Scaling electromechanical sensors down to the nanoscale

Michael A. Cullinana, Robert M. Panasb, Christopher M. DiBiasioc, Martin L. Culpepperac

* Massachusetts Institute of Technology, Department of Mechanical Engineering, Cambridge, MA 02139, United States
b National Institute of Technology, Intelligent Systems Division, 100 Bureau Drive, Gaithersburg, MD 20899, United States
c C.S. Draper Laboratory, 555 Technology Sq, Cambridge, MA 02139, United States

Abstract

Traditional MEMS sensing systems do not scale down well to the nanoscale due to resolution and fabrication limitations. Therefore, new sensing systems need to be developed in order to meet the range and resolution requirements of nanoscale mechanical systems. Several nanoscale mechanical sensing systems have emerged that take advantage of nanoscale phenomena to improve the quality of nanoscale sensors. In this paper, we will discuss some of the fundamental limitations in scaling mechanical sensors down to the nanoscale and some of the emerging technologies for nanoscale sensing.

© 2012 Elsevier B.V. All rights reserved.

1. Introduction

Nanoscale mechanical devices offer the potential to overcome many of the speed and sensitivity limitations of macro- and microscale devices due to their small size and low mass. However, it is difficult to measure and control the motion of these devices.
because of the limited dynamic range of nanoscale sensors. Therefore, one of the greatest challenges for mechanical devices as they move toward the nanoscale is the incorporation of fine resolution sensors into these devices.

The incorporation of sensors into nanomechanical devices is critical in numerous applications because the ability to measure and control forces and displacements in mechanical systems can greatly improve the accuracy, disturbance rejection and bandwidth of these systems. However, sensors are difficult to integrate into nanoscale devices because they are either too large to fit into nanoscale systems or are not sensitive enough to measure nanoscale motions over large ranges. For example, traditional MEMS sensing techniques are not adequate for many nanoscale sensing applications because most MEMS sensing techniques suffer from poor resolution or become increasingly difficult to fabricate when scaled down to the nanoscale. Therefore, new sensors must be developed in order to meet the requirements of nanoscale sensing. In this paper we will examine possibilities for scaling down traditional MEMS sensing techniques to the nanoscale and will discuss new measurement techniques only available at the nanoscale.

1.1. NEMS

Nanoelectromechanical systems (NEMS) are critical for the development of medical diagnostics, displays, and data storage. Because of their size, NEMS can outperform their macro-scale counterparts in terms of speed and resolution by trading off payload size and range. For example, NEMS devices can help to enable the testing and manipulation of individual biomolecules, which is not possible at the macroscale. In addition, NEMS can operate where larger machines will generally not fit, such as inside the human body or in other instrumentation. This is important because it can enable real-time diagnostics of biological, mechanical and chemical systems.

Another major advantage of NEMS devices is that they eliminate the scale mismatch problems that occur when macro- and microscale devices are used to perform nano-scale tasks. For example, motions in NEMS devices caused by thermal expansion or mechanical vibrations are typically in the sub-nanometer range. However, similar phenomena in macroscale devices can cause micron-scale motions that can damage nanoscale features. Elimination of this scale mismatch problem could allow nano-scale devices to have a large impact on a number of fields where interactions with nanoscale structures are important, such biology, chemistry, physics, and nanofluidics, by allowing machines to better interact with individual molecules.

2. Scaling MEMS sensors down to the nanoscale

The first step in evaluating what types of sensors may be useful at the nanoscale is to look at the types of sensors that have been used at the microscale. In this section we will review several types of microscale sensors and analyze their underlying device physics. Based on the device physics, we can create physical scaling laws which we can use to determine how well various microscale sensors scale down to the nanoscale. The results of this scaling analysis are only summarized here; please refer to other previously published works [1-6] for more comprehensive reviews of the state-of-the-art in MEMS sensing technologies.

