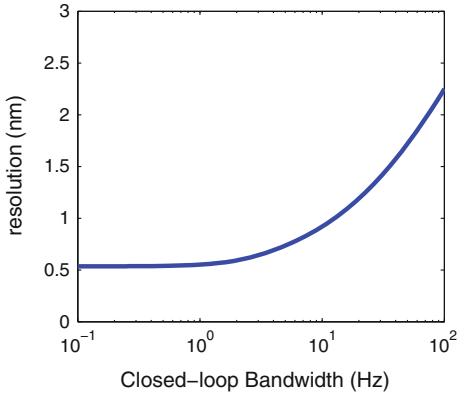


Fig. 13.19 Spectral density of the sensor noise $\sqrt{S_{n_s}(f)}$ in pm/$\sqrt{\text{Hz}}$

Fig. 13.20 6σ Positioning resolution versus closed-loop bandwidth



where $P(s)$ is the nanopositioner response plotted in Fig. 13.17. The sensitivity functions and position noise density due to each source can be computed using Eqs. (13.67) and (13.68). The resolution can then be found from Eqs. (13.64), (13.65) and (13.66).

In Fig. 13.20 the closed-loop positioning resolution is plotted against closed-loop bandwidth, which is equal to $\alpha/2\pi$. The minima of 0.5 nm occurs at 0 Hz which implies that feedforward control will result in the least positioning noise. In closed-loop, the positioning resolution becomes greater than twice the open-loop noise at frequencies greater than 15 Hz. At higher frequencies, the resolution increases proportionally to the square root of closed-loop bandwidth. That is, if the closed-loop bandwidth is doubled the positioning noise increases by a factor of $\sqrt{2}$.

This experiment confirms a well-known observation of scanning probe microscope users: Although large range piezoelectric tubes are suitable for atomic force microscopy, they cannot be used for scanning tunneling microscopy where atomic resolution is required. For such experiments, much smaller piezoelectric tubes are used with a travel range of typically 1 μm. This reduces the effect of amplifier voltage noise leading to an improvement in resolution.

13.9 Time-Domain Noise Measurements

In addition to the frequency domain techniques discussed previously, the position noise can also be estimated directly from time-domain measurements. To estimate the position noise, the measurements of amplifier and sensor noise are filtered by the noise sensitivity functions described in Sect. 13.5.1.

Compared to frequency domain techniques, the time-domain approach has a number of benefits:

- It is simpler;
- It less likely that an error due to units or scaling will occur;
- A spectrum analyzer is not required;
- The distribution histogram can be plotted directly;
- No assumptions about the distribution are required to estimate the peak-to-peak value or 6σ-resolution.

However, there are also a number of disadvantages:

- It is difficult to record signals with 1/f noise due to their high dynamic range.
- To capture both low- and high-frequency noise, long time records are required with high sampling rate;
- There is less insight into the nature of the noise;
- It more difficult to plot the resolution versus bandwidth.

In summary, time-domain noise recordings are simple but lack some of the intuition provided by frequency domain techniques. Time-domain noise measurement techniques are discussed in the following sections, then applied to a nanopositioning system.

13.9.1 Total Integrated Noise

A common method for reporting time-domain noise is known as the *total integrated noise*. The total integrated noise is the RMS value, or standard deviation over a particular measurement bandwidth. For example, a white noise process with a constant spectral density $\sqrt{A}$, has an RMS value σ that is related to the measurement bandwidth f_{bw} by,

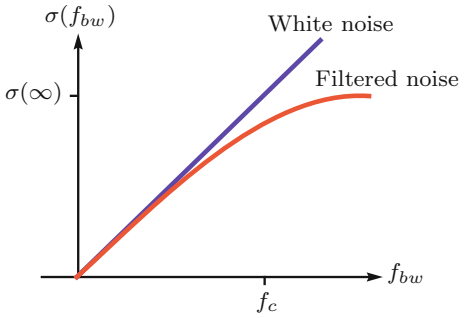


Fig. 13.21 The total integrated noise $\sigma(f_{\text{bw}})$ of a white-noise process with and without a first-order low-pass filter. f_c is the filter cut-off frequency and f_{bw} is the measurement bandwidth. With an unlimited measurement bandwidth, the total integrated noise of the filtered process approaches $\sigma(\infty) = \sqrt{A}\sqrt{1.57 f_c}$

$$\sigma(f_{\text{bw}}) = \sqrt{A}\sqrt{f_{\text{bw}}}, \tag{13.101}$$

where f_{bw} is the measurement bandwidth. In practice, a white noise source is usually filtered by a low-pass system $G(s)$. The total integrated noise is now:

$$\sigma(f_{\text{bw}}) = \sqrt{\int_0^{f_{\text{bw}}} A\,|G(j2\pi f)|^2\ \mathrm{d}f}. \tag{13.102}$$

If $G(s)$ is a first-order filter with cut-off frequency f_c, the total integrated noise is

$$\sigma(f_{\text{bw}}) = \sqrt{A}\sqrt{\int_0^{f_{\text{bw}}} \frac{f_c^2}{f^2 + f_c^2}\ \mathrm{d}f}. \tag{13.103}$$

Using the following integral pair from Poularikas (1999, 45.3.6.1),

$$\int \frac{1}{a + bf^2} = \frac{1}{\sqrt{ab}} \tan^{-1}\left(\frac{f\sqrt{ab}}{a}\right), \tag{13.104}$$

Equation (13.103) reduces to

$$\sigma(f_{\text{bw}}) = \sqrt{A}\sqrt{f_c\,\tan^{-1}(f_{\text{bw}} f_c)}. \tag{13.105}$$

Note that as the measurement bandwidth approaches ∞, $\tan^{-1}(f_{\text{bw}} f_c) \to 1.57$ and $\sigma(f_{\text{bw}})$ approaches the standard expression for the standard deviation of low-pass filtered white noise. The total integrated noise of white noise with, and without a low-pass filter is shown in Fig. 13.21.

The main benefit of total integrated noise is that it can be measured directly using simple instruments. For example, the plot in Fig. 13.21 can be constructed with a variable cut-off low-pass filter and RMS measuring instrument. The filter order should generally be greater than three so that errors resulting from the nonideal response are negligible. Refer to Sect. 13.9.3 for some other guidelines to ensure a statistically valid estimate.

13.9.2 Estimating the Position Noise

In the time-domain, the process of estimating position noise is analogous to the frequency domain techniques discussed in Sect. 13.5.2. Three possible techniques are discussed in the following.

13.9.2.1 Direct Measurement with an Ideal Sensor

The most straightforward and conclusive method for measuring the positioning noise of a nanopositioning system is simply to measure it directly. However, this approach is not often possible as an additional sensor is required with lower noise and a significantly higher bandwidth than the closed-loop system. This problem was discussed in Sect. 13.7.3.

For example, when measuring the position $\mathrm{d}(t)$, the measurement $y(t)$ also contains the sensor noise $n_s(t)$, that is

$$y(t) = \mathrm{d}(t) + n_s(t). \tag{13.106}$$

The RMS value of $y(t)$ is equal to the square-sum of the actual position $\mathrm{d}(t)$ and measurement noise $n_s(t)$, i.e.,

$$\sigma_y = \sqrt{\sigma_{\mathrm{d}}^2 + \sigma_{n_s}^2}, \tag{13.107}$$

where $\sigma_y = \sqrt{E[y^2(t)]}$ is the RMS value of y. Note that the frequency domain analogy can be found in Eq. 13.91.

For an accurate estimate of σ_d, the sensor noise density needs to be comparable or lower than $\mathrm{d}(t)$ over the relevant frequency range (0.1 Hz to $5 f_V$). It is also possible to subtract $\sigma_{n_s}^2$ from the measurement, i.e.,

$$\sigma_d = \sqrt{\sigma_y^2 - \sigma_{n_s}^2}. \tag{13.108}$$

However, this approach is undesirable as it is sensitive to the accuracy of $\sigma_{n_s}^2$ and has the potential to deliver an underestimate of σ_d.

To avoid low-pass filtering and underestimating the noise, the sensor bandwidth must be at least five times greater than the position noise bandwidth. Due to these demanding requirements of the sensor, direct measurement is rarely an option since a suitable sensor may not be available. If such a sensor is available, a major benefit is that position noise will not be underestimated. This provides a high degree of confidence in the measured noise and also resolution.

13.9.2.2 Direct Measurement with Two Noisy Sensors

In Sect. 13.7.3 it was shown that the autocorrelation of positioning noise $\mathrm{d}(t)$ can be found using two, possibly noisy, auxiliary sensors. If the noise of the auxiliary sensors is stationary and uncorrelated, the autocorrelation of the position noise is equal to the cross-correlation of the noisy measurements, that is

$$R_{\mathrm{d}}(\tau) = R_{y_1 y_2}(\tau). \tag{13.109}$$

In many cases, only the RMS value of position noise is required, not the full autocorrelation matrix. Since the RMS value of $\mathrm{d}(t)$ is $\sqrt{R_{\mathrm{d}}(0)}$, the expression can be simplified to

$$\sigma_d = \sqrt{R_{\mathrm{d}}(0)}, \tag{13.110}$$

$$= \sqrt{E[y_1(t) \times y_2(t)]}, \tag{13.111}$$

where $y_1(t)$ and $y_2(t)$ are the noisy measurements of $\mathrm{d}(t)$.

