



A review of nanometer resolution position sensors: Operation and performance

Andrew J. Fleming*

School of Electrical Engineering and Computer Science, University of Newcastle, Callaghan, NSW 2308, Australia

ARTICLE INFO

Article history:

Received 10 July 2012

Received in revised form 12 October 2012

Accepted 12 October 2012

Available online 9 November 2012

Keywords:

Position sensors

Nanopositioning

Resolution

Linearity

Noise

Metrology

ABSTRACT

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 the machine, the sensor performance is a foremost consideration. The first goal of this article is to define concise performance metrics and to provide exact and approximate expressions for error sources including non-linearity, 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.

© 2012 Elsevier B.V. All rights reserved.

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 [1,2] or nanofabrication [3,4], 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.

This article 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 article is aligned with the definitions and methods reported in the ISO/IEC 98:1993 Guide to the Expression of Uncertainty in Measurement [5], and the ISO 5723 Standard on Accuracy (Trueness and Precision) of Measurement Methods and Results [6].

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 Section 2. The variance is then utilized to define a concise statistical description of the resolution, which is a straight-forward function of the noise density, bandwidth, and $1/f$ corner frequency.

The second goal of this article 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 Section 3.1. These devices are often used for closed-loop control of piezoelectric actuators but are limited by temperature dependence and low sensitivity [7]. Piezoresistive and piezoelectric strain sensors provide improved sensitivity but at the cost of stability and DC performance.

* Tel.: +61 02 49216493; fax: +61 02 4921 6993.

E-mail address: Andrew.Fleming@newcastle.edu.au

The most commonly used sensors in nanopositioning systems [8] are the capacitive and eddy-current sensors discussed in Sections 3.4 and 3.6. Capacitive and eddy-current sensors are more complex than strain sensors but can be designed with sub-nanometer resolution, albeit with comparably small range and low bandwidth. They are used extensively in applications such as atomic force microscopy [2,9–11] and nanofabrication [12,4]. The linear variable displacement transformer (LVDT) described in Section 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 Section 3.8. Although bulky and costly, the interferometer has been the sensor of choice for applications such as IC wafer steppers [13,14] and metrological systems [15]. New fibre 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 Section 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 [16], 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 [17]. 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 non-linearities 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 [18,19]. However, to date, the linearity and hysteresis of this approach has not been reported.

2. Sensor characteristics

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

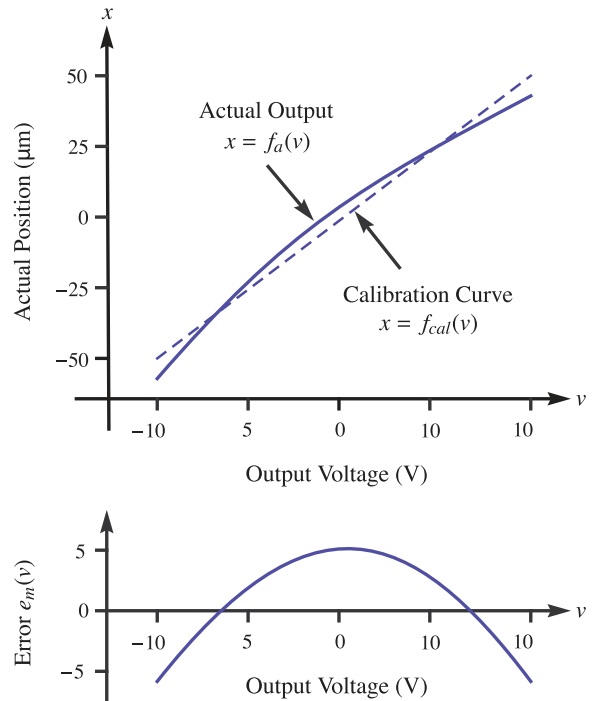


Fig. 1. The actual position versus the output voltage of a position sensor. The calibration function $f_{cal}(v)$ is an approximation of the sensor mapping function $f_a(v)$ where v is the voltage resulting from a displacement x . $e_m(v)$ is the residual error.

nonlinearity. These effects must be measured and accounted for in order to minimize the uncertainty in position.

The typical output voltage curve for a capacitive position sensor is illustrated in Fig. 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

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

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. 1, is a straight line of best fit,

$$f_{cal}(v) = \theta_0 + \theta_1 v. \quad (2)$$

In the above equation, the sensor offset is θ_0 and the sensitivity is θ_1 $\mu\text{m}/\text{V}$. More complex mapping functions are also commonly used, including the higher order polynomials

$$f_{cal}(v) = \theta_0 + \theta_1 v + \theta_2 v^2 + \theta_3 v^3 \dots \quad (3)$$

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 (1), that is

$$e_m(v) = f_a(v) - f_{cal}(\theta^*, v). \quad (4)$$

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 $e_m(v)$ [20], this may underestimate the positioning error if $e_m(v)$ is not symmetric. A more conservative definition of the mapping error (e_m) is

$$e_m = \pm \max |e_m(v)|. \quad (5)$$

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

$$\text{mapping error (\%)} = \pm 100 \frac{\max |e_m(v)|}{\text{FSR}}. \quad (6)$$

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 [21]. However, sensors that exhibit hysteresis have poor repeatability and are generally not considered for precision sensing applications.

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,

$$f_a(v) = (1 + k_s)f_a^*(v) + k_o, \quad (7)$$

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

$$e_d(v) = (1 + k_s)f_a^*(v) + k_o - f_{cal}(v). \quad (8)$$

Eqs. (7) and (8) are illustrated graphically in Fig. 2. If the nominal mapping error is assumed to be small, the expression for error can be simplified to

$$e_d(v) = k_s f_{cal}(v) + k_o. \quad (9)$$

That is, the maximum error due to drift is

$$e_d = \pm(k_s \max |f_{cal}(v)| + k_o). \quad (10)$$

Alternatively, if the nominal calibration can not 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).

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° causes a measurement error of 10% FSR.

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

$$e_{bw}(s) = x(s) - v(s), \quad (11)$$

which is equal to

$$e_{bw}(s) = x(s)(1 - P(s)), \quad (12)$$

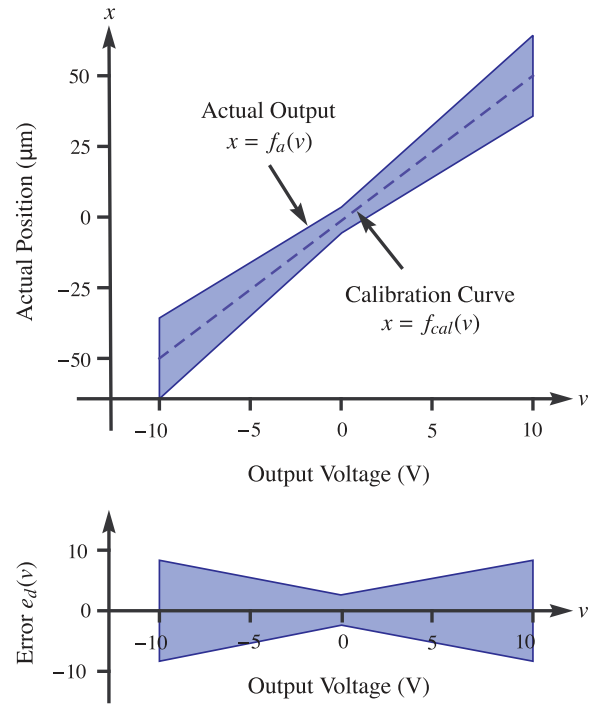


Fig. 2. The worst case range of a linear mapping function $f_a(v)$ for a given error in sensitivity and offset. In this example the greatest error occurs at the maximum and minimum of the range.

where $P(s)$ 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

$$e_{bw} = \pm A |1 - P(s)|. \quad (13)$$

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

$$e_{bw} = \pm \frac{\text{FSR}}{2} |1 - P(s)|. \quad (14)$$

The error resulting from a Butterworth response is plotted against normalized frequency in Fig. 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

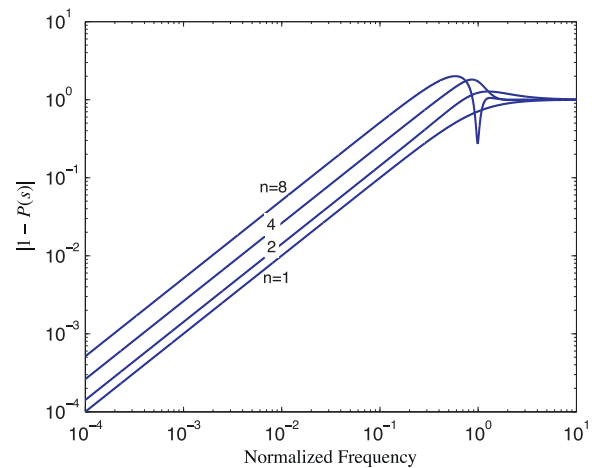


Fig. 3. The magnitude of error caused by the sensor dynamics $P(s)$. The frequency axis is normalized to the sensor 3 dB bandwidth. Lower order sensor dynamics result in lower error but typically result in significantly lesser bandwidths. In this example the dynamics are assumed to be n th order Butterworth.

the cut-off frequency, the low-frequency magnitude of $|1 - P(s)|$ is approximately

$$|1 - P(s)| \approx n \frac{f}{f_c}, \quad (15)$$

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

$$e_{bw} \approx \pm An \frac{f}{f_c}. \quad (16)$$

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 10,000 times greater than the signal frequency.

In the above derivation, the position signal was assumed to be sinusoidal, for different trajectories, the maximum error must be found by simulating Eq. (12). Although the RMS error can be found analytically by applying Parseval's equality, there is no straight-forward 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.