Nanoscale mechanical devices typically rely on flexure beams to guide their motion and to set the stiffness of the system. Based on the stiffness of the flexure it is generally possible to trade range for resolution. For example, compliant flexures are generally good for fine resolution force measurements but tend to fail at relatively low load levels. Therefore, it is the ratio of range to resolution, known

![Fig. 1. Comparison of sensor dynamic ranges versus size for MEMS sensing.](Image)

as dynamic range, that is the best metric of nanoscale sensor performance. Throughout this paper the viability of different sensor systems is judged on the basis of the dynamic range of the sensor. Typically, nanoscale sensing applications require a minimum dynamic range at least a 40 dB (range to resolution ratio of 100:1). However, dynamic ranges of greater than 80 dB are desirable for many positioning [7] and force-sensing applications [8].

2.1. Overview of MEMS sensors

The most common sensors used to measure motions in MEMS devices are optical, capacitive, inductive, Hall effect, piezoelectric, piezoresistive and tunneling sensors. Unfortunately, the dynamic range of sensors of these types tends to scale with the sensor footprint as shown in Fig. 1. Therefore, when these sensors are scaled down to the nanoscale they become ineffective at making the fine precision motion measurements required in many nanoscale systems. As we will show in this paper, however, by incorporating nanoscale materials and new measurement techniques into NEMS devices it should be possible to achieve dynamic ranges of greater than 80 dB in nanoscale devices.

In this paper we start by investigating how traditional MEMS devices scale down to the nanoscale in order to determine their utility for NEMS devices. The MEMS sensors examined in this paper are compared in Table 1 based on the general categories of resolution (Res), dynamic range (DR), sensor footprint (Size), bandwidth (BW), sensitivity to environment (Env), and sensitivity to temperature fluctuations around room temperature (T).

All of the values presented in Table 1 are approximate due to the tradeoffs between performance characteristics in each of the sensors. This makes it especially difficult to evaluate the “best case” performance of each sensor technology in the literature, as the results presented typically maximize only one sensor attribute while trading off others. For example, by increasing the size of the sensor and decreasing the bandwidth it is often possible to increase both the resolution and the dynamic range of the sensor.

In building this table, a high performance resolution was considered for each sensor and used as a driver in determining the performance parameters listed in the other columns. Resolution and dynamic range for each sensor type are calculated using the equations presented in the following sections. The footprint of the sensors is estimated based on the size necessary to achieve the resolution presented in the table, however, the size of additional electronics are not included. Smaller sensors can typically be made, but they are unable to meet the resolution and dynamic range shown in Table 1. Similarly, the bandwidth in Table 1 is the maximum bandwidth at which the displacement resolution in Table 1 can be achieved. The environmental sensitivity column in Table 1 lists outside factors that may affect the performance of the sensor,
such as electromagnetic noise, humidity, and pressure variations, but excluding temperature, which is presented separately. If “low” is shown, it means that the sensor is not greatly affected by environmental factors other than temperature. Temperature sensitivity is calculated as the as the percent change in output due to a degree change in temperature at room temperature.

Based on the information presented in Table 1, interferometry offers the highest performance in terms of dynamic range if cost and system complexity are not concerns. If cost is a concern but size is not, then capacitive sensors may be able to achieve the highest dynamic ranges since their dynamic range scales with their area. However, for nanoscale sensing systems, where size is a major restriction, piezoresistive and tunneling sensors are the most effective sensing systems, as shown in Fig. 1. Therefore, we will start by investigating these types of sensors.

### 2.2. Piezoresistive sensors

Piezoresistive sensors are commonly used to measure displacements on the micro- and nanoscale due to their small size [3,11,12]. The electrical resistance of piezoresistive sensors changes as strain is applied to the sensor and is typically measured using a Wheatstone bridge. This measurement of the strain in the sensor, and thus the strain in the beam, makes it possible to determine the displacement of the beam. A typical application of a piezoresistive sensor is shown in Fig. 2, where six sets of polysilicon-based piezoresistors are used to measure the in-plane and out-of-plane motions of a six-axis mesoscale nanopositioner.