The result in Eq. (13.111) is simply the Root-Mean value of $y_1(t) \times y_2(t)$. In addition to the standard considerations for measuring an RMS value, a longer time recording may be required to provide an acceptable variance for the estimate of σ_d. Although the required number of samples can be calculated analytically, it is much simpler to implement a cumulative computation for σ_d and keep recording until the estimate variance is satisfactory.

13.9.2.3 Prediction Based on Measured Noise

In many cases, it is not possible to measure the position noise directly as auxiliary sensors with suitable performance may not be available or can not be accommodated. In such cases, the position noise can be predicted from measurements of the amplifier and sensor noise. A benefit of this approach is that the the closed-loop noise can be predicted for a number of different bandwidths and controllers, much like frequency domain techniques.

Referring to the feedback diagram in Fig. 13.10, the signals of interest are the amplifier noise V_o, and the sensor noise n_s. As the position noise will be calculated by superposition, the amplifier noise should be measured with the input signal grounded

and the output connected to the nanopositioner. Conversely, the sensor noise should be measured with a dedicated test-rig to avoid the influence of external disturbances. If the sensor noise must be measured in-situ, all of the nanopositioner actuators should be disconnect from their sources and short-circuited.

After the constituent noise sources have been recorded, the position noise can be predicted simply by filtering the noise signals by the sensitivity functions of the control loop. That is, the position noise is:

$$\mathrm{d}(t) = n_s(t)\frac{-C(s)P(s)}{1+C(s)P(s)} + V_o(t)\frac{P(s)}{1+C(s)P(s)}. \tag{13.112}$$

The RMS value of the position noise can now be computed and plotted for a range of different controller-gains and closed-loop bandwidths.

Although the data sizes in time-domain experiments must be necessarily large to guarantee statistical validity, this is not a serious impediment since a range of numerical tools are readily available for extracting the required information.

For example, in Matlab, the RMS value of a vector d can be calculated using `RMS = std(d)` or `RMS=sqrt(mean(y.^2))`. The 6σ-resolution can be found using the function `Res=2*quantile(abs(d),0.997)`. It is also informative to plot the probability density function using `ksdensity` or with the basic histogram function:

```
xi = linspace(-range,range,Ny);
dx = 2*range/Ny;
[y,x] = hist(d,xi);
plot(x,y/(length(d)*dx))
```

where `-range` and `range` encompass the minimum and maximum values of d and `Ny` is the number of x-axis points in the probability density, e.g., 1,000.

13.9.3 Practical Considerations

Many of the considerations for frequency domain noise measurements discussed in Sect. 13.7 are also valid for time-domain measurements. Of particular importance is the need for preamplification and the removal of offset voltages, which were discussed in detail in Sect. 13.7.1.

After a suitable preamplification scheme has been implemented, the position noise can be estimated from recordings of the sensor and amplifier noise. This requires a choice of the recording length and sampling rate.

The length of each recording is defined by the lowest spectral component under consideration. As discussed in Sect. 13.7.1, the lower frequency limit in nanopositioning applications is usually considered to be 0.1 Hz, or less. To obtain a statistically meaningful estimate of the RMS value, a record length of at least ten times the mini-

Table 13.7 Recommended parameters for time-domain noise recordings

Record length	100 s
Amplifier bandwidth	f_V
Antialiasing filter cut-off frequency	$7.5 \times f_V$
Sampling rate	$15 \times f_V$

mum period is required, which implies a minimum recording length of at least 100 s. A longer record length is preferable, but usually not practical.

A more rigorous method for selecting the record length is to calculate the estimator variance as a function of the record length. This relationship was described in Fleming and Moheimani (2003), however, assumptions are required about the autocorrelation or power spectral density. In most cases, the simple rule-of-thumb discussed above will be sufficient.

When selecting the sampling rate, the highest significant frequency that influences position noise should be considered. Since the sensor noise is low-pass filtered by the closed-loop response of the control loop, the highest significant frequency is usually the bandwidth of the voltage amplifier. A good choice of sampling rate is 15 times the amplifier bandwidth. This allows a nonideal antialiasing filter to be utilized with a cut-off frequency of five times the amplifier bandwidth. Since the noise power of a first-order amplifier drops to 3.8 % at five times the bandwidth, this technique captures the majority of noise power. The recommended parameters for time-domain noise recordings are summarized in Table 13.7.

13.9.4 Experimental Demonstration

In this section, the frequency domain noise analysis presented in Sect. 13.8 will be repeated in the time-domain. The same piezoelectric tube nanopositioner, capacitive sensor and high-voltage amplifier will be used.

The bandwidth of the high-voltage amplifier is 2 kHz, so the sampling rate is chosen to be 30 kHz. The preamplifier is also used for antialiasing with a cut-off frequency of 10 kHz as recommended in Table 13.7. With a record length of 100 s, the data will contain 3×10^6 samples.

The distribution and total integrated noise of the voltage amplifier and sensor are plotted in Fig. 13.22. The RMS value of the amplifier noise is 0.14 mV over the 0.1 Hz to 10 kHz measurement bandwidth which corresponds to a predicted 6σ-resolution of 0.84 mV. The measured 6σ-resolution was 0.86 mV which supports the assumption of approximate Gaussian distribution.

The RMS noise and 6σ-resolution of the capacitive sensor was measured to be 3.6 nm and 20 nm, respectively. The capacitive sensor also exhibits an approximately Gaussian distribution, albeit with a slightly greater dispersion than the voltage amplifier.

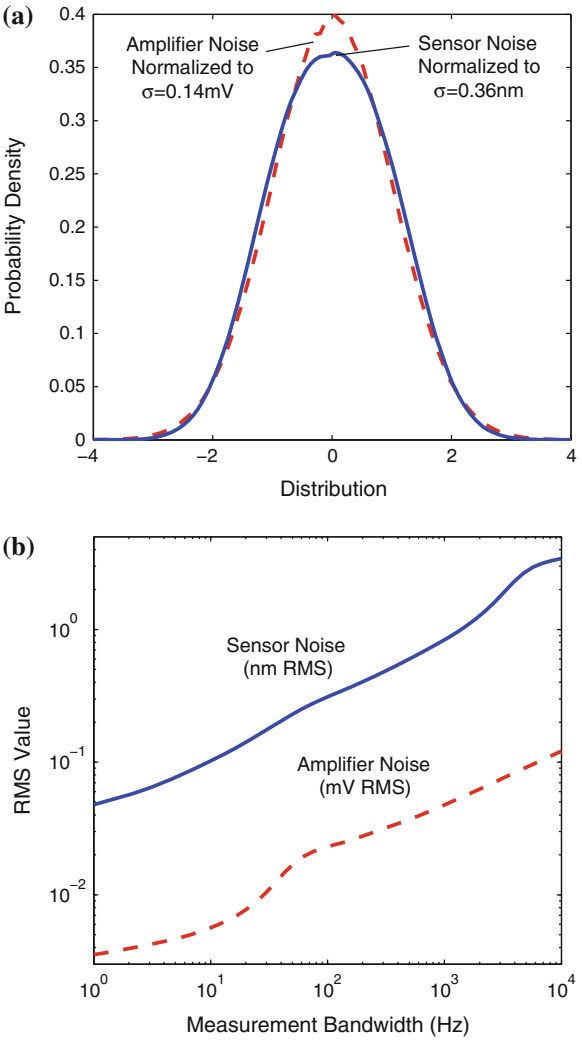


Fig. 13.22 The distribution and total integrated noise of the voltage amplifier and capacitive sensor. Both of the sensors exhibit an approximately Gaussian distribution

For the sake of comparison, the inverse controller discussed in Sect. 13.8 will be used. That is,

$$C(s) = \frac{\alpha}{s}\frac{1}{P(s)}, \tag{13.113}$$

where $P(s)$ is the second-order model of the nanopositioner and α is the closed-loop bandwidth. The position noise can now be simulated using the noise recordings and Eq. (13.112).

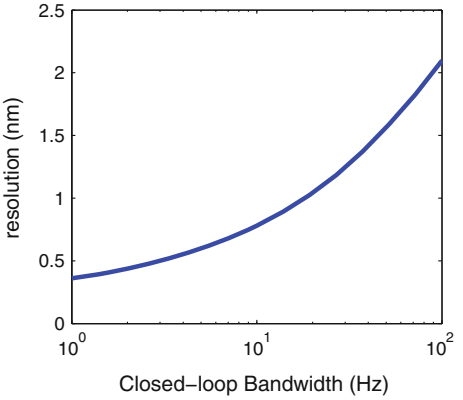


Fig. 13.23 6σ Positioning resolution versus closed-loop bandwidth

Table 13.8 The predicted closed-loop resolution using frequency and time-domain measurements

Bandwidth (Hz)	Frequency domain (nm)	Time-domain (nm)
100	2.2	2.1
10	0.92	0.78
1	0.55	0.36

At low closed-loop bandwidths, the transient response time of the system is significant. For this reason, only the second half of the simulated output is used to calculate the resolution. For the same reason, it is not practical to simulate closed-loop bandwidths of less than approximately 1 Hz. This is an additional disadvantage of time-domain measurements.

The predicted resolution is plotted against closed-loop bandwidth in Fig. 13.23. As expected, this plot closely resembles Fig. 13.20 which was obtained from frequency domain data. The time and frequency domain results are compared below in Table 13.8. With a closed-loop bandwidth of 100 Hz, the predictions are identical, however, at low closed-loop bandwidth, some discrepancy exists. This is due to the long transient response at which tends to underestimate the positioning noise. If necessary, a more accurate result can be achieved by significantly increasing the recording length, however this is not usually desirable or practical.