2.4. Noise

In addition to the actual position signal, all sensors produce some additive measurement noise. In many types of sensor, the majority of noise arises from the thermal noise in 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

$$R_{\mathcal{X}}(\tau) = E[\mathcal{X}(t)\mathcal{X}(t + \tau)], \quad (17)$$

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,

$$\text{Var}\mathcal{X} = E[(\mathcal{X} - E[\mathcal{X}])^2]. \quad (18)$$

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

$$\sigma_{\mathcal{X}} = \text{standard deviation of } \mathcal{X} = \sqrt{\text{Var}\mathcal{X}} \quad (19)$$

The standard deviation is also the root-mean-square (RMS) value of a zero-mean random process. 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,

$$2 \times E \left[\frac{1}{T} \left| \mathcal{F} \{ \mathcal{X}_T(t) \} \right|^2 \right] \Rightarrow S_{\mathcal{X}}(f) \text{ as } T \Rightarrow \infty. \quad (20)$$

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. 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

$$S_{\mathcal{X}}(f) = 2\mathcal{F} \{ R_{\mathcal{X}}(\tau) \} = 2 \int_{-\infty}^{\infty} R_{\mathcal{X}}(\tau) e^{-j2\pi f\tau} d\tau, \quad (21)$$

and

$$R_{\mathcal{X}}(\tau) = \frac{1}{2} \mathcal{F}^{-1} \{ S_{\mathcal{X}}(f) \} = \frac{1}{2} \int_{-\infty}^{\infty} S_{\mathcal{X}}(f) e^{j2\pi f\tau} df. \quad (22)$$

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

$$\sigma_{\mathcal{X}}^2 = E[\mathcal{X}^2(t)] = R_{\mathcal{X}}(0) = \int_0^{\infty} S_{\mathcal{X}}(f) df. \quad (23)$$

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,

$$\text{spectral density} = \sqrt{S_{\mathcal{X}}(f)}. \quad (24)$$

The units of $\sqrt{S_{\mathcal{X}}(f)}$ are units/ $\sqrt{\text{Hz}}$ rather than units²/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

$$S(f) = A \frac{f_{nc}}{|f|} + A, \quad (25)$$

where A is power spectral density/ Hz and f_{nc} is the noise corner frequency illustrated in Fig. 4. The variance of this process can be found by evaluating Eq. (23). That is,

$$\sigma^2 = \int_{f_i}^{f_h} A \frac{f_{nc}}{|f|} + A df, \quad (26)$$

where f_i and f_h define the bandwidth of interest. Extremely low-frequency noise components are considered to be drift. In positioning applications, f_i is typically chosen between 0.01 Hz and 0.1 Hz. By solving Eq. (26), the variance is

$$\sigma^2 = Af_{nc} \ln \frac{f_h}{f_i} + A(f_h - f_i). \quad (27)$$

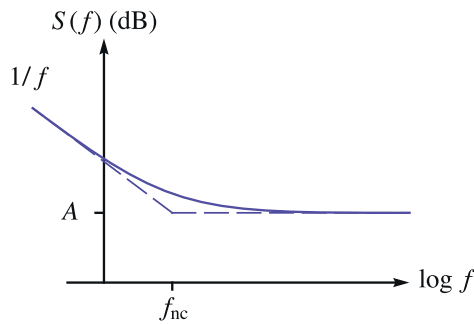


Fig. 4. A constant power spectral density that exhibits $1/f$ noise at low frequencies. The dashed lines indicate the asymptotes.

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

$$\sigma^2 = Af_{nc} \ln \frac{f_h}{f_l} + Ak_e f_h, \tag{28}$$

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 [22].

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 artefacts. 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 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 Section 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 [23] states that there

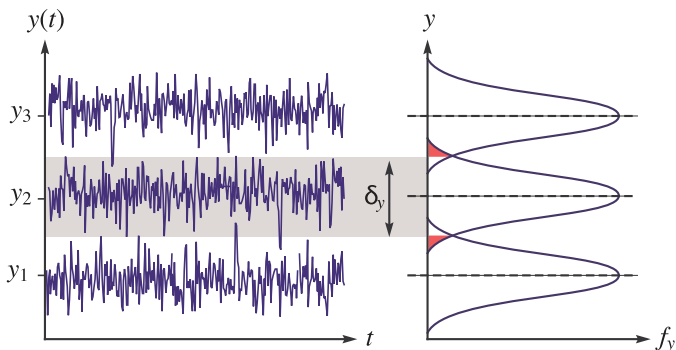


Fig. 5. The time-domain recording $y(t)$ of a position sensor at three discrete positions y_1, y_2 and y_3 . The large shaded area represents the resolution of the sensor and the approximate peak-to-peak noise of the sensor. The probability density function f_y of each signal is shown on the right.

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 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 [6] 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 times the sensor bandwidth [24]. An anti-aliasing filter is required with a cut-off frequency 7.5 times 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 [25].

Another important parameter that must be specified when quoting resolution is the sensor bandwidth. In Eq. (28), the variance of a noise process is shown to be approximately proportional to the bandwidth f_h . By combining Eq. (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} ,

$$6\sigma\text{-resolution} = 6\sqrt{A} \sqrt{f_{nc} \ln \frac{f_h}{f_l} + k_e f_h}. \tag{29}$$

From Eq. (29), it can be observed that the resolution is approximately proportional to the square-root of bandwidth when $f_h \gg f_{nc}$. It is also clear that the $1/f$ corner frequency limits the improvement that can be achieved by reducing the bandwidth. Note that Eq. (29) relies on a noise spectrum of the form (25) which may not adequately represent some sensors. The resolution of sensors with irregular spectrum's can be found by solving (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. 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 kHz 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

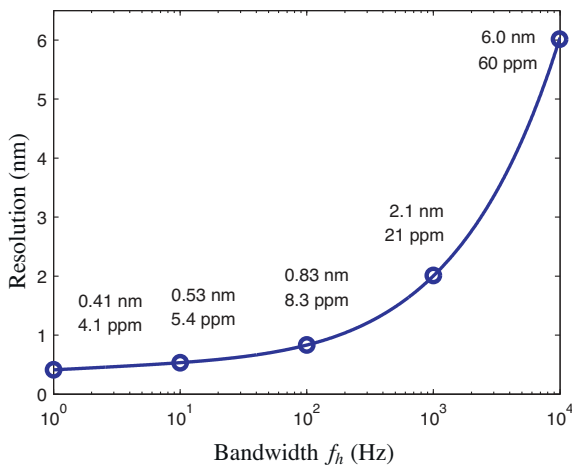


Fig. 6. The resolution versus bandwidth of a position sensor with a noise density of $10 \pm \sqrt{\text{Hz}}$ and a $1/f$ corner frequency of 10 Hz. ($f_l = 0.01$ Hz and $k_e = 1$). At low frequencies, the noise is dominated by $1/f$ noise; however, at high frequencies, the noise increases by a factor of 3.16 for every decade of bandwidth.

convenient method for reporting this ratio is in parts-per-million (ppm), that is

$$\text{DNR}_{\text{ppm}} = 10^6 \frac{6\sigma\text{-resolution}}{\text{full scale range}} \quad (30)$$

This measure is equivalent to the resolution in nanometers of a sensor with a range of 1 mm. In Fig. 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,

$$\text{DNR}_{\text{db}} = 20 \log_{10} \frac{\text{full scale range}}{6\sigma\text{-resolution}} \quad (31)$$

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 [26]. 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.

2.6. Combining errors

The exact and worst-case errors described in Section 2 are summarized in Table 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

Table 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 e_m(v) $
Drift e_d	$(1 + k_s)f_a^*(v) + k_o - f_{cal}(v)$	$\pm (k_s \max f_{cal}(v) + k_o)$
Bandwidth e_{bw}	$\mathcal{F}^{-1}\{x(s)(1 - P(s))\}$	$\pm \frac{A_{nf}}{f_c} (\text{sine-wave})$
Noise δ	NA	$6\sqrt{A} \sqrt{f_{nc} \ln \frac{f_h}{f_l} + k_e f_h}$

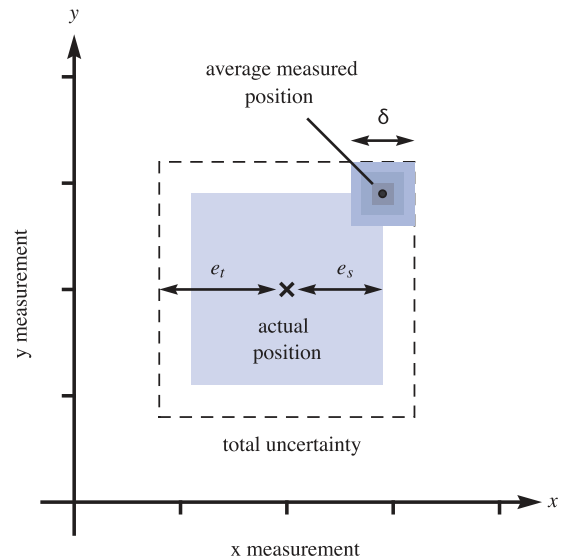


Fig. 7. The total uncertainty of a two-dimensional position measurement is illustrated by the dashed box. The total uncertainty e_t is due to both the static trueness error e_s and the noise δ .

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

$$e_t = e_m + e_d + e_{bw} + \frac{\delta}{2}, \quad (32)$$

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. 7.

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” [27]. The reference for a distance measurement is the meter standard, defined by the distance travelled by light in vacuum over $1/299\,792\,458$ s. 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 [27]. 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 [5].

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 (BSI), United Kingdom.
- National Institute of Standards and Technology (NIST), USA.

3. Nanometer position sensors

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. 14(a). Other application examples can be found in references [28–31].

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

$$GF = \frac{\Delta R/R_G}{\epsilon}, \quad (33)$$

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 in 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. 14(a), a 10 mm Noliac SCMAP07 piezoelectric actuator is pictured with a strain gauge bonded to each of the two non-electrode sides. The strain gauges are Omega SGD-3/350-LY13 gauges, with a nominal resistance of 350 Ω and package dimensions of 7 mm \times 4 mm. The electrical wiring of the strain gauges is illustrated in Fig. 8. The two-varying-element bridge circuit is completed by two dummy 350 Ω wire wound resistors and excited by a 5 V 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

$$V_s = \frac{A_v V_b}{2} \left(\frac{\Delta R}{R_G + \Delta R/2} \right), \quad (34)$$

where $A_v = 2000$ is the differential gain and $V_b = 5$ V is the excitation voltage. By substituting (33) into (34) and neglecting the small

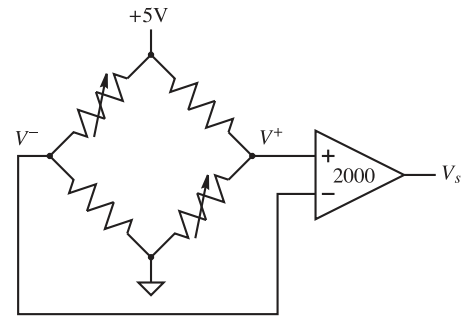


Fig. 8. A two-varying-element bridge circuit that contains two fixed resistors and two strain dependent resistors. All of the nominal resistance values are equal. A simultaneous change in the two varying elements produces a differential voltage across the bridge.

bridge non-linearity,¹ the measured voltage is proportional to the strain ϵ and displacement d by

$$V_s = \frac{1}{2} A_v V_b G F \epsilon, \quad (35)$$

$$V_s = \frac{1}{2L} A_v V_b G F d, \quad (36)$$

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/ μ m which implies a full scale voltage of 5 V from a displacement of 10 μ m. The actual sensitivity was found to be 0.3633 V/ μ m [31].