The major noise sources in piezoresistive sensors are flicker noise caused by conductance fluctuations and Johnson noise caused by thermal agitation of electrons in the conductor [13,14]. The total noise is the sum of the Johnson, flicker, and amplifier input noise variance as seen in Eq. (1) where $k_B$ is the Boltzmann constant, $T$ is the temperature, $R$ is the bridge resistance, $f$ is the frequency in Hertz, $\alpha$ is the Hooge constant, $N$ is the number of charge carriers, $S_{\text{Vai}}$ is the voltage noise spectral density of the amplifier input, and $V_s$ is the voltage across the resistors [15].

$$\sigma_v = \sqrt{4k_BTR(f_{\text{max}} - f_{\text{min}}) + \frac{V_s^2}{4N} \frac{\alpha}{\epsilon} \ln \left( \frac{f_{\text{max}}}{f_{\text{min}}} \right) + S_{\text{Vai}}(f_{\text{max}} - f_{\text{min}})}$$

The dynamic range of a piezoresistive sensor is given by Eq. (2) where $\sigma_v$ and $E$ are the yield stress and Young’s Modulus of the flexure material, respectively, $\eta$ is the safety factor (usually 3), $N_e$ is the bridge strain number (e.g. 1/4 for quarter bridge), $V_s$ is the supply voltage, and $G_F$ is the gauge factor.

$$\text{DR} = \frac{\sigma_v V_s N_e G_F}{\eta \sigma_v}$$

(2)

For doped silicon piezoresistors, the dynamic range can be on the order of 120 dB [16], but both the resistance and gauge factors of piezoresistors are sensitive to thermal variations. Therefore, it is critical to either include thermal compensation, such as span temperature compensation [17] and a thermally symmetric bridge design [12], or to precisely control the temperature of the environment in order to make accurate measurements. As piezoresistive sensors are scaled down to the nanoscale, flicker noise becomes the dominate noise source and the sensor’s dynamic range starts to scale with the inverse of the square root of the resistor volume. This typically limits the performance of nanoscale piezoresistors to less than 60 dB dynamic range.

### 2.3. Tunneling sensors

Tunneling sensors have shown great success in measuring subatomic displacements but have been limited at the macroscale due to their lack of range. Tunneling sensors generally consist of a sharp tip placed within a nanometer of a moving surface. As the gap between the tip and the surfaces changes, the tunneling current changes as given by Eq. (3), where $\rho_0(E_F)$ is the local electron density of states, $E_F$ is the Fermi energy, $V$ is the DC bias voltage, $\kappa$ is decay constant for the electron wave-function within the gap, and $d$ is the gap size [2].

$$i = \rho_0(E_F) Ve^{-\kappa d}$$

(3)

This results in a high displacement sensitivity but small range, due to the exponential decline in tunneling current as the gap size increases. For example, an electronic circuit with a 1% variation in a 1 nA current from a 100 MΩ source is capable of detecting displacements as small as 0.3 pm but has a range of less than 1 nm [18]. In order to overcome these range limitations, feedback...
control systems are often used to maintain a fixed gap distance over a range of displacements [19]. These feedback sensors, however, are not able to maintain the high displacement resolution of the tunneling sensor over the full range of the device, since large motions must be measured by an external sensor on the feedback platform.

While non-feedback tunneling sensors are limited at the macro- and microscales, they may have some utility at the nanoscale. As devices are scaled down to the nanoscale, the necessary range of these devices also tends to decrease as well. Therefore, the limited range of the tunneling sensors may not be that limiting in nanoscale sensing. Non-feedback tunneling sensors can also easily be scaled down to the nanoscale, since the tunneling occurs between the nearest pair of atoms on each side of the tip-surface gap [19]. Therefore, the tunneling sensor does not need to be any larger than a few atoms and the sensor range and resolution do not scale with the sensor size. This makes tunneling sensors ideal for nanoscale sensing applications that require high resolution but do not have a large range.