13.10 A Simple Method for Measuring the Resolution of Nanopositioning Systems

Thus far, a range of time and frequency domain approaches have been discussed for the measurement and prediction of nanopositioner resolution. These techniques can provide a detailed prediction of resolution as a function of closed-loop bandwidth. However, these techniques also require careful measurement practices and involved

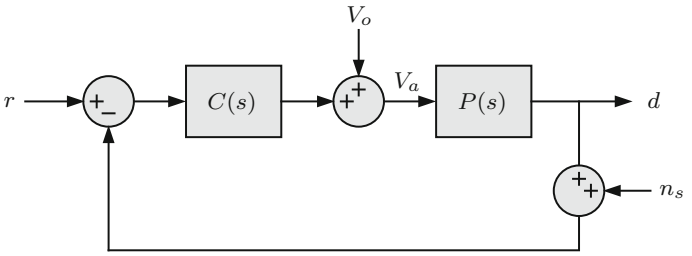


Fig. 13.24 The closed-loop voltage V_a applied to the nanopositioner can be measured to predict the effect of the amplifier and sensor noise, V_o and n_s

processing of the measured data. Specialized equipment and software is also required that may preclude the application of these techniques.

A need exists for a simple procedure to estimate the closed-loop positioning resolution of a nanopositioning system. In the following, the closed-loop output of the high-voltage amplifier is proposed as a suitable measurement signal. This signal, or its spectrum can be filtered by the open-loop response of the plant to reveal the closed-loop positioning resolution.

As shown in Fig. 13.24 the position d is simply the amplifier output voltage V_a filtered by the plant model. This measurement is straightforward and does not require any additional sensors. Either time or frequency domain measurements can be used and will be discussed in the following.

The considerations described in Sects. 13.7 and 13.9.3 are also applicable here. A preamplifier is required with a gain of approximately 1,000 and an AC-coupling frequency of 0.03 Hz or less. A simple protection circuit may also be required to avoid exceeding the voltage range of the preamplifier.

For a time-domain recording, the sampling rate should be greater than fifteen times the amplifier bandwidth and the record length should be 100 s or more. The actual position noise can then be estimated by filtering the recording by a model of the plant. The portion of the simulated displacement that is effected by the transient response should be excised before calculating the RMS value and resolution.

In the case of frequency domain measurement, the spectrum should be split into two or three decades to provide sufficient resolution and range. For example: 0 to 12, 12 to 1.2 kHz, and 1.2 to 12 kHz. The data should preferably be recorded in units of $V/\sqrt{\text{Hz}}$ and have a frequency range of at least five times the amplifier bandwidth. The RMS value and resolution can then be found by evaluating the integral

$$\sigma = \int_0^\infty \sqrt{S_{V_a}(f)}\ |P(2\pi f)|\ \mathrm{d}f \tag{13.114}$$

where $\sqrt{S_{V_a}(f)}$ is the spectral density of V_a.

In the following, the "applied voltage" technique will be used to estimate the resolution of the piezoelectric tube nanopositioner described in Sect. 13.8. To reduce external disturbances, the apparatus is mounted on an isolating table with an acoustic enclosure. A simple analog integral controller is then used to provide a closed-loop bandwidth of 10 Hz. After setting the reference input to zero, the voltage applied to the nanopositioner was preamplified by an SR560 amplifier with a gain of 500 and a AC-coupling frequency of 0.03 Hz. This signal was recorded for 100 s with a sampling rate of 30 kHz.

The noise recording was filtered by the plant model to estimate the closed-loop positioning noise. The distribution is plotted in Fig. 13.25a which has an RMS value of 0.24 nm and a 6σ resolution of 1.4 nm. Since 1.4 nm is greater than 6×0.24 nm, the distribution is slightly wider than a Gaussian distribution. The data can also be used to visualize the expected two-axis performance. In Fig. 13.25b, nine 100 mS sets of data were taken randomly from the estimated position noise and plotted on a constellation diagram with a spacing equal to the prescribed resolution. The 6σ definition of resolution can be observed to be a true prediction of the minimum reasonable spacing between two distinct points.

The measurement of resolution can also be compared with the values predicted by the techniques in Sects. 13.8 and 13.9.4. Note that these simulations must be modified to treat an integral controller rather than an inverse controller. The predicted resolution versus closed-loop bandwidth is plotted in Fig. 13.26. With a bandwidth of 10 Hz, the predicted resolution is 1.5 nm, which closely correlates with the above measurement of 1.4 nm.

13.11 Techniques for Improving Resolution

The obvious methods for improving resolution include reducing the noise density and corner frequency of the amplifier and sensor noise, however, these parameters may be fixed. In Sect. 13.5.3, is was observed that the amplifier bandwidth should not be unnecessarily greater than the closed-loop bandwidth. Since a piezoelectric actuator is primarily capacitive, the bandwidth can be arbitrarily reduced by installing a resistor in series with the load. The resulting first-order cut-off frequency is $f_c = 1/(2\pi RC)$. This simple technique can be used to restrict the bandwidth and avoid unnecessary high-frequency noise which may excite uncontrolled mechanical resonances.

A significant source of positioning noise is the excitation of mechanical resonance due to sensor noise. If the mechanical resonance is lightly damped, it may become the dominant noise contributor. This limitation can be aleviated through the use of model-based (Salapaka et al. 2002; Sebastian and Salapaka 2005) or inverse controllers. However, notch filters and inverse controllers are sensitive to variations in resonance frequency (Leang and Devasia 2007; Abramovitch et al. 2008). Damping control is an alternative technique that provides improved robustness. Suitable damping controllers for nanopositioning applications include polynomial based control (Aphale

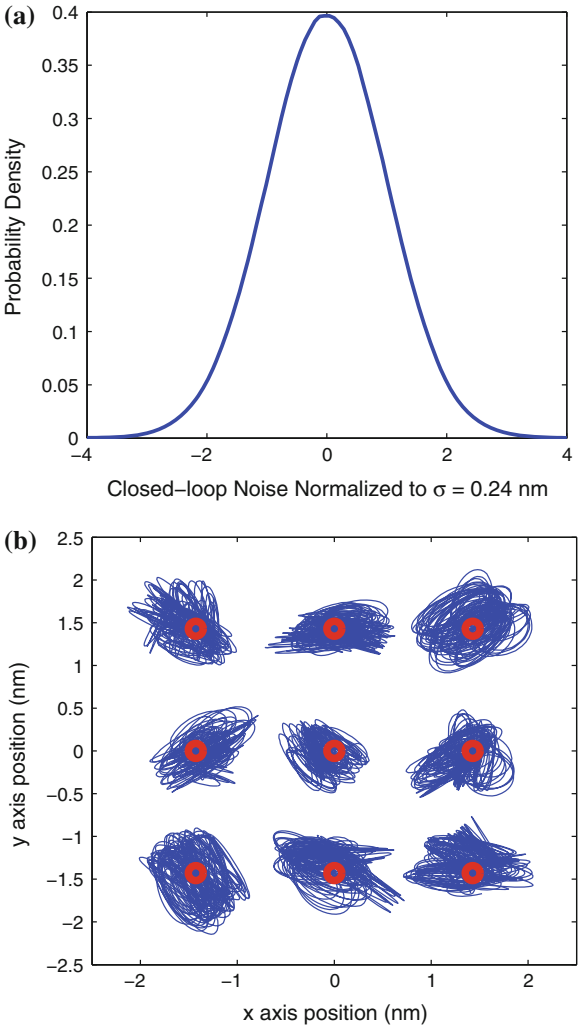


Fig. 13.25 The predicted position noise using the applied voltage technique. **a** Distribution of the predicted noise. **b** Constellation diagram of nine two-dimensional points spaced by the measured 6σ-resolution

et al. 2008), shunt control (Fleming and Moheimani 2006; Fleming et al. 2002), resonant control (Sebastian et al. 2008), Force Feedback (Fleming 2010; Fleming and Leang 2010), and Integral Resonance Control (IRC) (Aphale et al. 2007; Bhikkaji and Moheimani 2008).

The resolution can also be improved by reducing the closed-loop bandwidth, which may be possible if a feedforward controller is used to compensate for the

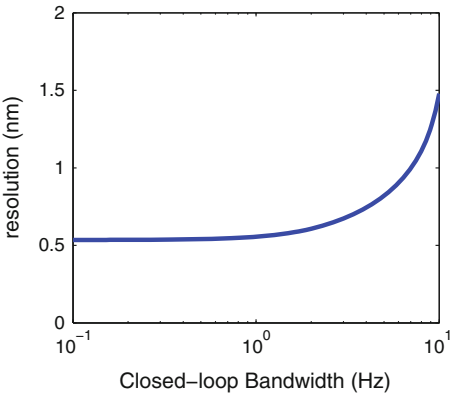


Fig. 13.26 The predicted 6σ resolution of an integral controller versus closed-loop bandwidth. This prediction is based on the frequency domain measurements in Sect. 13.8

reduction of servo bandwidth (Leang et al. 2009; Clayton et al. 2009; Eielsen et al. 2012; Butterworth et al. 2012). The noise sensitivity can also be reduced if the reference trajectory is periodic, which commonly occurs in nanopositioning applications (Kenton and Leang 2012). Periodic trajectories can be effectively controlled using repetitive (Shan and Leang 2012) or iterative controllers (Li and Bechhoefer 2008; Kim and Zou 2008). Both of these techniques provide excellent tracking performance with less noise than a standard control loop with similar tracking error.