The bridge configuration shown in Fig. 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. [32]. 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 [32,30]. 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

$$S(f) = 4kTRV^2/\text{Hz}, \quad (37)$$

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

The strain gauge pictured in Fig. 14(a) has a resistance of 350 Ω , hence the spectral density is 2.4 nV/ $\sqrt{\text{Hz}}$. Since the sensitivity is 0.3633 V/ μ m, the predicted spectral density is 13 pm/ $\sqrt{\text{Hz}}$. This figure agrees with the experimentally measured spectral density

¹ In a two-varying-element bridge circuit, the non-linearity due to $\Delta R/2$ in Eq. (34) is 0.5% non-linearity per percent of strain [32]. Since the maximum strain of a piezoelectric actuator is 0.1%, the maximum non-linearity is only 0.05% and can be neglected. If this magnitude of non-linearity is not tolerable, compensating circuits are available [32].

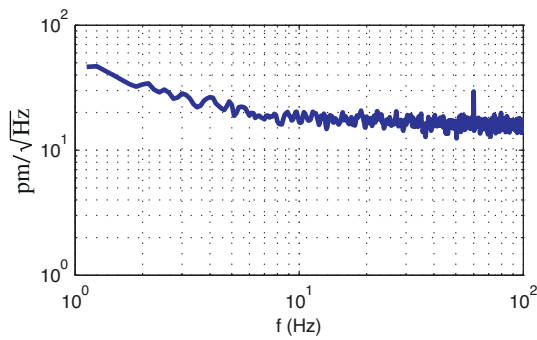


Fig. 9. The noise density of the strain sensor and instrumentation. The spectrum can be approximated by a constant spectral density and $1/f$ noise.

plotted in Fig. 9. The sensor exhibits a noise density of approximately $15 \pm \sqrt{\text{Hz}}$ 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 $1 \text{ pm}/\sqrt{\text{Hz}}$ for a range of $10 \mu\text{m}$. 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, $f_i = 0.01 \text{ Hz}$, and a first-order bandwidth of 1 kHz ($k_e = 1.57$). The resolution predicted by Eq. (29) is 580 nV or 230 ppm. In other words, if the full-scale range was $100 \mu\text{m}$, the resolution would be 23 nm, which is not competitive.

3.2. Piezoresistive strain sensors

In 1954, a visiting researcher at Bell Laboratories, C.S. Smith, demonstrated that ‘exceptionally large’ resistance changes occur in silicon and germanium when subjected to external strain [33]. This discovery was the foundation for today’s semiconductor piezoresistive sensors that are now ubiquitous in applications such as integrated pressure sensors and accelerometers [34].

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 non-linearity (1%) [34]. The elimination of these artefacts requires a more complicated conditioning circuit than metal foil strain gauges; however, integrated circuits are now available that partially compensate for non-linearity, offset, and temperature dependence, for example, the Maxim MAX1450.

As shown in Fig. 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 [34]. 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,

$$\Delta R = R_G[\pi_L \sigma_{xx} + \pi_T(\sigma_{yy} + \sigma_{zz})], \quad (38)$$

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 [34].

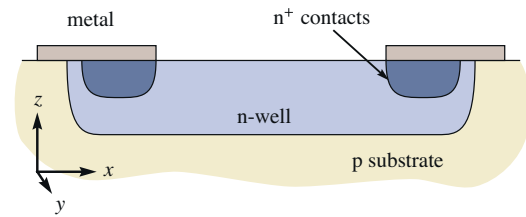


Fig. 10. A cross-section of a piezoresistive strain sensor. Deformation of the semiconductor crystal causes a resistance change 100 times that of a resistive strain gauge.

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 nanositioners [35] and MEMS devices [36]. Discrete piezoresistive sensors are also available for standard macro-scale nanositioning 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 $2.4 \text{ mm} \times 0.4 \text{ mm}$. The sensitivity is typically specified in the same way as a metal foil sensor, by the gauge factor defined in Eq. (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 non-linearity. 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 [34]. However, since piezoresistive sensors are semiconductors, the $1/f$ noise can be substantially worse [34]. Consider the Micron Instruments SS-095-060-350PU piezoresistive sensor which has a gauge factor of 120 and a resistance of 350Ω . In a two-varying-element bridge with 2 V excitation, Eq. (36) predicts that a full-scale strain of 0.1% develops 120 mV. The thermal noise due to the resistance is $2.4 \text{ nV}/\sqrt{\text{Hz}}$. 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 $100 \mu\text{m}$, 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.

3.3. Piezoelectric strain sensors

In addition to their actuating role, piezoelectric transducers are also widely utilized as high sensitivity strain sensors [37–43].

The basic operation of a piezoelectric strain sensor is illustrated in Fig. 11(a). In this case the applied force F and resulting strain $\Delta h/h$ is aligned in the same axis as the polarization vector. 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

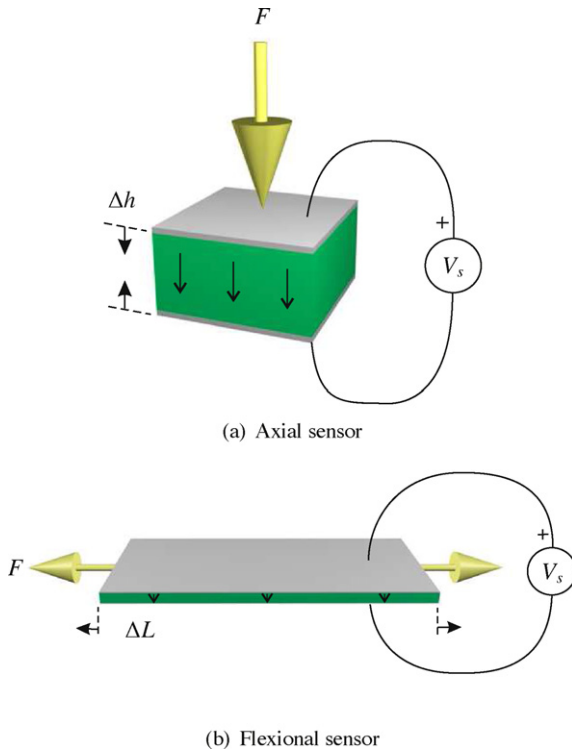


Fig. 11. A piezoelectric stack and plate strain sensor. The polarization vector is shown as a downward arrow. Axial sensors are typically used to measure dynamic forces while flexional sensors are used to measure changes in strain or curvature. (a) Axial sensor, (b) Flexional sensor.

form of the piezoelectric constituent equations, the developed electric field E is

$$E = q_{33} \frac{\Delta h}{h}, \quad (39)$$

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

$$q_{33} = \frac{d_{33}}{\epsilon^T s^D}, \quad (40)$$

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 (40) by the thickness h , the measured voltage can be written as

$$V_s = q_{33} \Delta h. \quad (41)$$

If there are multiple layers, the voltage is

$$V_s = \frac{q_{33}}{n} \Delta h, \quad (42)$$

where n is the number of layers. The developed voltage can also be related to the applied force [31].

$$V_s = \frac{nd_{33}}{C} F, \quad \text{or} \quad V_s = \frac{d_{33}h}{n\epsilon^T A} F, \quad (43)$$

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. 11(b) 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

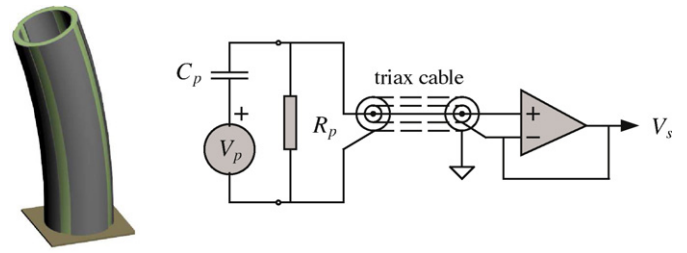


Fig. 12. A piezoelectric tube actuator with one electrode utilized as a strain sensor. The electrical equivalent circuit consists of the induced piezoelectric voltage V_p in series with the transducer capacitance. The dielectric leakage and input impedance of the buffer circuit are modeled by the parallel 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.)

is much larger than the width and thickness, the developed voltage is

$$V_s = \frac{-g_{31}}{L} F, \quad (44)$$

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

$$V_s = -g_{31} k \frac{\Delta L}{L}, \quad (45)$$

$$V_s = \frac{-g_{31} A}{s^D L} \frac{\Delta L}{L}, \quad (46)$$

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 [44,45,37]. 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 [39,40,42]. This application is illustrated in Fig. 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 $N_{V_s}(\omega)$ 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. 13. Neglecting the leakage resistance R , the noise density of the sensor voltage is

$$N_{V_s}(\omega) = i_n \frac{1}{C\omega}, \quad (47)$$

where N_{V_s} 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. 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. 15. The

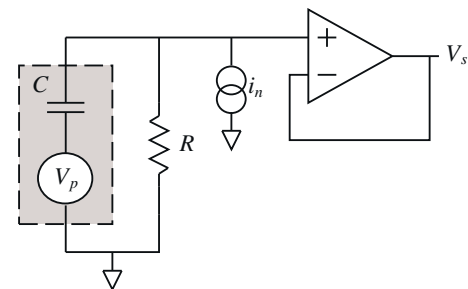


Fig. 13. The electrical model of a piezoelectric force sensor. The open-circuit voltage V_p is high-pass filtered by the transducer capacitance C and leakage resistance R . The current source i_n represents the current noise of a high-impedance buffer.