2.4. Capacitance sensors

Capacitance sensors are widely used at the macro-, meso-, and microscale for measuring displacements on the order of nanometers [20–22]. There are two architectures for capacitive sensors at the MEMS scale [14]: comb drive fingers and parallel plate capacitors. For example, in Fig. 3, parallel plate capacitors are used to measure the out-of-plane motions and the comb drive fingers are used to measure the in-plane motions of a 6 axis force-torque sensor.

Comb drives produce a linear change in capacitance in response to a change in displacement, while parallel plate capacitors produce a nonlinear change. The dynamic range, DR, of a parallel plate capacitor can be found to scale roughly as Eq. (4) [14].

$$DR \approx \frac{\varepsilon \cdot A}{g \cdot \Delta C}$$

(4)

In Eq. (4), $g$ is the initial gap size, $\varepsilon$ is the permittivity of the medium, $A$ is the area of one of the parallel plates, and $\Delta C$ is the change in capacitance at maximum range due to a change of $\delta$ in position. At large ranges (gap sizes), the sensitivity, $\Delta C/\delta$, of the sensor decreases. The main noise source in capacitive sensors is stray capacitance in the environment, which can vary from 1 fF [21] to 180 pF [24] depending on system shielding. Parallel plate capacitive sensors do not scale down well to the nanoscale because the dynamic range of these sensors is dependent on the sensor area. For example, sensor areas of 10's mm$^2$ to 10's cm$^2$ are required in order to achieve an 80 dB dynamic range when the initial gap size is $\sim$100 nm.

The dynamic range in a comb drive sensor can be found to scale roughly as Eq. (5) [14].

$$DR \approx \frac{(N - 1) \cdot L \cdot t}{g \cdot \Delta C}$$

(5)

In Eq. (5), $N$ is the total number of comb fingers, $L$ is the overlapping length of the comb fingers, and $t$ is the device thickness. Eq. (5) shows that as the range of the sensor is increased, the sensitivity, $\Delta C/\delta$, is not changed. In order to achieve an 80 dB dynamic range, a large number of fingers per axis are necessary ($10^2$–$10^8$). Typically, in MEMS devices, the gap size is on the order of 100 nm to 1 $\mu$m and the fingers are on the order of 1–10 $\mu$m wide. In addition, the fingers are generally an order of magnitude longer than the overlapping length on each axis. These sensors, therefore, need to have a footprint of approximately 1 cm $\times$ 100 $\mu$m in order to achieve an 80 dB dynamic range. Again, because the dynamic range of these sensors scales with sensor area, comb drive sensors become impractical at the nanoscale.

2.5. Hall effect sensors

Hall effect sensors have also been used to measure sub-micron displacements in MEMS devices [4,25]. Eq. (6) shows the sensitivity of the Hall effect sensor output voltage, $V_H$, to motion of the sensor, $\delta$. This sensitivity scales with an applied magnetic field, $B(x)$, which varies as a function of the distance, $x$, from the magnet. In Eq. (6), $R_H$ is the Hall coefficient, $I_s$ is the energizing current, and $t$ is the sensor thickness [26].

$$\partial V_H = \frac{R_H I_s}{t} \cdot \frac{\partial B(x)}{\partial x} \cdot \delta$$

(6)

The main sources of noise in a Hall effect sensor are Johnson noise and 1/f noise [25,26], as described by Eq. (1). Typically these noise sources range from 100's nV to $\mu$V [26]. The maximum Hall coefficient in doped silicon is approximately $1.4 \times 10^{-3}$ m$^2$/C for a dopant concentration of $4.5 \times 10^{15}$ dopants/cm$^3$ [26]. For a MEMS sensor with a thickness of approximately 100 $\mu$m and a typical current of roughly 1–10 mA the sensor minimum detectable field is approximately 7 $\mu$T [9,25]. The dynamic range of this sensor can be approximated by Eq. (7), assuming linearity of the field over the range, $r$, of the sensor.