Further noise advantages can be achieved if the reference trajectory is also narrowband. For example, AFM scan trajectories can be sprial (Hung 2008; Mahmood et al. 2010) or sinusoidal (Fleming et al. 2010; Tuma et al. 2012; Bazaei et al. 2012). In such cases, the controller bandwidth can be essentially reduced to a single, or a small number of frequenies (Sebastian et al. 2008).

Multiple sensors can also be used collaboratively to provide both high resolution and wide bandwidth. For example, a low-noise piezoelectric sensor can be used for active resonance damping while a capacitive sensor is used for low-frequency tracking (Yong et al. 2013; Fleming 2010). Magnetoresistive sensors have also shown promise for low-noise high-bandwidth position sensing (Sahoo et al. 2011; Kartik et al. 2012). Multiple sensors can be combined by complementary filters (Fleming 2010) or by an optimal technique in the time (Fleming et al. 2008) or frequency domain (Sebastian and Pantazi 2012).

13.12 Chapter Summary

Resolution is a key performance specification of nanopositioning systems. In this chapter, resolution is defined as the maximum peak-to-peak position variation or the minimum distance between two distinct locations. As the position variation is predominantly the sum of multiple random processes, the peak-to-peak variation is defined as the 99.7 % probability that a single observation will lie within thesebounds.

If the contributing noise sources are Gaussian random processes, the peak-to-peak variation is equal to six times the standard deviation or RMS value (6σ).

The foremost noise sources in a nanopositioning system were identified as the amplifier voltage noise and the displacement sensor noise. The simulation examples demonstrate that the minimum position noise usually occurs in open-loop or with very low closed-loop bandwidth. This implies that combined feedback and feedforward control can achieve the best positioning resolution. Such techniques are discussed in Chap. 9.

Both frequency and time-domain techniques were described for measuring and predicting the closed-loop resolution of a nanopositioning system. Although frequency domain techniques provide a more intuitive understanding of the noise sources, time-domain recordings may be easier to perform. For practical application, both techniques require careful experimental planning. A number of guidelines were discussed to ensure the procurement of statistically valid estimates.

Although the above techniques can predict resolution as a function of closed-loop bandwidth, this process may be too involved for some applications in both academia and industry. To meet the need for a straightforward process, it was demonstrated that the voltage applied to a nanopositioner can be recorded and used to predict the closed-loop resolution. The "applied voltage" technique requires only one recording and one filtering operation to predict the closed-loop resolution. Experimental results demonstrate an excellent correlation between the applied voltage technique and other methods.

References

Abramovitch D, Franklin G (2002) A brief history of disk drive control. IEEE Control Syst 22(3): 28–42

Abramovitch DY, Hoen S, Workman R (2008) Semi-automatic tuning of PID gains for atomic force microscopes. In: American control conference, WA, June, Seattle, pp 2684–2689

Al Mamun A, Ge SS (2005) Precision control of hard disk drives. IEEE Control Syst 25(4):14–19

Aphale SS, Bhikkaji B, Moheimani SOR (2008) Minimizing scanning errors in piezoelectric stack-actuated nanopositioning platforms. IEEE Trans Nanotechnol 7(1):79–90

Aphale SS, Fleming AJ, Moheimani SOR (2007) Integral control of resonant systems with collocated sensor-actuator pairs. IOP Smart Mater Struct 16:439–446

Bazaei A, Yong YK, Moheimani SOR (2012) High-speed lissajous-scan atomic force microscopy: Scan pattern planning and control design issues. Rev Sci Instrum 83(6):063701

Bhikkaji B, Moheimani SOR (2008) Integral resonant control of a piezoelectric tube actuator for fast nano-scale positioning. IEEE Trans Mechatron 13(5):530–537

Brown RG, Hwang PYC (1997) Introduction to random signals and applied kalman filtering. Wiley, New York

Butterworth JA, Pao LY, Abramovitch DY (2012) Analysis and comparison of three discrete-time feedforward model-inverse control techniques for nonminimum-phase systems. Mechatronics 22(5):577–587

Clayton GM, Tien S, Leang KK, Zou Q, Devasia S (2009) A review of feedforward control approaches in nanopositioning for high-speed SPM. J Dynamic Syst Meas Control 131:061 101 (1–19)

Eielsen AA, Gravdahl JT, Pettersen KY (2012) Adaptive feed-forward hysteresis compensation for piezoelectric actuators. Rev Sci Instrum 83(8):085001
Einstein A (1914) Mthode pour la dtermination de valeurs statistiques d'observations concernant des grandeurs soumises des fluctuations irrgulires. Archives des Sciences Physiques et Naturelles 37(4):254–256
Einstein A (1987) Method for the determination of statistical values of observations regarding quantities subject to irregular fluctuations. IEEE ASSP Mag 4(4):6–6
van Etten WC (2005) Introduction to noise and random processes. Wiley, West Sussex
Fleming AJ, Behrens S, Moheimani SOR (2002) Optimization and implementation of multi-mode piezoelectric shunt damping systems. IEEE/ASME Trans Mechatron 7(1):87–94
Fleming AJ (2010) Nanopositioning system with force feedback for high-performance tracking and vibration control. IEEE Trans Mechatron 15(3):433–447
Fleming AJ, Kenton BJ, Leang KK (August 2010) Bridging the gap between conventional and video-speed scanning probe microscopes. Ultramicroscopy 110(9):1205–1214
Fleming AJ, Leang KK (2010) Integrated strain and force feedback for high performance control of piezoelectric actuators. Sens Actuators A 161(1–2):256–265
Fleming AJ, Moheimani SOR (2003) Adaptive piezoelectric shunt damping. IOP Smart Mater Struct 12(1):18–28
Fleming AJ, Moheimani SOR (2006) Sensorless vibration suppression and scan compensation for piezoelectric tube nanopositioners. IEEE Trans Control Syst Technol 14(1):33–44
Fleming AJ, Wills AG, Moheimani SOR (November 2008) Sensor fusion for improved control of piezoelectric tube scanners. IEEE Trans Control Syst Technol 15(6):1265–6536
Fleming AJ (2012) A method for measuring the resolution of nanopositioning systems. Rev Sci Instrum 83(8):086101
Hicks TR, Atherton PD, Xu Y, McConnell M (1997) The nanopositioning book. Queensgate Intstruments Ltd., Berkshire
Horowitz P, Hill W (1989) The art of electronics. Cambridge University Press, Cambridge
Hung S-K (2008) Spiral scanning method for atomic force microscopy. In: Proceedings on tip based nanofabrication workshop, Taipei, Taiwan, October 2008, pp 10(1–10)
ISO (1994) ISO 5725—accuracy (trueness and precision) of measurement methods and results
Kartik V, Sebastian A, Tuma T, Pantazi A, Pozidis H, Sahoo DR (2012) High-bandwidth nanopositioner with magnetoresistance based position sensing. Mechatronics 22(3):295–301
Kenton BJ, Leang KK (2012) Design and control of a three-axis serial-kinematic high-bandwidth nanopositioner. IEEE/ASME Trans Mechatron 17(2):356–369
Kim K, Zou Q (2008) Model-less inversion-based iterative control for output tracking: piezo actuator example. In: American control conference, WA, June, Seattle, pp 2710–2715
Leang KK, Devasia S (2007) Feedback-linearized inverse feedforward for creep, hysteresis, and vibration compensation in afm piezoactuators. IEEE Trans Control Syst Technol 15(5):927–935
Leang KK, Zou Q, Devasia S (2009) Feedforward control of piezoactuators in atomic force microscope systems. Control Syst Mag 29(1):70–82
Li Y, Bechhoefer J (2008) Feedforward control of a piezoelectric flexure stage for AFM. In: American control conference. WA, June, Seattle, pp 2703–2709
Mahmood IA, Moheimani SOR, Bhikkaji B (2011) A new scanning method for fast atomic force microscopy. IEEE Trans Nanotechnol 10(2):203–216
Poularikas AD (1999) The handbook of formulas and tables for signal processing. CRC Press, Boca Raton
Proakis JG, Manolakis DG (1996) Digital signal processing. Principles, algorithms, and applications, 3rd edn. Prentice Hall Inc, Upper Saddle River
Sahoo DR, Sebastian A, Häberle W, Pozidis H, Eleftheriou E (2011) Scanning probe microscopy based on magnetoresistive sensing. Nanotechnology 22(14):145501
Salapaka S, Sebastian A, Cleveland JP, Salapaka MV (2002) High bandwidth nano-positioner: a robust control approach. Rev Sci Instrum 75(9):3232–3241

Sebastian A, Pantazi A (March 2012) Nanopositioning with multiple sensors: a case study in data storage. IEEE Trans Control Syst Technol 20(2):382–394

Sebastian A, Pantazi A, Moheimani SOR, Pozidis H, Eleftheriou E (2008) A self servo writing scheme for a MEMS storage device with sub-nanometer precision. In: Proceedings on IFAC World Congress, Seoul, Korea, July, pp 9241–9247

Sebastian A, Pantazi A, Pozidis H, Elefthriou E (2008) Nanopositioning for probe-based data storage. IEEE Control Syst Mag 28(4):26–35

Sebastian A, Salapaka S (2005) Design methodologies for robust nano-positioning. IEEE Trans Control Syst Technol 13(6):868–876

Shan Y, Leang KK (2012) Accounting for hysteresis in repetitive control design: nanopositioning example. Automatica 48(8):1751–1758

Tuma T, Lygeros J, Kartik V, Sebastian A, Pantazi A (2012) High-speed multiresolution scanning probe microscopy based on lissajous scan trajectories. Nanotechnology 23:185501

Wulff C (2006) On spectral densities

Yong YK, Fleming AJ, Moheimani SOR (2013) A novel piezoelectric strain sensor for simultaneous damping and tracking control of a high-speed nanopositioner. IEEE/ASME Trans Mechatron 18(3):1113–1121

Chapter 14
Electrical Considerations

Due to their high stiffness, small dimensions and low mass, piezoelectric stack actuators are capable of developing large displacements with bandwidths of greater than 100 kHz. However, due to their large electrical capacitance, the associated driving amplifier is usually limited in bandwidth to a few kHz.