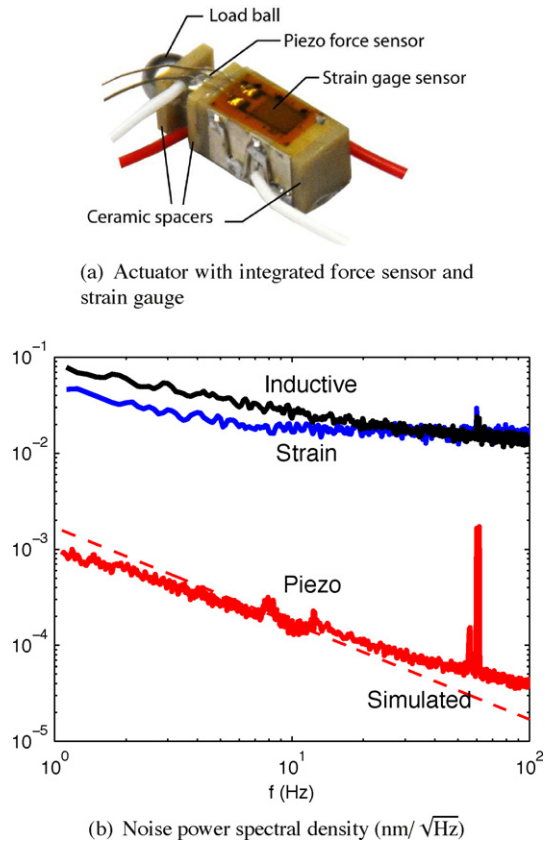


Fig. 14. (a) A piezoelectric stack actuator with an integrated force sensor and two resistive strain gauges bonded to the top and bottom surface (the bottom gauge is not visible). In (b), the noise density of the piezoelectric sensor is compared to the resistive strain gauge and a Kaman SMU9000-15N inductive sensor, all signals are scaled to nm/√Hz. The simulated noise of the piezoelectric force sensor is also plotted as a dashed line. (a) Actuator with integrated force sensor and strain gauge and (b) noise power spectral density (nm/√Hz).

sensor has a capacitance of 30 nF and the voltage buffer (OPA606) has a noise density of 2 fA/√Hz. Further details on the behavior of piezoelectric force sensors can be found in [31].

In Fig. 14(b) 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

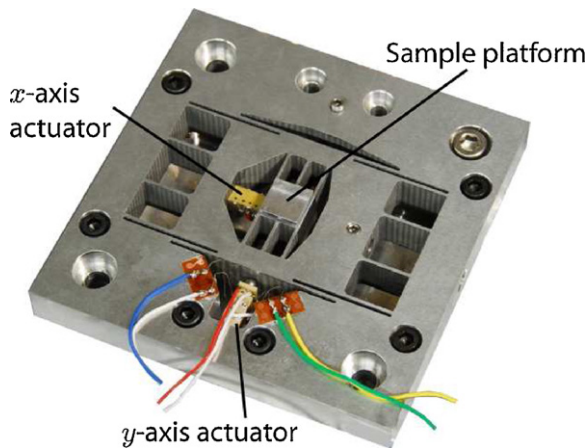


Fig. 15. A nanopositioning platform with two varying-element strain gauge fitted to the y-axis actuator [31]. The nanopositioner is driven by two piezoelectric stack actuators that deflect the sample platform by a maximum of 10 μm in the x and y lateral axes.

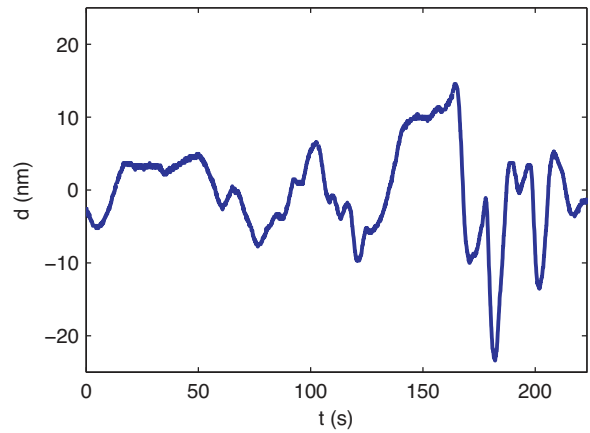


Fig. 16. Low-frequency noise of the piezoelectric sensor pictured in Fig. 14(a), scaled to nanometers. The peak-to-peak noise over 220 seconds is 38 nm or 26 mV.

noise density also continues to reduce at higher frequencies. However, at low-frequencies the noise of the piezoelectric force sensor eventually surpasses the noise of 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. 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. 13 is high-pass filtered by the internal transducer capacitance C and the leakage resistance R . The cut-off frequency is

$$f_{hp} = \frac{1}{2\pi RC} \text{ Hz.} \quad (48)$$

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°, the piezoelectric force sensor cannot be used to measure frequencies of less than 0.5 Hz.

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.

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 [46]. However, due to the electronics required for measuring the capacitance and deriving position, capacitive sensors are inherently more complex than sensors such as resistive strain gauges. Larger ranges can be achieved with the use of an encoder-style electrode array [47].

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

$$C = \frac{\epsilon_0 \epsilon_r A}{h}, \quad (49)$$

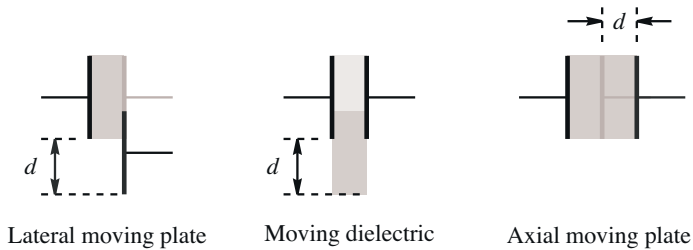


Fig. 17. Types of capacitive sensor. The axial moving plate produces the highest sensitivity but the smallest practical travel range. Lateral moving plate and moving dielectric sensors are most useful in long-range applications.

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. 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

$$C = \frac{\epsilon_0 \epsilon_r A}{d}, \quad (50)$$

hence, the sensitivity is

$$\frac{dC}{dd} = \frac{C_0}{d_0} \text{ F/m}, \quad (51)$$

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 [20].

A practical parallel plate capacitive sensor is illustrated in Fig. 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. A summary of correction terms for different guard electrode geometries can be found in references [20] and [46].

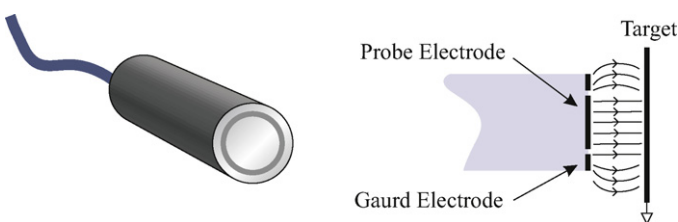


Fig. 18. A capacitive sensor probe and electrode configuration. The guard electrode is driven at the same potential as the probe in order to linearize the electric field and reduce fringing effects.

To measure the capacitance and thus derive the position, a wide variety of circuits are available [21,46]. 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,

$$I = j\omega VC, \quad (52)$$

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

$$V = \frac{I}{j\omega C}, \quad (53)$$

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. 19(a). 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. 19(b). A synchronous demodulator can be constructed from a filter and voltage controlled switch [21,46]. 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. 19(b) eliminates the large DC offset produced by single ended demodulators, such as Fig. 19(a). 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 [20] concluded

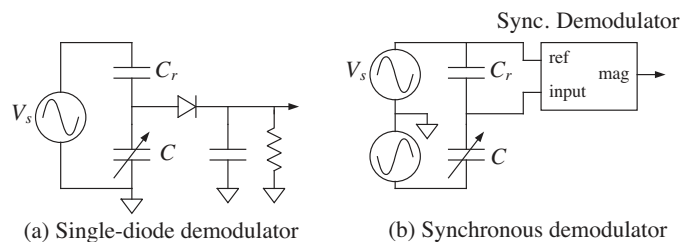


Fig. 19. Demodulation circuits for measuring capacitance. The linearity, temperature sensitivity and noise performance of the synchronous detector is significantly better than the single-diode envelope detector. (a) Single-diode demodulator and (b) Synchronous demodulator.

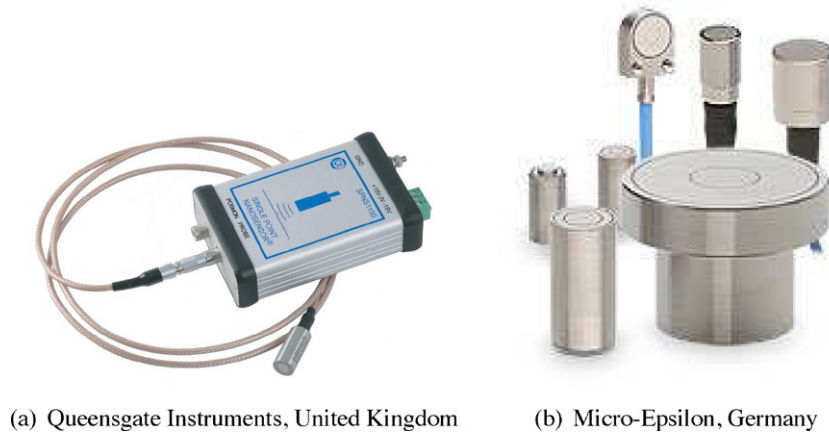


Fig. 20. An example of two commercially available capacitive sensors. Photos courtesy of Queensgate Instruments, UK and Micro-Epsilon, Germany. (a) Queensgate Instruments, United Kingdom, (b) Micro-Epsilon, Germany.

that the worst sources of non-linearity 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 [20] is contained in Table 2. Considering that the linearity of an capacitive sensor in ideal conditions can 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 (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.

Table 2

A summary of error sources in a parallel plate capacitive sensor studied in [20]. The sensor has a gap of 100 μm , a radius of 6 mm, and a nominal capacitance of 10 pF.

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

A typical commercial capacitive sensor with a range of 100 μm has a noise density of approximately of 20 pm/ $\sqrt{\text{Hz}}$ [40]. 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. (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. 20.

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. 21(a) is a common variety found in a number of nanopositioning applications, for example [48,49]. 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. 21(b). 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 [46,50].

To increase the range of motion beyond a single inter-electrode spacing, the configuration in Fig. 21(c) uses withdrawn electrodes to form a capacitive incremental encoder [51–53]. The slider can now move freely in either direction, limited only by the length of

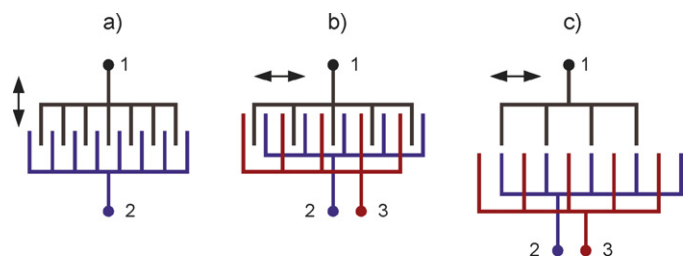


Fig. 21. Three examples of MEMs capacitive sensor geometries. (a) Standard comb sensor; (b) differential comb sensor; and (c) incremental capacitive encoder.

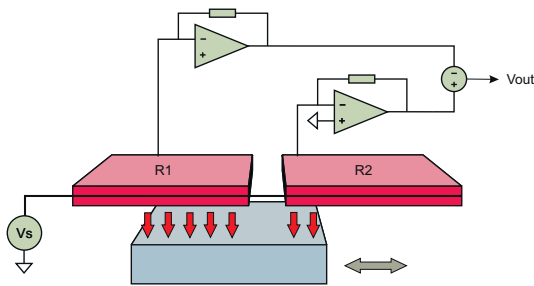


Fig. 22. An electrothermal position sensor. The two stationary micro-heaters are driven by a constant voltage source V_s . The rate of heat transfer and the resulting temperature is proportional to the overlap between the heater and the heatsink. The position of the heatsink can be estimated by measuring the current difference between the two micro-heaters which indicates the difference in resistance and temperature.