$$DR \approx \frac{R_H I_s f_s}{\sigma_v t} \cdot \frac{\partial B(r)}{\partial x}$$

(7)

In Eq. (7), $r_s$ is the separation of the sensor from the field source, $\sigma_v$ is the electronic noise in the system, and $f_s$ is the fraction of the sensor to magnet separation over which the field gradient falls within the bounds required for the sensor to function effectively. Note that there may be significant variation in the magnetic field.
gradient over the range of the sensor. A common bound is that a minimum sensitivity as calculated through Eq. (6) must be met over the whole sensor range, which means that the gradient will have a lower bound. Magnetic fields used for Hall effect sensing typically decay at a high (2nd, 3rd) power, and thus the gradient condition is met at all displacements up to a maximum, \( r_s \), so that \( f_r \) approaches unity [9]. The main challenge in implementing Hall effect sensors in nanoscale systems is minimizing the nonlinear stray magnetic field from the environment as a positioning stage moves through a work volume. If these fields exceed the minimum detectable magnetic field they can adversely affect the sensor accuracy and/or resolution.

### 2.6. Piezoelectric sensors

Piezoelectric sensors are commonly used to measure forces and strains corresponding to micro–scale motion, particularly those of a dynamic nature, such as vibrations, accelerations and oscillations [1,27–29]. These sensors operate through the re-orientation of the crystalline bulk dipole density (polarization vector) in response to stress, which acts to produce an effective charge density on the crystal face [29]. A common model of a piezoelectric sensor is a capacitor in parallel with a current source whose output is the derivative of the charge produced [1,27]. This charge output is often read by an amplifier which outputs a voltage proportional to the captured charge [28]. Piezoelectric sensors are ‘active’ sensors that do not require a power source to operate, which reduces heating issues compared to ‘passive’ piezoresistive sensors [29]. A typical application of piezoelectric sensors as accelerometers is shown in Fig. 4, where piezoelectric sensors are used to measure acceleration. The device in Fig. 4A is a clamped-clamped beam micro-machined out of silicon, and a \( 1 \) \( \mu \)m thick layer of ZnO has been sputtered on to the beam to act as the sensing element [30]. The device in Fig. 4B is an accelerometer surface micro-machined out of 500 nm thick ZnO, with each set of legs acting at the piezoelectric sensing circuit [31]. The device in Fig. 4C is an accelerometer that uses interdigitated electrodes for higher sensitivity [1].

The main sources of noise in a piezoelectric sensor are: (i) the current noise from the amplifier, (ii) the Johnson noise from the resistance in the amplifier feedback circuit, and (iii) the Johnson noise from the equivalent leakage path resistance (in parallel with the capacitance and current source) in the sensor [27,29]. Current noise variance is transformed into charge noise using a \( (2\pi f)^2 \) term, which heavily biases the PSD toward the lower frequency. The total noise, \( \sigma_Q \), is shown in Eq. (9) where \( S_{\text{int}} \) is the current spectral noise density of the amplifier input, \( \rho_c \) and \( \tau_c \) are the resistivity and thickness of the piezoelectric material film, respectively, and \( V_{\text{max}} \) is the charge amplifier voltage output limit (usually 10 V).

\[
\sigma_Q = \sqrt{\frac{S_{\text{int}}}{4\pi f_{\text{min}}} + \frac{k_B T \rho_c}{\rho_s \tau_c f_{\text{min}}} + \frac{2k_B T Q_{\text{max}}}{\pi V_{\text{max}}}} \tag{9}
\]

The dynamic range of a piezoelectric sensor is shown in Eq. (10).

\[
\text{DR} = \frac{Q_{\text{max}}}{\sigma_Q} \tag{10}
\]

A PZT-5H piezoelectric MEMS sensor of about 100 \( \mu \)m on each side and 1 \( \mu \)m thick, using a high performance amplifier (AD743 [33]) with high-pass cutoff frequency at 0.001 Hz, could reach a dynamic range of roughly 80 dB. This bandwidth cutoff frequency is sufficient for \( \approx 1 \) s measurements to be carried out with \( \approx 1\% \) error in a DC value. The dynamic range for piezoelectric sensors scales with the area of the sensor at small sizes, as this limits the charge produced. The amplifier charge noise dominates in these situations.