In this chapter, the limiting characteristics of piezoelectric drives are discussed. These are found to be signal bandwidth, output impedance, cable inductance, and power dissipation. For applications that require extremely high speed, the *dual-amplifier* (Fleming 2008) is introduced that exhibits a bandwidth of 2 MHz with a 100-nF capacitive load. Experiments demonstrate a 20-V 300-kHz sine wave faithfully reproduced across a 100 nF load with negligible phase delay and a peak-to-peak current of 3.8 A. Although the peak output voltage and current is 200 V and 1.9 A, the worst-case power dissipation is only 30 W.

14.1 Introduction

Piezoelectric transducers are the actuator of choice in applications requiring precision motion and force control. They are compact, light-weight, and high in stiffness. These properties permit high mechanical resonance frequencies, typically in the tens or hundreds of kilohertz.

Many applications utilize the high speed and precision offered by piezoelectric actuators. Examples include nanofabrication systems (Rubio-Sierra et al. 2005; Tseng et al. 2008, 2005), high-speed micro-mechanical systems (Uchino and Giniewica 2003), Scanning Probe Microscopes (SPMs) (Bonnell 2001), and vibration control systems (Inman 2006; Moheimani and Fleming 2006; Preumont 2006).

Both scanning probe microscopes and nanofabrication systems use piezoelectric tube or stack based scanners for sample positioning and probe control. With increasing SPM imaging speed and nanofabrication throughput, greater demands are placed on the bandwidth of the positioning stages (Zou et al. 2004). These demands have

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2_14,

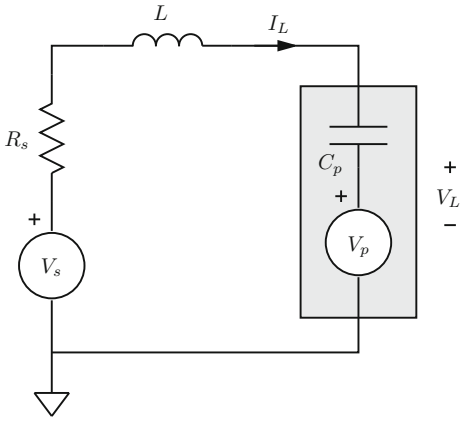


Fig. 14.1 A voltage source V_s driving a piezoelectric load. The actuator is modeled by a capacitance C_p and strain-dependent voltage source V_p. The resistance R_s and inductance L are the output impedance and cable inductance, respectively

necessitated the use of small, high capacitance, multilayer actuators to achieve the required stiffness and resonance frequency (Rost et al. 2005; Schitter et al. 2007).

Unfortunately, due to the high capacitance of stacked and multilayer actuators, in practice, system bandwidth is usually dictated by driving electronics. The first contribution of this paper is to identify the limitations of piezoelectric drive electronics. The foremost limitations established in the following section are signal bandwidth, output impedance, cable inductance, and power dissipation.

To circumvent the limitations identified, a new amplifier is described in Sect. 14.3. The *dual-amplifier* comprises a standard high-voltage amplifier and secondary low-voltage amplifier that increases performance at high frequencies. Experimental results in Sect. 14.4 demonstrate a bandwidth of 2 MHz with a 100-nF load capacitance.

14.2 Bandwidth Limitations

14.2.1 Passive Bandwidth Limitations

Two major causes of bandwidth limitation are the amplifiers output impedance and the inductance of cables and connectors. Consider the electrical circuit shown in Fig. 14.1, where a voltage source is connected to a piezoelectric actuator. The actuator is modeled as a capacitance C_p is series with a strain-dependent voltage source V_p. The resistance R_s and inductance L are the source impedance and cable inductance respectively.

The cable inductance per meter L_m can be calculated from the characteristic impedance Z_0 and capacitance per meter C_m using the equation (Horowitz and Hill 1989):

$$L_m = Z_0^2 C_m. \tag{14.1}$$

Table 14.1 Bandwidth limitation imposed by source impedance (**a**) and cable inductance (**b**)

(a) Bandwidth due to R_s				(b) Resonance frequency due to L			
	C_p				C_p		
R_s	100 nF	1 uF	10 uF	L	100 nF	1 uF	10 uF
1 Ω	1.6 MHz	160 KHz	16 kHz	25 nH	3.2 MHz	1 MHz	320 kHz
10 Ω	160 KHz	16 kHz	1.6 kHz	250 nH	1 MHz	320 kHz	100 kHz
10 Ω	16 kHz	1.6 kHz	160 Hz	2500 nH	320 kHz	100 kHz	32 kHz

The inductance of standard RG-58 coaxial cable is 250 nH/m, this is lower than typical speaker cable which has an inductance of around 600 nH/m. Both are commonly used as interconnects between amplifiers and actuators.

The amplifier source impedance refers to the high-frequency output impedance of the amplifier. In commercially available amplifiers, R_s is typically between 10 and 100 Ω. When considering the effects of both output impedance and cable inductance, the transfer function from source voltage V_s to load voltage V_L is:

$$\frac{V_L(s)}{V_s(s)} = \frac{\frac{1}{LC_p}}{s^2 + \frac{R_s}{L}s + \frac{1}{LC_p}}. \tag{14.2}$$

This is a unity-gain second-order resonant low-pass filter with resonance frequency f_r and damping ratio ξ defined by:

$$f_r = \frac{1}{2\pi\sqrt{LC_p}}, \quad \xi = \frac{R_s\sqrt{LC_p}}{2L}. \tag{14.3}$$

If inductance is neglected, the first-order cut-off frequency resulting from the source resistance is

$$f_c = \frac{1}{2\pi R_s C_p}. \tag{14.4}$$

In Table 14.1, the first-order cut-off frequency and resonance frequency is tabulated for a range of typical values for R_s, C_p and L. Clearly, the output impedance is of primary concern. This is because the bandwidth imposed by source impedance is inversely proportional to both resistance and capacitance. Reductions in both of these parameters can achieve significant bandwidth improvements. Alternatively, the resonance frequency is inversely proportional to both $\sqrt{L}$ and $\sqrt{C_p}$, so a fourfold reduction in L or C_p is required to double the bandwidth.

Although it is difficult to achieve improvements in resonance frequency, all effort should be expended in doing so, as a lightly damped resonance in the transfer function is highly undesirable. In addition to oscillations induced by wide band input signals, the gain peaking and phase-lag can severely limit the performance of feedback control systems in which the amplifier and actuator are enclosed.

14.2.2 Amplifier Bandwidth

The most obvious bandwidth limitation is the small-signal bandwidth of the amplifier. In commercial devices, this can range from 1 kHz to 100 kHz. Unfortunately, these figures are load dependent. The highly capacitive impedance and resonant nature of piezoelectric loads introduces phase-lag into the feedback path. This reduces bandwidth, decreases phase margin, and can lead to instability. For standard voltage-feedback amplifiers with dominant pole compensation, a rough rule of thumb is that closed-loop bandwidth cannot exceed one-tenth the cut-off frequency of the pole formed by the amplifiers output impedance R_s and load capacitance C_p. Typical frequencies for this pole are shown in Table 14.1a. Thus, with standard voltage-feedback amplifiers, the dominant limitation is the output pole. To improve performance, this pole will either have to be increased in frequency, or removed from the closed-loop transfer function of the amplifier, or both.

Further bandwidth restrictions are imposed by the maximum slew rate of the amplifier. This is the maximum rate at which the output voltage can change and is usually expressed in Volts per microsecond V/μs. For sinusoidal signals, the amplifiers slew rate must exceed

$$SR_{\sin} = V_{p\text{-}p}\pi f, \tag{14.5}$$

where $V_{p\text{-}p}$ is the peak-to-peak voltage and f is the frequency. Triangular signals, used in scanning systems require a lesser slew rate of

$$SR_{\text{tri}} = V_{p\text{-}p}2f. \tag{14.6}$$

If a 300-kHz sine wave is to be reproduced with an amplitude of 10 V, the required slew rate is 20 V/μs. This value is proportional to both frequency and amplitude. Although slew rate limitations can be critical when considering resistive or inductive loads; when dealing with capacitive loads, the current limit is usually exceeded well before the slew rate limit. This is discussed in the following section.