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 [54,55].

Electrothermal sensors are an alternate class of position sensors first utilized in nanopositioning applications by IBM in 2005 [56]. An example of a differential electrothermal position sensor is illustrated in Fig. 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. 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 [57–59] and nanopositioning [60,60]. 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 [49].

With a range of $100\ \mu\text{m}$, a thermal position sensing scheme achieved a noise density of approximately $10\ \text{pm}/\sqrt{\text{Hz}}$ with a $1/f$ corner frequency of approximately 3 kHz [60]. 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 [60].

3.6. Eddy-current sensors

Eddy-current, or inductive proximity sensors, operate on the principle of electromagnetic induction [62,63]. As illustrated in Fig. 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 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 [64,62,21].

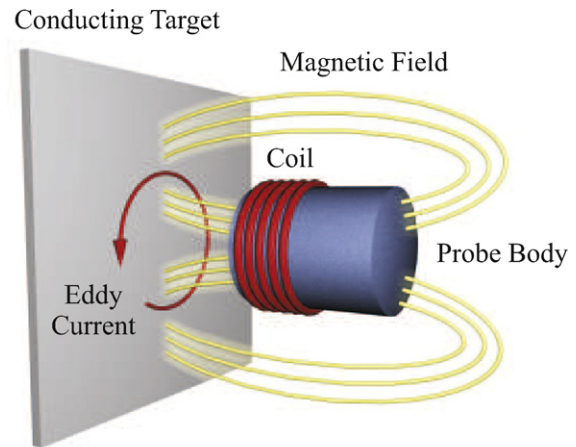


Fig. 23. The operating principle of an eddy-current sensor. An alternating current in the coil induces eddy currents in the target. Increasing the distance between the probe and target reduces the eddy currents and also the effective resistance of the coil.

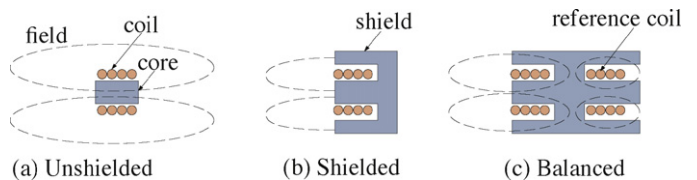


Fig. 24. Types of eddy-current sensor. The unshielded type has the greatest range but is affected by nearby fields and conductors. A shield makes the magnetic field more directional but reduces the range. A reference coil can be used to reduce the sensitivity to temperature. (a) Unshielded, (b) shielded, and (c) balanced.

Three common types of eddy-current sensor are depicted in Fig. 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 non-inductive coil that is used to null the effect of temperature variation [65]. The second coil is used in a divider or bridge configuration such as that illustrated in Fig. 25.

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

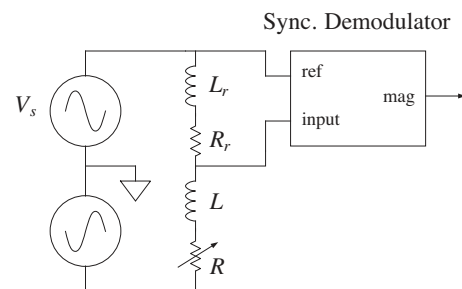


Fig. 25. Synchronous demodulation circuit for a balanced eddy-current sensor. L_r and R_r are the inductance and resistance of the reference coil.

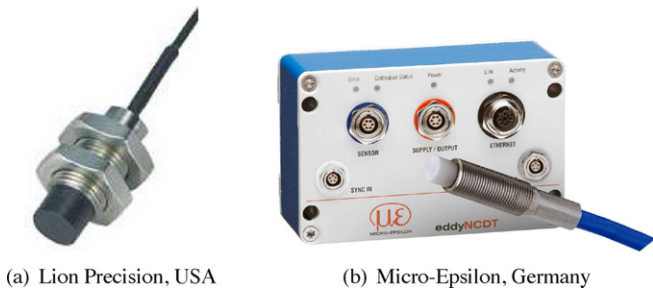


Fig. 26. Two commercially available eddy-current sensors. Photos courtesy of Lion Precision, USA and Micro-Epsilon, Germany. (a) Lion Precision, USA, (b) Micro-Epsilon, Germany.

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 μm 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. 14(b). The 1/f corner frequency is approximately 20 Hz and the constant density is approximately 20 pm/√Hz. Eq. (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 to 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. 26

3.7. Linear variable displacement transformers (LVDTs)

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 [21].

The operating principle of an LVDT is illustrated in Fig. 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

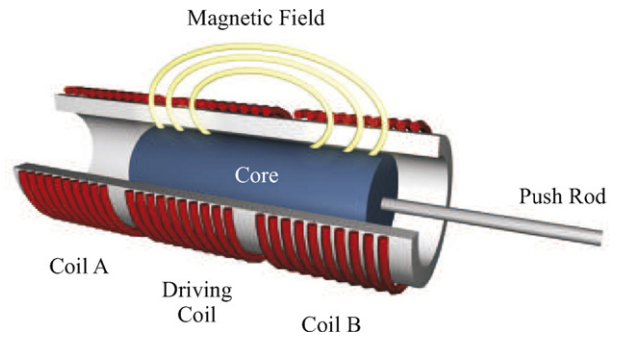


Fig. 27. The operating principle of a linear variable displacement transducer (LVDT). Changes in the core position produce a linear differential change in the coupling between the driving coil and the pick-up coils.

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 illustrated in Fig. 28.

In addition the components in Fig. 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 non-magnetic 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 [21]. 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. 29 [21]. The square-wave oscillator is replaced by a sine-wave oscillator if the electronics and LVDT are not physically collocated. Other demodulation circuits

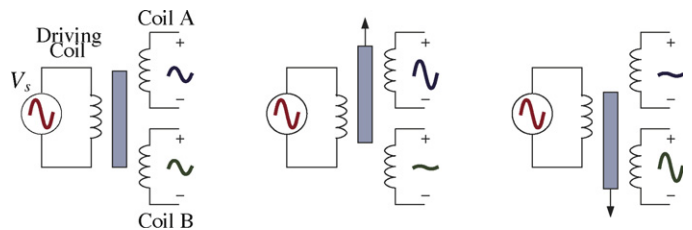


Fig. 28. The relationship between the sensor coil voltage and core position in an LVDT. The coil voltage is proportional to the amount of core it contains.

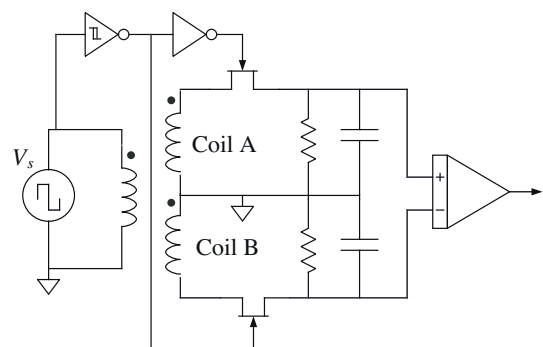


Fig. 29. A LVDT conditioning circuit with a synchronous demodulator and differential amplifier [21].

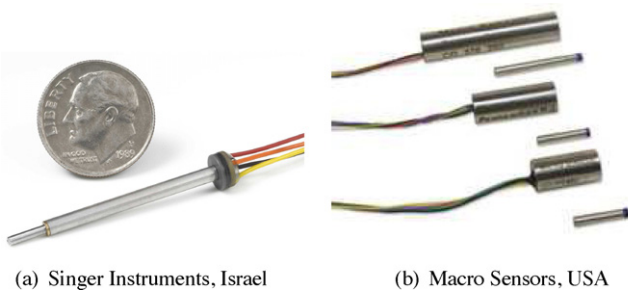


Fig. 30. Two commercially available LVDT sensors. Photos courtesy of Singer Instruments, Israel and Macro Sensors, USA. (a) Singer Instruments, Israel, (b) Macro Sensors, USA.

include the single-diode demodulator in Fig. 19(a) and the AD630 based demodulator in Fig. 19(b).

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. Non-linearity 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 non-contact 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 [66]. 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 [66].

The theoretical resolution of LVDT sensors is limited primarily by the Johnson noise of the coils and Barkhausen noise in the magnetic materials [66]. However, standard conditioning circuits like the Analog Devices AD598 produce electronic noise on the order of 50 $\mu\text{Vp-p}$ 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. 30.

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 [67]. 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 [67]. 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} [67]. 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 Section 2.7.

The operating principle of a Michelson interferometer is described in Fig. 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.

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 [68]. 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 [69]. 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.

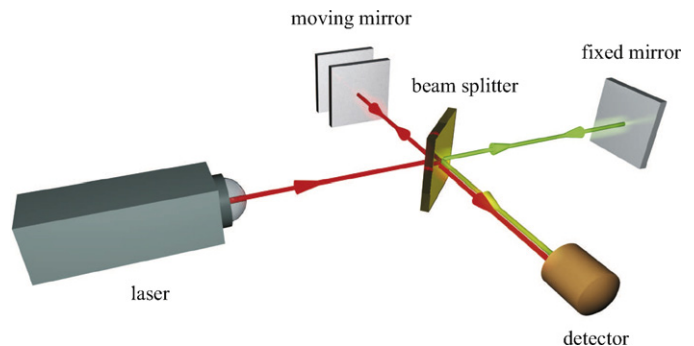


Fig. 31. The operation of a Michelson interferometer. The laser light is split into two paths, one that encounters a moving mirror and another that is fixed. The two beams are recombined and interfere at the detector. If the distance between the paths is an integer number of wavelengths, constructive interference occurs.

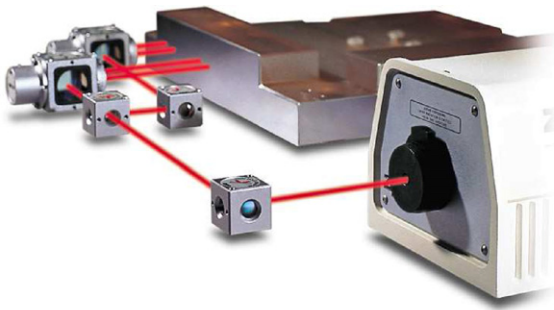


Fig. 32. A ZMI™ two-axis heterodyne interferometer with a single laser source for measuring the angle and displacement of a positioning stage. Courtesy of Zygo, USA.