Piezoelectric sensors function as high-pass filters due to the finite insulation resistance of capacitors and other circuitry [1,27–29]. This cutoff frequency is generally determined by the RC decay time in the amplifier feedback circuit [28,29]. The frequency can be made as low as 10–6 Hz for meso- to macroscale sensors (>1 mm²), but scales inversely with sensor area below this threshold due to limited capacitor insulation resistance [29]. Several other low frequency errors such as drift and dielectric capacitance act to reduce accuracy for signals around these frequencies [29]. Achieving DC measurement errors of <1% requires measurement periods corresponding to frequencies at least 3 decades above the high-pass cutoff frequency, effectively inhibiting the abilities of microscale sensors to measure quasi-statically.

### 2.7. Inductance sensors

There are two types of inductance sensors that are commonly used to measure displacements, linear variable differential transformers (LVDT) and eddy current proximity sensors [9]. In order for LVDTs to achieve nanometer resolution, the motion in the other orthogonal axes must be minimized. Therefore, multi-degree-of-freedom devices are difficult to implement with LVDT sensing.

In eddy current proximity sensors, an AC current is driven over a reference coil to generate a sensing magnetic field. This
field induces eddy currents in the target, which itself produces a magnetic field in opposition to the sensing field. The opposing field, which is a function of the coil-to-target separation, alters the effective inductance of the reference coil. The coil inductance can be measured by the sensing electronics to determine the coil to target separation [9]. MEMS eddy current sensors have been shown to have resolutions of 100’s of nanometers [34–36] when used with metallic targets. MEMS eddy current sensors have also shown to be sensitive to motion in the axes orthogonal to the measurement axis [34–36]. Due to their reliance on magnetic fields for accurate location measurement, eddy current sensors are sensitive to magnetic interference. Eddy current sensors become hard to fabricate at the nanoscale and the maximum current in the coil decreases significantly, which limits the sensor resolution. The inductance of the eddy current sensor scales down with the overall size of the device while the resistance rises at smaller scales. This leads to difficulty in capturing high precision measurements of the changing inductance for microscale eddy-current sensors. Therefore, eddy current sensors are not considered practical for nanoscale sensing applications.

2.8. Interferometry

Optical interferometry is mainly used in high precision equipment, especially in cases that require both large range and high resolution displacement measurement [9,37–41]. The basic form for this method of sensing relies on splitting coherent light into two beams, arranging the beam paths such that one path length is dependent on the displacement of interest, and then recombining the beams to generate an interference pattern [9,41–43]. This pattern will vary with the displacement measurement, and can be measured to infer the change in the path length. A common interferometry setup is the Michelson-type heterodyne interferometer [9,39,41,42], which passes lasers of different frequencies along the measurement and reference optical paths, combining them at the end to generate a signal at the beat frequency determined by the frequency difference between the two beams. The lower frequency beat signal allows for greater range and is more easily measured by the electronics, resulting in reduced noise [9,42]. This type of interferometry setup is shown in Fig. 5, where the interferometer is used to measure the motion of a 6 degree-of-freedom stage. As can be seen in the figure, the footprint of the sensor is quite large (10’s to 100’s of cm²), due to the required optics and laser source. Therefore, it is difficult to scale this type of sensor down to the nanoscale.