14.2.3 Current and Power Limitations

Neglecting the piezoelectric strain voltage, i.e., when driving the actuator off-resonance, the current delivered to a piezoelectric actuator is approximately

$$I_L(s) = V_L(s)C_p s, \tag{14.7}$$

or in the time domain

$$I_L(t) = C_p \frac{\mathrm{d}\, V_L(t)}{\mathrm{d}t}. \tag{14.8}$$

For sinusoidal signals, the maximum positive and negative current is equal to

Table 14.2 Minimum current requirements for a 10 V sinusoid

	C_p		
f	100 nF	1 uF	10 uF
30 Hz	0.19 mA	1.9 mA	19 mA
3 kHz	19 mA	190 mA	1.9 A
300 kHz	1.9 A	19 A	190 A

$$I_L^{\max} = V_{p-p}\pi f C_p. \tag{14.9}$$

For triangular signals, the maximum current is

$$I_L^{\max} = V_{p-p}2 f C_p. \tag{14.10}$$

Examples of current requirements for different load capacitances and frequencies are shown in Table 14.2.

When selecting the required current limit of an amplifier, the key parameter is the maximum allowable power dissipation. The power dissipation in the output stage of a linear amplifier is equal to the product of current through, and voltage across the power transistors. That is, the power dissipation P_d is

$$P_d = I_L(V_{\text{rail}} - V_L), \tag{14.11}$$

where V_{rail} the is the internal power supply voltage.

In an amplifier designed to tolerate output short circuits, i.e., $V_L = 0$, the maximum power dissipation typically sets the limit on maximum allowable output current. Fold-back current limiting can increase maximum output current but is not suitable for capacitive loads where maximum current can be required with low or no output voltage.

14.3 Dual-Amplifier

14.3.1 Circuit Operation

High-voltage amplifiers cannot achieve bandwidths over 100 kHz with large capacitive loads for two main reasons. First, the output impedance of high-voltage output stages is typically around 10 Ω,[1] this limits the small-signal bandwidth to the values shown in Table 14.1a with $R_s = 10\,\Omega$. Second, the power dissipation, even for output

[1] Due to second-breakdown in bipolar transistors (Horowitz and Hill 1989), MOSFET transistors are the only option in high-voltage circuits where power dissipation of greater than a few Watts is required. The best output impedance at high frequencies is obtained by a complementary class AB MOSFET push–pull stage. Although Class A stages achieve the lowest output impedance, they are impractical due to their low efficiency and high quiescent dissipation (Horowitz and Hill 1989).

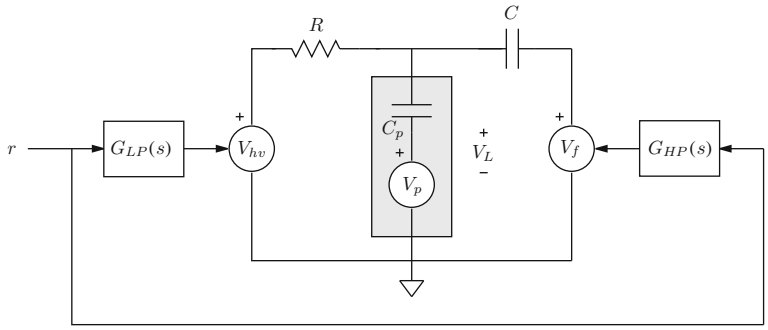


Fig. 14.2 Schematic diagram of a high-speed dual-amplifier. The reference input r is applied simultaneously to a slow high-voltage amplifier V_{hv}, and a fast low-voltage amplifier V_f. The two amplifiers are coupled to the load through the resistance R and capacitance C

voltages as low as 10 V, is in the hundreds of Watts at frequencies above 100 kHz. For these reasons, high-voltage amplifiers driving capacitive loads of 100 nF or greater are usually restricted in bandwidth to around 10 kHz.

In the following, a *dual-amplifier* amplifier is described that alleviates the problems associated with high-voltage amplifiers by adding an auxiliary low-voltage stage to improve output impedance and drop power dissipation at high frequencies (Fleming 2008). A schematic diagram of the dual-amplifier is shown in Fig. 14.2. Essentially the dual-amplifier comprises two amplifiers, a standard high-voltage amplifier V_{hv} and a fast low-voltage amplifier V_f. The low-voltage stage drives the actuator at high frequencies but with reduced range. Due the lower supply voltage (say ± 15 V) the low-voltage stage dissipates less power and can use bipolar transistors under heavy forward bias to provide an output impedance in the milliOhm range.

The two amplifiers are coupled to the load through the resistor R and capacitor C. This network ensures that the load voltage V_L receives low-frequency power from V_{hv} and high-frequency power from V_f. As a function of V_{hv} and V_f, the load voltage is equal to

$$V_L(s) = \frac{\alpha}{s+\alpha} V_{hv}(s) + \frac{C}{C+C_p} \frac{s}{s+\alpha} V_f(s) \qquad \text{where, } \alpha = \frac{1}{R(C+C_p)}. \tag{14.12}$$

That is, the response from the high-voltage side is a low-pass filter, while the response from the low-voltage side is a high-pass filter with attenuation. A key observation is that the filters $\frac{\alpha}{s+\alpha}$ and $\frac{s}{s+\alpha}$ are complementary, i.e., if the attenuation due to C and C_p is accounted for, a signal applied to both amplifiers will be perfectly reproduced across the load. Low-frequency power is supplied by V_{hv} while high-frequency power is supplied by V_f, which is exactly the situation desired.

A design issue for the circuit shown in Fig. 14.2 is the choice of R and C. Choosing C is straight forward. C should be chosen so that the ratio $\frac{C}{C+C_p}$ is close to unity

and that the combination of C and C_p does not unnecessarily load the high-voltage stage. A reasonable compromise is $C = 10C_p$. With C fixed, the selection of R controls the low and high-pass cutoff frequencies. A good choice is to design the cut-off frequency and R so that the high-voltage amplifier can be fully utilized, that is, so that at the cut-off frequency, the high-voltage amplifier is about to reach current limit. If the peak-to-peak voltage and current from the high-voltage stage is V_{pp} and I_{pp}, the corresponding impedance Z is

$$\frac{V_{pp}}{I_{pp}} = Z = \left| R + \frac{-j}{2\pi f(C + C_p)} \right| \tag{14.13}$$

$$\frac{V_{pp}}{I_{pp}} = \sqrt{R^2 + \frac{1}{4\pi^2 f^2 (C + C_p)^2}} \tag{14.14}$$

The frequency where this occurs can be set to the filter cutoff frequency by substituting $f = \frac{\alpha}{2\pi}$. Simplification yields

$$\frac{V_{pp}}{I_{pp}} = \sqrt{2R^2}. \tag{14.15}$$

Hence, R and the cutoff frequency in Hertz F_c are

$$R = \sqrt{\frac{1}{2}\left(\frac{V_{pp}}{I_{pp}}\right)^2} \qquad F_c = \frac{1}{2\pi R(C + C_p)}. \tag{14.16}$$

14.3.2 Range Considerations

Although the low-voltage stage significantly improves high-frequency performance, it is important to note that the penalty is reduced range at high frequencies. However, in many applications this does not present a significant drawback as there is no requirement to drive the actuator at full range above 100 kHz. Indeed, the majority of piezoelectric stack actuators would be destroyed by inertial forces and dielectric heating.

The full voltage range of the amplifier can only be realized within the bandwidth of the high-voltage stage, i.e., from DC to $\frac{\alpha}{2\pi}$ Hz. More precisely, the full voltage range is reduced by $\frac{1}{\sqrt{2}}$ at $\frac{\alpha}{2\pi}$ Hz.

To avoid saturation of the low-voltage stage in the frequency band where full range is available, the additional first-order complementary filters $G_{HP}(s)$ and $G_{LP}(s)$ shown in Fig. 14.2 are required. While $G_{HP}(s)$ removes low-frequency signal content to avoid saturation of the low-voltage stage, $G_{LP}(s)$ ensures that both stages remain complementary.

The cut-off frequency is determined by the difference in range between the high- and low-voltage stages. $G_{HP}(s)$ should be high enough in frequency to ensure that the range of the low-voltage stage is not exceeded at $\frac{\alpha}{2\pi}$ Hz. For example, if the low-voltage stage has one-tenth the range of the high-voltage stage, the cut-off frequency should be ten times greater than the RC cutoff in (14.12), i.e.,

$$G_{HP}(s) = \frac{s}{s+\beta} \qquad \text{where } \beta = 10\alpha. \tag{14.17}$$

Once $G_{HP}(s)$ is decided, the filter $G_{LP}(s)$ is calculated to maintain complimentary signal paths through the low- and high-voltage stages. It is easily verified that this condition is satisfied when

$$G_{LP}(s) = \frac{s+\alpha}{a}\left(1 - \frac{s}{s+\beta}\frac{s}{s+\alpha}\right) = \frac{(\beta+\alpha)}{\alpha}\frac{s+\frac{\beta\alpha}{\beta+\alpha}}{s+\beta} \tag{14.18}$$

The output range versus frequency of the dual-amplifier can be estimated from the dominant poles in the system and the range of the low- and high-voltage stages, i.e.,

$$\text{Range}(\omega) = \left|\frac{\alpha}{s+\alpha}R_{HV} + \frac{s}{s+\beta}R_{LV}\right|_{s=j\omega} \tag{14.19}$$

where R_{HV} and R_{LV} are the full output voltage ranges of the high- and low-voltage stages.