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. 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 fibre interferometer, pictured in Fig. 33(a), was developed [70]. The measuring head contains a single-mode optical fibre with a 9 μm core diameter coupled to a collimator lens. Approximately 4% of the applied light is immediately reflected off the fibre termination and is returned down the fibre, forming the reference beam. The transmitted light passes through the collimator lens and is reflected off the slightly angled target mirror back towards the fibre surface but away from the core. As the fibre surface is a poor reflector, only 4% of the incident light is reflected from the fibre surface. This reflected light travels back through the lens, is reflected off the mirror and is coupled directly to the fibre core, thus forming a Fabry-Perot interferometer with a cavity length equal to twice the distance between the fibre 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. 33(b). 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. 33(c).

Since the miniature fibre 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 fibre surface, the fibre 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

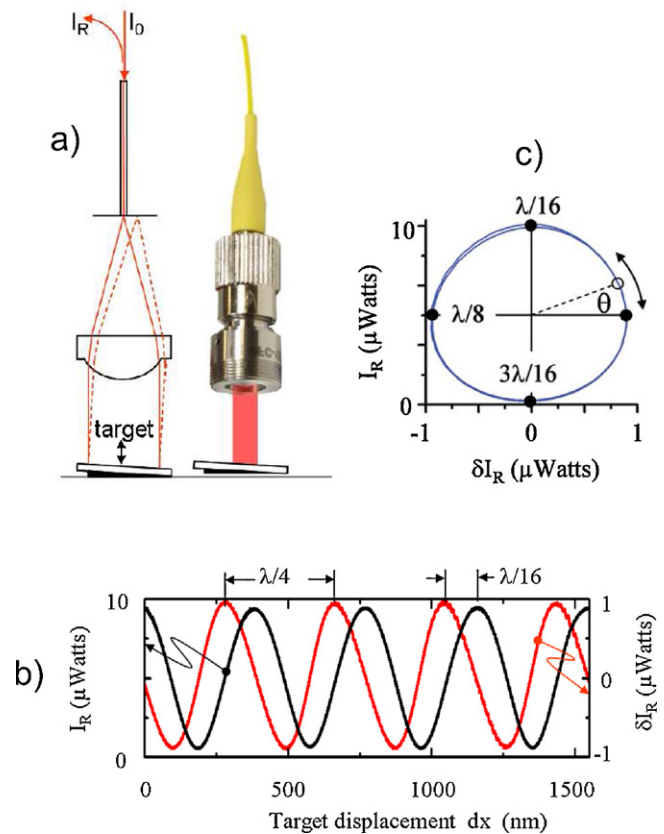


Fig. 33. The operating principle of an Attocube FPS miniature fibre interferometer [70], courtesy of Attocube, Germany. In (a) the transmitted light is reflected from the mirror, the fibre surface, the mirror again, and is then focused onto the fibre core. The interferogram plotted in (b) shows the direct 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. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

require metrological precision, for example, the metrological AFM described in [15].

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 centimetres 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 reference [70] has a $1/f$ noise corner frequency of approximately 10 Hz and a noise density of approximately $2 \text{ pm}/\sqrt{\text{Hz}}$. This results in a resolution of approximately 1.6 nm with a 12 kHz bandwidth. Eq. (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 (Fibre Interferometer); Keyence, Japan (Fibre Interferometer); Renishaw, UK; Sios, Germany; and Zygo,

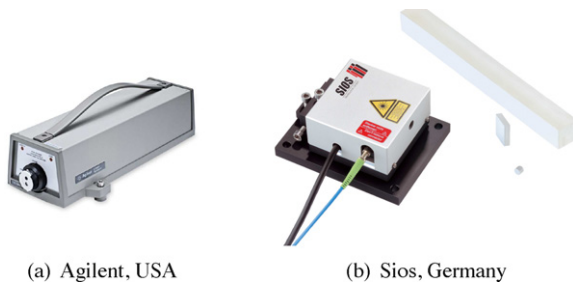


Fig. 34. Two commercially available Laser Interferometers. Photos courtesy of Agilent, USA and Sios, Germany. (a) Agilent, USA, (b) Sios, Germany.

USA. Instruments from these manufacturers are pictured in Figs. 33(a) and 34.

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 [21]. Although simple, the constant contact between the brush and scale meant a very limited life and poor reliability.

In the 1950s 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 cannot 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 callipers. 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. 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 to the distance between the reflective bars. In between the peaks, the position can be estimated from the received power. Rather than partial

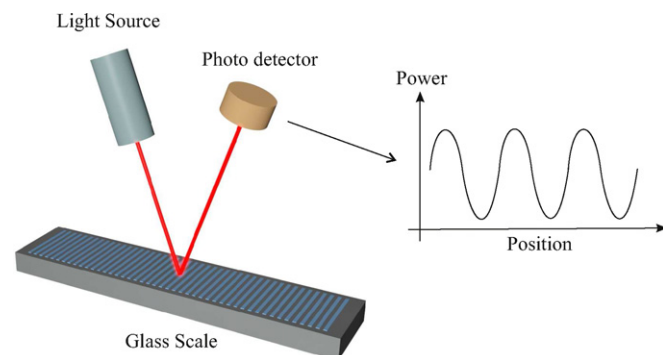


Fig. 35. The operation of a simple reflective optical encoder. The peaks in the received power correspond to the distance between reflective bars.

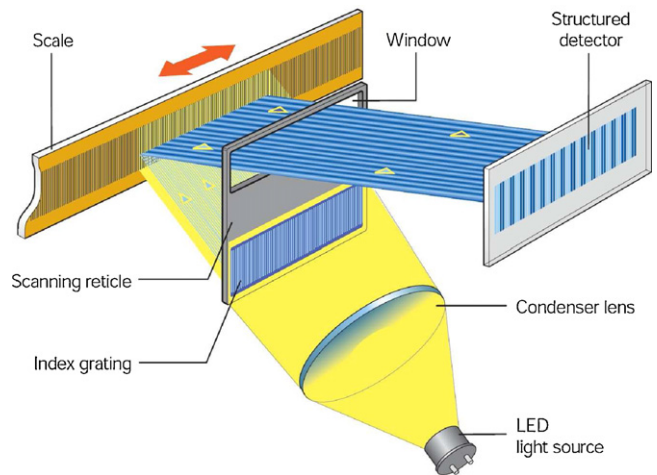


Fig. 36. The image scanning technique is used for reference scales with a grating pitch of between $10\ \mu\text{m}$ and $200\ \mu\text{m}$. Image courtesy of Heidenhain, Germany.

reflection, other gratings contain height profiles that modulate the proximity and thus received power [71].

There are two major difficulties with the design illustrated in Fig. 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.

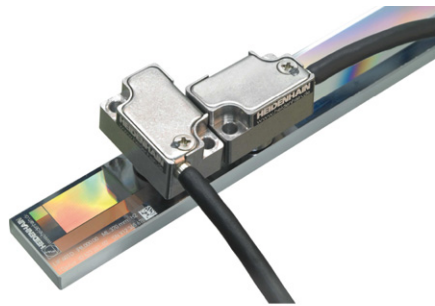
To provide immunity to dust and contamination, commercial optical encoders use a large number of parallel measurements to effectively average out errors. This principle relies on the Moiré phenomenon [72] and is illustrated by the image scanning technique shown in Fig. 36. In Fig. 36 a parallel beam of light is projected onto a reflective scale through a scanning reticle. The reflected Moiré pattern is essentially the binary product of the scanning reticle and the scale and is 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 $1\ \text{nm RMS}$, or $6\ \text{nm peak-to-peak}$. The overall accuracy is around $5\ \mu\text{m/m}$ [73], however accuracies as high as $0.5\ \mu\text{m/m}$ are possible with ranges up to $270\ \text{mm}$ [74].

The highest resolution optical encoders operate on the principle of interference [74,75]. 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 [74]. Since this technique operates on the principle of diffraction, extremely small signal periods of down to $128\ \text{nm}$ 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 [76] and MEMS mass storage devices [58].

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. 37.



(a) Heidenhain, Germany



(b) Renishaw, UK

Fig. 37. Two commercially available optical linear encoders. Photos courtesy of Heidenhain, Germany and Renishaw, UK. (a) Heidenhain, Germany; (b) Renishaw, UK.

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 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 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 $100\ \mu\text{m}$ and a first-order bandwidth of 1 kHz. The low-frequency limit is assumed to be $f_l = 0.01\ \text{Hz}$.
- The quoted accuracy is the typical static trueness error defined in Section 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 $500\ \mu\text{m}$. The parameters in Table 3 pertain to the example of a two varying element bridge discussed in Section 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 $1\ \mu\text{m}$. 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 non-contact.

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 Section 3.4.

Eddy current sensors can provide excellent resolution for travel ranges greater than $100\ \mu\text{m}$. 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 Section 3.6.

LVDT sensors are among the most popular in industrial applications requiring a range from a few millimetres 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 $1\ \mu\text{m}$. 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 3 is associated with the Fabry-Perot interferometer discussed in Section 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 $5\ \mu\text{m}/\text{m}$ 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.

Table 3

Summary of position sensor characteristics. The dynamic range (DNR) and resolution are approximations based on a full-scale range of $100\ \mu\text{m}$ and a first-order bandwidth of 1 kHz.

Sensor type	Range	DNR	Resolution	Max. BW	Accuracy
Metal foil	10–500 μm	230 ppm	23 nm	1–10 kHz	1% FSR
Piezoresistive	1–500 μm	4.9 ppm	0.49 nm	>100 kHz	1% FSR
Capacitive	10 μm to 10 mm	24 ppm	2.4 nm	100 kHz	0.1% FSR
Electrothermal	10 μm to 1 mm	100 ppm	10 nm	10 kHz	1% FSR
Eddy current	100 μm to 80 mm	10 ppm	1 nm	40 kHz	0.1% FSR
LVDT	0.5–500 mm	10 ppm	5 nm	1 kHz	0.25% FSR
Interferometer	Meters		0.49 nm	>100 kHz	1 ppm FSR
Encoder	Meters		6 nm	>100 kHz	5 ppm FSR

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 [13,14]. 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 [77–80] and stepping mechanisms [81,82]. Capacitive sensors can be adapted for this purpose by using a periodic array of electrodes [55]. 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 [83,84]; nanofabrication [12,4,3,85], and imaging of fast biological and physical processes [86–90,10]. 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 [43,41]. Technologies such as piezoresistive sensors [91] 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 [18,19].

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 reference [40] 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 [41,31].

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 reference [58] 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 [49]. Multiple sensors can be combined by complementary filters [41] or by an optimal technique in the time domain [40] or frequency domain [60]. 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.

Acknowledgements

This work was supported financially by the Australian Research Council (DP0986319) and the Center for Complex Dynamic Systems and Control.

References

- [1] D.Y. Abramovitch, S.B. Andersson, L.Y. Pao, G. Schitter, A tutorial on the mechanisms, dynamics, and control of atomic force microscopes, in: Proc. American Control Conference, New York City NY, New York City, NY, 2007, pp. 3488–3502.
- [2] S.M. Salapaka, M.V. Salapaka, Scanning probe microscopy, *IEEE Control Systems* 28 (2) (2008) 65–83.
- [3] A.A. Tseng (Ed.), *Nanofabrication: Fundamentals and Applications*, World Scientific, Singapore, 2008.
- [4] J.A. Vicary, M.J. Miles, Pushing the boundaries of local oxidation nanolithography: short timescales and high speeds, *Ultramicroscopy* 108 (10) (2008) 1120–1123.
- [5] ISO/IEC Guide 98:1993 Guide to the Expression of Uncertainty in Measurement, International Organization for Standardization, 1994.
- [6] ISO 5725 – Accuracy (Trueness and Precision) of Measurement Methods and Results, International Organization for Standardization, 1994.
- [7] G. Schitter, R.W. Stark, A. Stemmer, Sensors for closed-loop piezo control: strain gauges versus optical sensors, *Measurement Science and Technology* 13 (2002) N47–N48.
- [8] S. Devasia, E. Eleftheriou, S.O.R. Moheimani, A survey of control issues in nanopositioning, *IEEE Transactions on Control Systems Technology* 15 (5) (2007) 802–823.
- [9] K.K. Leang, Q. Zou, S. Devasia, Feedforward control of piezoactuators in atomic force microscope systems, *IEEE Control Systems* 29 (1) (2009) 70–82.
- [10] A.J. Fleming, B.J. Kenton, K.K. Leang, Bridging the gap between conventional and video-speed scanning probe microscopes, *Ultramicroscopy* 110 (9) (2010) 1205–1214, <http://dx.doi.org/10.1016/j.ultramic.2010.04.016>.
- [11] A.J. Fleming, S.S. Aphale, S.O.R. Moheimani, A new method for robust damping and tracking control of scanning probe microscope positioning stages, *IEEE Transactions on Nanotechnology* 9 (4) (2010) 438–448, doi:10.1109/TNANO.2009.2032418.
- [12] A.A. Tseng, S. Jou, A. Notargiacomo, T.P. Chen, Recent developments in tip-based nanofabrication and its roadmap, *Journal of Nanoscience and Nanotechnology* 8 (5) (2008) 2167–2186.
- [13] H. Butler, Position control in lithographic equipment, *IEEE Control Systems* 31 (5) (2011) 28–47.
- [14] S. Mishra, J. Coaplen, M. Tomizuka, Precision positioning of wafer scanners. segmented iterative learning control for nonrepetitive disturbances, *IEEE Control Systems* 27 (4) (2007) 20–25, <http://dx.doi.org/10.1109/MCS.2007.384130>.
- [15] R. Merry, M. Uyanik, R. van de Molengraft, R. Koops, M. van Veghel, M. Steinbuch, Identification control and hysteresis compensation of a 3 DOF metrological AFM, *Asian Journal of Control* 11 (2) (2009) 130–143.
- [16] Y. Shan, J. Speich, K. Leang, Low-cost IR reflective sensors for submicrolevel position measurement and control, *Mechatronics, Transactions on IEEE/ASME* 13 (6) (2008) 700–709, <http://dx.doi.org/10.1109/TMECH.2008.2005407>.
- [17] S. Parkin, X. Jiang, C. Kaiser, A. Panchula, K. Roche, M. Samant, Magnetically engineered spintronic sensors and memory, *Proceedings of the IEEE* 91 (5) (2003) 661–680, <http://dx.doi.org/10.1109/JPROC.2003.811807>.
- [18] D.R. Sahoo, A. Sebastian, W. Häberle, H. Pozidis, E. Eleftheriou, Scanning probe microscopy based on magnetoresistive sensing, *Nanotechnology* 22 (14) (2011) 145501.
- [19] V. Kartik, A. Sebastian, T. Tuma, A. Pantazi, H. Pozidis, D.R. Sahoo, High-bandwidth nanopositioner with magnetoresistance based position sensing, *Mechatronics* 22 (3) (2012) 295–301.
- [20] T.R. Hicks, P.D. Atherton, Y. Xu, M. McConnell, *The Nanopositioning Book*, Queensgate Instruments Ltd., Berkshire, UK, 1997.
- [21] D.S. Nyce, *Linear Position Sensors. Theory and Application*, John Wiley and Sons, Hoboken, New Jersey, 2004.
- [22] W.C. van Etten, *Introduction to Noise and Random Processes*, John Wiley and Sons, West Sussex, England, 2005.
- [23] R.G. Brown, P.Y.C. Hwang, *Introduction to Random Signals and Applied Kalman Filtering*, John Wiley and Sons, Hoboken, New Jersey, 1997.
- [24] A.J. Fleming, A method for measuring the resolution of nanopositioning systems, *Review of Scientific Instruments* 83 (8) (2012) 086101.
- [25] A.J. Fleming, Measuring picometer nanopositioner resolution, in: Proc. Actuator, Bremen, 2012.
- [26] A.J. Fleming, Estimating the resolution of nanopositioning systems from frequency domain data, in: Proc. IEEE International Conference on Robotics and Automation, St. Paul, MN, 2012, pp. 4786–4791.
- [27] JCGM 200. International Vocabulary of Metrology – Basic and General Concepts and Associated Terms (VIM), 3rd ed., 2008.
- [28] T.-F. Lu, D. Handley, Y.K. Yong, Position control of a 3 DOF compliant micro-motion stage, in: Proc. Control, Automation, Robotics and Vision Conference Vol. 2, 2004, pp. 1274–1278, <http://dx.doi.org/10.1109/ICARCV.2004.1469029>.