14.4 Electrical Design

In this section, the implementation of a dual-amplifier is described. The goal is to drive a 100-nF load with a full voltage range of 0–200 V, and a high-frequency range of 20-V peak-to-peak. The completed device is pictured in Fig. 14.3.

14.4.1 High-Voltage Stage

The high-voltage stage is a basic voltage-feedback amplifier constructed from an Apex Microtechnology PA98 450-V op amp. This is supplied by an International Power IHB200-0.12 215-V power supply with a current rating of 120 mA. With sufficient storage capacitance, the 120 mA output current is sufficient to supply the PA98 with its required 20 mA quiescent current and allow for an amplifier current limit of ±200 mA so long as the load is capacitive. The worst-case power dissipation into a capacitive load is 20 W, safely within the rated limit of 30 W.

Fig. 14.3 Amplifier enclosure with connected cable and 100 nF capacitive load

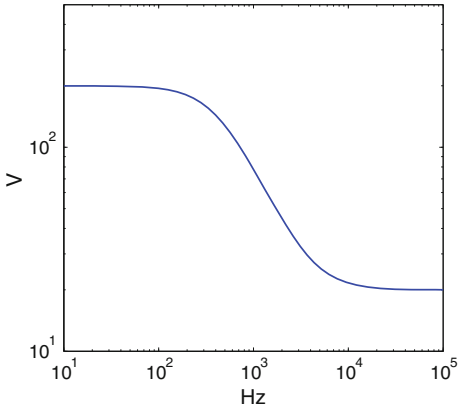


Fig. 14.4 The peak-to-peak output voltage range of the dual-amplifier

As discussed in Sect. 14.3.1, a load capacitance of 100 nF requires a filter capacitance C of approximately 1 μF. R can be calculated from Eq. (14.16) as 353 Ω. The resulting filter cutoff frequency is 409 Hz.

To avoid saturation of the low-voltage stage, which has only one-tenth the high-voltage range, the filter cutoff frequency of $G_{HP}(s)$ should be approximately 5 kHz, i.e., $\beta = 2\pi 5000$. As discussed in Sect. 14.3.2, the resulting range versus frequency is plotted in Fig. 14.4.

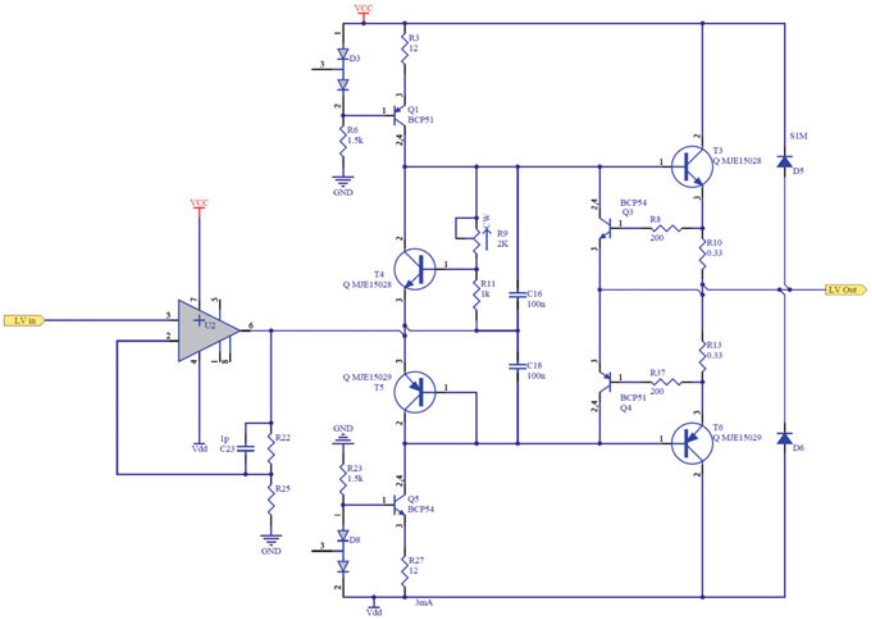


Fig. 14.5 Schematic of the low-voltage amplifier stage. This stage comprises an op amp amplifier to achieve the correct signal gain followed by a bipolar class AB output

14.4.2 Low-Voltage Stage

The necessary amplifier performance requirements for achieving high bandwidth with capacitive loads were identified in Sects. 14.2.1 and 14.2.2. With a 100-nF load, a 1-MHz bandwidth requires an output impedance of less than 1 Ω and a small-signal bandwidth well in excess of 1 MHz.

With high-voltage supplies and MOSFET output transistors, low-output impedance is extremely difficult to achieve. However, due to the reduced supply voltage, high-performance bipolar transistors can be utilized. These transistors can be operated under heavy forward bias to drop the output impedance to well below 1 Ω.

The schematic of a wide-bandwidth low-voltage amplifier is shown in Fig. 14.5. This circuit utilizes a high-speed op amp to achieve the gain of 21 and bandwidth in excess of 10 MHz. The following biasing stage provides drive for the output transistors and protects the op amp from excessive current load. The output transistors are high-performance audio types and provide excellent linearity in open-loop. With sufficient bias, the output stage is sufficiently linear to negate the need for output voltage feedback. This is a major advantage as the gain stage is not limited in bandwidth by the pole formed by the output impedance and load capacitance.

A further advantage of the low-voltage stage is that it can provide large output currents at high frequency without major power dissipation. For example, a 300-kHz

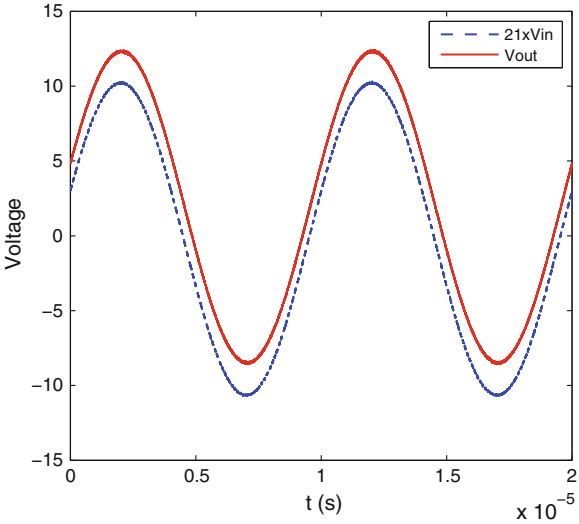


Fig. 14.6 Measured amplifier input and output with a 300-kHz sine wave reference and 100-nF load. The peak-to-peak current is 3.8 A. For clarity, the output is offset by 2-V

20-V peak-to-peak sine with 100-nF load requires 1.9-A peak current. With a 2 A current limit, the worst-case power dissipation is around 30 W. In comparison, a 200-V amplifier would dissipate 400 W in the same scenario, which is highly impractical.

With the high- and low-voltage stages complete, the assembled dual-amplifier was tested for frequency response and drive capability with a 100-nF load. In Fig. 14.6, the amplifier's response to a 300-kHz 20-V peak-to-peak sine wave is plotted. The phase-lag between input and output is extremely low. The frequency response is plotted in the next section after a discussion of cable and interconnect inductance.

14.4.3 Cabling and Interconnects

In Sect. 14.2.1, the cable inductance was identified as a major limitation of amplifier bandwidth when driving highly capacitive loads. As cable inductance is proportional to the area enclosed in the loop between the two conductors, it is desirable to position the conductors as closely together as possible. Simply using twisted small diameter wire is not sufficient as the resistance of the conductors is also of importance. A better solution is to use copper foil for each conductor separated by a thin insulating layer, this configuration yields minimal loop area, low resistance, and low characteristic impedance (14.1).

Cables with the geometry discussed above have been developed for audio applications, one manufacturer is Alphacore. The MI-2 cable pictured in Figs. 14.3 and 14.7 contains two copper foil conductors, a polyester dielectric and a polycarbonate outer

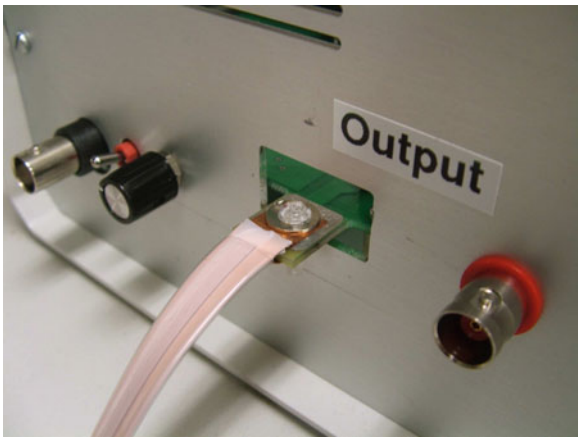


Fig. 14.7 Close-up of low-inductance connection between amplifier circuit board and cable

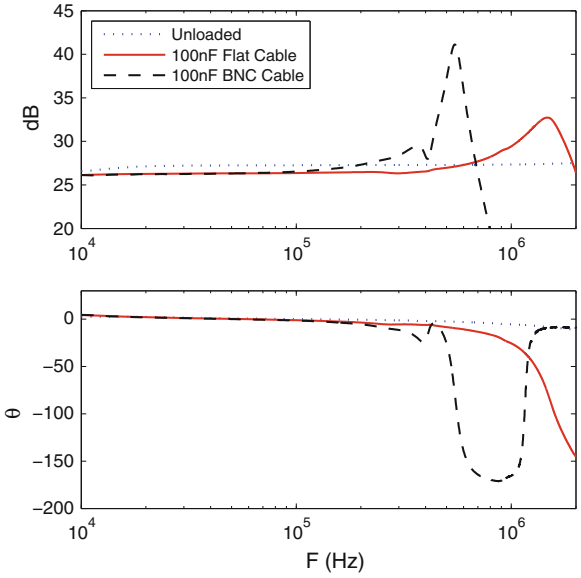


Fig. 14.8 Amplifier frequency response with open-circuit and capacitive loads

layer. The width and thickness is 9.5×0.254 mm which provides a satisfactory resistance of 14-mΩ per meter. Thanks to the high dielectric strength of the insulator, the conductors are separated by only 76 μm which results in a cable inductance of only 33 nH/m, approximately an order of magnitude less than RG-58 coaxial cable (which is 250 nH/m).