- [29] W. Dong, L.N. Sun, Z.J. Du, Design of a precision compliant parallel positioner driven by dual piezoelectric actuators, *Sensors and Actuators A* 135 (1) (2007) 250–256.
- [30] G. Schitter, P.J. Thurner, P.K. Hansma, Design and input-shaping control of a novel scanner for high-speed atomic force microscopy, *Mechatronics* 18 (5–6) (2008) 282–288.
- [31] A.J. Fleming, K.K. Leang, Integrated strain and force feedback for high performance control of piezoelectric actuators, *Sensors and Actuators A* 161 (1–2) (2010) 256–265, <http://dx.doi.org/10.1016/j.sna.2010.04.008>.
- [32] W. Kester, *Sensor Signal Conditioning*, Analog Devices, 2002.
- [33] C.S. Smith, Piezoresistance effect in germanium and silicon, *Physical Review* 94 (1) (1954) 42–49, <http://dx.doi.org/10.1103/PhysRev.94.42>.
- [34] A. Barlian, W.-T. Park, J. Mallon, A. Rastegar, B. Pruitt, Review: semiconductor piezoresistance for microsystems, *Proceedings of the IEEE* 97 (3) (2009) 513–552, <http://dx.doi.org/10.1109/JPROC.2009.2013612>.
- [35] C.M. DiBiao, M.L. Culpepper, Design of a meso-scale six-axis nanopositioner with integrated position sensing, in: *Proc. Fifth Annual International Symposium on Nanomanufacturing*, Singapore, 2008.
- [36] R. Messinger, Q. Aten, T. McLain, L. Howell, Piezoresistive feedback control of a mems thermal actuator, *Journal of Microelectromechanical Systems* 18 (6) (2009) 1267–1278, <http://dx.doi.org/10.1109/JMEMS.2009.2035370>.
- [37] J. Sirohi, I. Chopra, Fundamental understanding of piezoelectric strain sensors, *Journal of Intelligent Material Systems and Structures* 11 (2000) 246–257.
- [38] A.J. Fleming, S.O.R. Moheimani, Control oriented synthesis of high performance piezoelectric shunt impedances for structural vibration control, *IEEE Transactions on Control Systems Technology* 13 (1) (2005) 98–112.
- [39] J. Maess, A.J. Fleming, F. Allgöwer, Simulation of dynamics-coupling in piezoelectric tube scanners by reduced order finite element models, *Review of Scientific Instruments* 79 (2008), <http://dx.doi.org/10.1063/1.2826428>, 015105(1–9).
- [40] A.J. Fleming, A.G. Wills, S.O.R. Moheimani, Sensor fusion for improved control of piezoelectric tube scanners, *IEEE Transactions on Control Systems Technology* 15 (6) (2008) 1265–1278, <http://dx.doi.org/10.1016/j.mechatronics.2007.07.006>.
- [41] A.J. Fleming, Nanopositioning system with force feedback for high-performance tracking and vibration control, *IEEE Transactions on Mechatronics* 15 (3) (2010) 433–447.
- [42] Y.K. Yong, B. Ahmed, S.O.R. Moheimani, Atomic force microscopy with a 12-electrode piezoelectric tube scanner, *Review of Scientific Instruments* 81 (2010), 033701(1–10).
- [43] Y.K. Yong, A.J. Fleming, S.O.R. Moheimani, A novel piezoelectric strain sensor for simultaneous damping and tracking control of a high-speed nanopositioner, *IEEE/ASME Transactions on Mechatronics*, <http://dx.doi.org/10.1109/TMECH.2012.2193895>.
- [44] S.O.R. Moheimani, A.J. Fleming, *Piezoelectric Transducers for Vibration Control and Damping*, Springer-Verlag, London, 2006.
- [45] A. Preumont, *Mechatronics Dynamics of Electromechanical and Piezoelectric Systems*, Springer, Dordrecht, The Netherlands, 2006.
- [46] L.K. Baxter, *Capacitive Sensors: Design and Applications*, IEEE Press, Piscataway, NJ, 1997.
- [47] M. Kim, W. Moon, E. Yoon, K.-R. Lee, A new capacitive displacement sensor with high accuracy and long-range, *Sensors and Actuators A: Physical* 130–131 (14) (2006) 135–141, <http://dx.doi.org/10.1016/j.sna.2005.12.012>.
- [48] L.L. Chu, Y.B. Gianchandani, A micromachined 2d positioner with electrothermal actuation and sub-nanometer capacitive sensing, *Journal of Micromechanics and Microengineering* 13 (2) (2003) 279–285.
- [49] Y.K. Zhu, S.O.R. Moheimani, M.R. Yuce, Simultaneous capacitive and electrothermal position sensing in a micromachined nanopositioner, *Electron Device Letters* 32 (8) (2011) 1146–1148.
- [50] G.T.A. Kovacs, *Micromachined Transducers Sourcebook*, McGraw Hill, New York City, New York, 1998.
- [51] A.A. Kuipers, G.J.M. Krijnen, R.J. Wiegerink, T.S.J. Lammerink, M. Elwenspoek, 2d-finite-element simulations for long-range capacitive position sensor, *Journal of Micromechanics and Microengineering* 13 (4) (2003) S183–S189.
- [52] A.A. Kuipers, G.J.M. Krijnen, R.J. Wiegerink, T.S.J. Lammerink, M. Elwenspoek, A micromachined capacitive incremental position sensor: part 1. analysis and simulations, *Journal of Micromechanics and Microengineering* 16 (6) (2006) S116–S124.
- [53] A.A. Kuipers, G.J.M. Krijnen, R.J. Wiegerink, T.S.J. Lammerink, M. Elwenspoek, A micromachined capacitive incremental position sensor: part 2. experimental assessment, *Journal of Micromechanics and Microengineering* 16 (6) (2006) S125–S134.
- [54] J.-I. Lee, X. Huang, P. Chu, Nanoprecision MEMS capacitive sensor for linear and rotational positioning, *Journal of Microelectromechanical Systems* 18 (3) (2009) 660–670, <http://dx.doi.org/10.1109/JMEMS.2009.2016275>.
- [55] S.-C. Lee, R.D. Peters, Nanoposition sensors with superior linear response to position and unlimited travel ranges, *Review of Scientific Instruments* 80 (4) (2009) 045109, <http://dx.doi.org/10.1063/1.3123407>.
- [56] M.A. Lantz, G.K. Binnig, M. Despont, U. Drechsler, A micro-mechanical thermal displacement sensor with nanometre resolution, *Nanotechnology* 16 (8) (2005) 1089–1094.
- [57] A. Pantazi, A. Sebastian, G. Cherubini, M. Lantz, H. Pozidis, H. Rothuizen, E. Eleftheriou, Control of mems-based scanning-probe data-storage devices, *IEEE Transactions on Control Systems Technology* 15 (5) (2007) 824–841, <http://dx.doi.org/10.1109/TCST.2006.890286>.
- [58] A. Sebastian, A. Pantazi, H. Pozidis, E. Eleftheriou, Nanopositioning for probe-based data storage, *IEEE Control Systems* 28 (4) (2008) 26–35.
- [59] A. Sebastian, D. Wiesmann, Modeling and experimental identification of silicon microheater dynamics: a systems approach, *Journal of Microelectromechanical Systems* 17 (4) (2008) 911–920, <http://dx.doi.org/10.1109/JMEMS.2008.926980>.
- [60] A. Sebastian, A. Pantazi, Nanopositioning with multiple sensors: a case study in data storage, *IEEE Transactions on Control Systems Technology* 20 (2) (2012) 382–394, <http://dx.doi.org/10.1109/TCST.2011.2177982>.
- [60] Y. Zhu, A. Bazaee, S.O.R. Moheimani, M. Yuce, Design modeling and control of a micromachined nanopositioner with integrated electrothermal actuation and sensing, *IEEE/ASME Journal of Microelectromechanical Systems* 20 (3) (2011) 711–719.
- [62] J. Fraden, *Handbook of Modern Sensors: Physics Designs and Applications*, Springer, New York, 2004.
- [63] S. Fericean, R. Droxler, New noncontacting inductive analog proximity and inductive linear displacement sensors for industrial automation, *IEEE Sensors Journal* 7 (11) (2007) 1538–1545.
- [64] S.D. Roach, Designing and building an eddy current position sensor, *Sensors* September (1998).
- [65] Q. Li, F. Ding, Novel displacement eddy current sensor with temperature compensation for electrohydraulic valves, *Sensors and Actuators A: Physical* 122 (1) (2005) 83–87, <http://dx.doi.org/10.1016/j.sna.2005.04.008>.
- [66] R. Proksch, J. Cleveland, D. Bocek, US Patent 7,262,592 linear variable differential transformers for high precision position measurements (2007).
- [67] P. Hariharan, *Basics of Interferometry*, 2nd ed., Academic Press, Burlington, MA, 2007.
- [68] J.N. Dukes, G.B. Gordon, A two-hundred-foot yardstick with graduations every microinch, *Hewlett-Packard Journal* 21 (2) (1970) 2–8.
- [69] G.E. Sommargren, A new laser measurement system for precision metrology, in: *Proc. Precision Engineering Conference*, Dallas, TX, 1986.
- [70] K. Karrai, P. Braun, Miniature long-range laser displacement sensor, in: *Proc. Actuator, Bremen, Germany*, 2010, pp. 285–288.
- [71] A. Khat, F. Lamarque, C. Preme, P. Pouille, M. Leester-Schädel, S. Büttgenbach, Two-dimension fiber optic sensor for high-resolution and long-range linear measurements, *Sensors and Actuators A: Physical* 158 (1) (2010) 43–50, <http://dx.doi.org/10.1016/j.sna.2009.12.029>.
- [72] R.S. Sirohi, *Optical Methods of Measurement – Wholefield Techniques*, CRC Press, Taylor and Francis Group, Boca Raton, FL, 2009.
- [73] FASTRACK high-accuracy linear encoder scale system. data sheet I-9517-9356-01-b. Online: www.renishaw.com
- [74] Heidenhain exposed linear encoders. Online: www.heidenhain.com
- [75] J.-Y. Lee, H.-Y. Chen, C.-C. Hsu, C.-C. Wu, Optical heterodyne grating interferometry for displacement measurement with subnanometric resolution, *Sensors and Actuators A: Physical* 137 (1) (2007) 185–191, <http://dx.doi.org/10.1016/j.sna.2007.02.017>.
- [76] B.M. Chen, T.H. Lee, K. Peng, V. Venkatarmanan, Hard Disk Drive Servo System, Springer, London, 2006.
- [77] Y. Michellod, P. Mullhaupt, D. Gillet, Strategy for the control of a dual-stage nano-positioning system with a single metrology, in: *Proc. Robotics, Automation and Mechatronics*, 2006, pp. 1–8, <http://dx.doi.org/10.1109/RAMECH.2006.252649>.
- [78] L. Chassigne, M. Wakim, S. Xu, S. Topçu, P. Ruaux, P. Juncar, Y. Alayli, A 2d nano-positioning system with sub-nanometric repeatability over the millimetre displacement range, *Measurement Science and Technology* 18 (11) (2007) 3267–3272.
- [79] A.J. Fleming, Dual-stage vertical feedback for high speed-scanning probe microscopy, *IEEE Transactions on Control Systems Technology* 19 (1) (2011) 156–165, <http://dx.doi.org/10.1109/TCST.2010.2040282>.
- [80] J. Zheng, A. Salton, M. Fu, Design and control of a rotary dual-stage actuator positioning system, *Mechatronics* 21 (6) (2011) 1003–1012, <http://dx.doi.org/10.1016/j.mechatronics.2011.03.011>.
- [81] C.-L. Chu, S.-H. Fan, A novel long-travel piezoelectric-driven linear nanopositioning stage, *Precision Engineering* 30 (1) (2006) 85–95, <http://dx.doi.org/10.1016/j.precisioneng.2005.05.002>.
- [82] R. Merry, M. Maassen, M. van de Molengraft, N. van de Wouw, M. Steinbuch, Modeling and waveform optimization of a nano-motion piezo stage, *IEEE/ASME Transactions on Mechatronics* 16 (4) (2011) 615–626, <http://dx.doi.org/10.1109/TMECH.2010.2050209>.
- [83] G. Borionetti, A. Bazzalia, R. Orizio, Atomic force microscopy: a powerful tool for surface defect and morphology inspection in semiconductor industry, *The European Physical Journal Applied Physics* 27 (1–3) (2004) 101–106.
- [84] A. Humphris, M. McConnell, D. Catto, A high-speed atomic force microscope capable of video-rate imaging, in: *Microscopy and Analysis – SPM Supplement*, 2006, pp. 29–31.
- [85] A. Ferreira, C. Mavroidis, Virtual reality and haptics for nanorobotics, *IEEE Robotics and Automation Magazine* 13 (3) (2006) 78–92.
- [86] G.E. Fantner, G. Schitter, J.H. Kindt, T. Ivanov, K. Ivanova, R. Patel, N. Holten-Andersen, J. Adams, P.J. Thurner, I.W. Rangelow, P.K. Hansma, Components for high speed atomic force microscopy, *Ultramicroscopy* 106 (2–3) (2006) 881–887.

- [87] M. Kobayashi, K. Sumitomo, K. Torimitsu, Real-time imaging of DNA-streptavidin complex formation in solution using a high-speed atomic force microscope, *Ultramicroscopy* 107 (2-3) (2007) 184–190.
- [88] G. Schitter, K.J. Åström, B.E. DeMartini, P.J. Thurner, K.L. Turner, P.K. Hansma, Design and modeling of a high-speed AFM-scanner, *IEEE Transactions on Control Systems Technology* 15 (5) (2007) 906–915.
- [89] L.M. Picco, L. Bozec, A. Ulcinas, D.J. Engledew, M. Antognozzi, M. Horton, M.J. Miles, Breaking the speed limit with atomic force microscopy, *Nanotechnology* 18 (4) (2007), 044030(1–4).
- [90] T. Ando, T. Uchihashi, T. Fukuma, High-speed atomic force microscopy for nano-visualization of dynamic biomolecular processes, *Progress in Surface Science* 83 (7-9) (2008) 337–437.
- [91] E. Guliyev, T. Michels, B. Volland, T. Ivanov, M. Hofer, I. Rangelow, High speed quasi-monolithic silicon/piezostack spm scanning stage, *Microelectronic Engineering* 98 (0) (2012) 520–523, <http://dx.doi.org/10.1016/j.mee.2012.07.059>.

Biography

Andrew J. Fleming graduated from The University of Newcastle, Australia (Callaghan campus) with a Bachelor of Electrical Engineering in 2000 and Ph.D. in 2004. He is presently an Australian Research Fellow and Senior Lecturer at the School of Electrical Engineering and Computer Science, The University of Newcastle, Australia. His research includes nanopositioning, high-speed scanning probe microscopy, micro-cantilever sensors, metrological position sensors, and optical nanofabrication. Academic awards include the University of Newcastle Vice-Chancellors Award for Researcher of the Year and the IEEE Control Systems Society Outstanding Paper Award for research published in the *IEEE Transactions on Control Systems Technology*. Dr. Fleming is the co-author of three books, several patent applications and more than 100 Journal and Conference papers.