The inductance of connectors between the cable and amplifer is also of importance. In this work, to minimize loop area, the cable is connected directly to the amplifier

circuit board by bolting or soldering it to the exposed traces on the top and bottom surfaces. This configuration is pictured in Fig. 14.7 where a nylon bolt is used to fix the cable onto the PCB.

The frequency response from amplifier input to load voltage is plotted in Fig. 14.8. With no load, the amplifier bandwidth is exceptional at around 8 MHz. With a 100-nF load and a standard 50-Ω coaxial cable, a cable resonance appears at 350 kHz. In contrast, the resonance frequency with MI-2 cable is 1.5 MHz.

14.5 Chapter Summary

In this chapter, the bandwidth limitations of standard piezoelectric drives were identified as:

- High-output impedance
- The presence of a pole in the voltage-feedback loop due to output impedance and load capacitance
- Insufficient current capacity due to power dissipation
- High cable and connector inductance.

These limitations were overcome by combining a standard high-voltage amplifier with a fast low-voltage stage. Due to the lesser supply voltage of the low-voltage amplifier, it was possible to utilize a heavily biased bipolar output stage to provide low-output impedance. The low supply voltage also allows an order of magnitude increase in output current with no significant increase in power or dissipation requirements.

Cable and interconnect inductance proved to be the greatest limitation to bandwidth. A flat foil cable with ultra-low inductance was proposed for maximum bandwidth.

The completed amplifier's bandwidth, with 100-nF load was measured at 2 MHz. The main limitation was a cable resonance at 1.5 MHz. The device was demonstrated to drive a 100-nF load with a 300-kHz 20-V peak-to-peak sine wave with negligible phase-lag. Although peak-to-peak current was 3.8 A, the worst-case power dissipation is only 30 W.

References

Bonnell D (ed) (2001) Scanning probe microscopy and spectroscopy. Theory, techniques, and applications, 2nd edn. Wiley, Hoboken

Fleming AJ (2008) Techniques and considerations for driving piezoelectric actuators at high-speed. In: Proceedings of the SPIE smart materials and structures, San Diego, CA

Horowitz P, Hill W (1989) The art of electronics. Cambridge University Press, Cambridge

Inman DJ (2006) Vibration with control. Wiley, Chichester

Moheimani SOR, Fleming AJ (2006) Piezoelectric transducers for vibration control and damping. Springer, Berlin

Preumont A (2006) Mechatronics, dynamics of electromechanical and piezoelectric systems. Springer, Heidelberg

Rost MJ, Crama L, Schakel P, van Tol E, van Velzen-Williams GBEM, Overgauw CF, ter Horst H, Dekker H, Okhuijsen B, Seynen M, Vijftigschild A, Han P, Katan AJ, Schoots K, Schumm R, van Loo W, Oosterkamp TH, Frenken JWM (2005) Scanning probe microscopes go video rate and beyond. Rev Sci Instrum 76(5):053 710-1–053 710-9

Rubio-Sierra FJ, Heckle WM, Stark RW (2005) Nanomanipulation by atomic force microscopy. Adv Eng Mater 7(4):193–196

Schitter G, Åström KJ, DeMartini BE, Thurner PJ, Turner KL, Hansma PK (2007) Design and modeling of a high-speed AFM-scanner. IEEE Trans Control Syst Technol 15(5):906–915

Tseng AA, Jou S, Notargiacomo A, Chen TP (2008) Recent developments in tip-based nanofabrication and its roadmap. J Nanosci Nanotechnol 8(5):2167–2186

Tseng AA, Notargiacomob A, Chen TP (2005) Nanofabrication by scanning probe microscope lithography: a review. J Vac Sci Technol 23(3):877–894

Uchino K, Giniewica JR (2003) Micromechatronics. Marcel Dekker, New York

Zou Q, Leang KK, Sadoun E, Reed MJ, Devasia S (2004) Control issues in high-speed AFM for biological applications: collagen imaging example. Asian J Control 6(2):164–176

Index

A
Acceleration, 285, 288
Actuation, 61
Actuator dynamics, 223
AFM imaging, 3, 195, 256, 261, 268, 308
Amplifier bandwidth, 398
Amplifier noise, 354
Analog implementation, 193
Atomic force microscope (AFM), 3

B
Bandwidth, 109, 396

C
Cables, 405
Calibration and nonlinearity, 105
Capacitive sensor, 127, 133
Charge drives, 317
Charge versus voltage, 332
Classical Prandlt-Ishlinksii model, 306
Closed-loop noise, 359
Closed-loop noise spectrum, 360
Command shaping, 275
Compliance, 65
Connectors, 405
Continuous random processes, 343
Correlation functions, 344
Creep, 7, 32, 35
Current, 398

D
Damping control, 230
Drift, 107
Dual-amplifier, 399
Dual-stage repetitive control, 201
Duhem model, 301
Dynamics inversion, 253

E
Eddy-current sensor, 134
Electrical considerations, 395
Electrothermal sensor, 133
Environment, 58
Expected value, 339
External noise, 354

F
Failure considerations, 74
Feedback control, 10, 277, 292
Feedback versus feedforward control, 369
Feedforward and feedback control, 258
Feedforward control, 12, 240, 251, 292
Feedforward hysteresis compensation, 307
Filtered random processes, 347
Finite element analysis, 75
Flexure hinges, 62
Force feedback control, 221
Force sensor dynamics, 225
Force sensor noise, 226
Fourier series, 282
Frequency domain cost functions, 282
Frequency weighted objectives, 288

G
Gaussian random processes, 344
Gaussian random variables, 341

A. J. Fleming and K. K. Leang, *Design, Modeling and Control of Nanopositioning Systems*, Advances in Industrial Control,
DOI: 10.1007/978-3-319-06617-2, © Springer International Publishing Switzerland 2014

H
Hysteresis, 7, 31, 205, 252
Hysteresis modeling and control, 299

I
Interferometer, 140
Inversion, 276
Iterative feedforward control, 261

J
Joint density functions, 343

L
Laser interferometer, 140
Linear encoder, 144
Linear variable displacement transformers (LVDTs), 137

M
Manufacturing, 78
Materials, 75
Maxwell slip model, 300
Mechanical design, 57
Mechanical dynamics, 227
MEMs sensors, 133
Metrological traceability, 117
Minimizing signal power, 283
Minimizing velocity and acceleration, 285
Minimum acceleration, 288
Minimum velocity, 287
Model-based ILC, 265
Modeling, 223, 252
Modeling hysteresis, 300
Moments, 339

N
Nanometer position sensors, 118
Nanopositioner types, 43
Noise, 188, 337, 351
Noise measurement, 370, 373, 382, 386
Noise performance, 245
Noise sensitivity functions, 359
Nonlinear ILC, 267

O
Optimal inputs, 279

P
Parallel kinematic, 83
PI control, 180, 364
PI control with IRC damping, 183
PI control with notch filters, 181, 366
Piezoelectric actuator mounting, 27, 84
Piezoelectric compositions, 20
Piezoelectric manufacture, 22
Piezoelectric sensor, 123
Piezoelectric stack, 47, 322
Piezoelectric transducers, 17, 23
Piezoelectric tube, 43, 45, 166, 325
Piezoelectricity, 17
Piezoresistive sensors, 121
Polynomial model, 300
Position sensors, 103
Power, 398
Power spectral density, 345
Prandlt-Ishlinksii model, 309
Preamplification, 370
Preisach model, 302
Preload, 29, 85
Presiach model, 307
Probability distributions, 339

R
Random processes, 338
Repetitive control, 196
Resistive strain sensors, 118
Resolution, 113, 188, 351
Resonance, 9
Root-mean-square (RMS), 339

S
Scan generation, 288
Scanning probe microscope (SPM), 3
Self heating, 30
Sensor characteristics, 105
Sensor comparison, 147
Sensor noise, 110, 353
Serial kinematic, 79, 83
Shunt circuit implementation, 164
Shunt circuit modeling, 157
Shunt control, 155
Shunt damping, 159
Shunt design, 163
Signal optimization, 280
Spectral density, 349
Spectrum estimation, 372
Speed limitations, 81
Stability, 107

Stationarity, 343
Synthetic impedance, 164

T
Temperature dependence, 33
Thermal drift, 8
Thermal sensor, 133
Thermal stability, 77
Time domain cost function, 286
Time domain noise, 379
Total integrated noise, 379
Tracking control, 232

V
Variance, 339
Velocity, 285, 287
Vibration, 35

W
White noise, 348