Optical heterodyne interferometers are sensitive to environmental conditions such as temperature and pressure, both of which change the optical properties of the medium through which the beam travels [9,41,43]. The range of measurement of an optical heterodyne interferometer is effectively decoupled from its resolution because the relative phase measurement cyclically shifts only through $0–2\pi$ as the displacement value changes. Interferometers are generally limited in range by either a scaling of error due to environmental sensitivity or the spatial coherence of the interferometer [9,41]. High performance interferometers are capable of reaching ranges of multiple meters [9,39,41,44]. Interferometry position sensors are only capable of accurate displacement measurement up to a maximum velocity on the scale of a 1–3 m/s [9,41], however, above which the fringe count rate exceeds the ability of the electronics to perform accurate counting. The displacement resolution, $\delta$, of an optical heterodyne interferometer can be written as a function of the phase resolution capability of the electronics, $\delta_{\min}$, lasing wavelength, $\lambda$, and of number of reflections of the beam through the measurement arm, $N$, as [9,41,44], as in Eq. (11).

$$\delta_{\min} = \frac{\Delta \phi_{\min} \lambda}{2\pi N}$$  \hspace{1cm} (11)

State-of-the-art interferometry systems are capable of resolving down to roughly 1 part in 5000 of the lasing wavelength [9,39–41,44], for which a He:Ne laser is a common choice [9] with a wavelength of around 630 nm. The resulting resolution is roughly 130 pm. Multi-axis interferometry set-ups generally utilize a single high precision large lasing source and multiplex this to allow multiple degree-of-freedom measurements [39–41]. This requires a large optical layout on the order of 10’s of cm to 1 m on a side [38,40,41]. The complexity of the optical set-up increases rapidly as the number of measured axes is increased, resulting in increased system environmental sensitivity. The cost for a single axis optical heterodyne interferometer is in the range of $10,000, and can rise upwards of $100,000 for a complex multi-axis system [9]. Therefore, interferometry is much more expensive than integrated sensing on a MEMS chip where the cost per sensing axis is generally less than $1000. A state-of-the-art optical heterodyne laser interferometer is capable of reaching and exceeding 180 dB dynamic range. The main issue prohibiting the integration of optical interferometers into NEMS is the lack of a true nano- to microscale stabilized and coherent laser source. Until significant advances are made in chip-scale laser sources the size of interferometry set-ups limit their utility at the nanoscale.

2.9. Vernier gauges

The use of integrated Vernier gauges in MEMS devices to measure displacement is quite common [45–47], especially in bench-level prototypes. Vernier gauges consist of two sets of scales placed on the moving object, and an inertially grounded reference. Each scale consists of a set of small lines/ticks with a well defined pitch. By measuring the relative position of the scales it is possible to measure the displacement of a moving object with respect to the grounded reference.

The scales are often micro-machined onto the surface of the MEMS device using standard lithographic techniques [45], though these features can also be written with advanced fabrication equipment such as a focused ion beam (FIB) [48]. Typically the device under test is operated in a scanning electron microscope (SEM) and the movement of the Vernier gauge is measured by the SEM operator. This precludes the use of Vernier gauges as a closed

![Fig. 5. Large range Michelson type heterodyne interferometer for 6-DOF position measurement [41].](image)
loop measurement system. Recent research has been conducted in order to reduce the pitch of the Vernier scales from the micron to the nanometer level [49]. This reduced pitch may suggest that Vernier gauges could be candidates for use in NEMS, but two main challenges exist with regards to imaging. The first is that measuring ticks only a few nanometers wide with nanometer level pitch pushes the boundary of what may be imaged using an SEM. Second, SEMs that are capable of imaging down to this level are much larger than those with less resolution. Clearly the main challenge for integrating Vernier gauges into NEMS is shrinking the imaging component of the system.

3. Emerging nanoscale sensing technologies

Most traditional MEMS sensor systems do not scale down well to the nanoscale due to noise, sensitivity and fabrication limits, as discussed in the previous section. Therefore, new types of specialized sensors and sensing systems need to be developed in order to meet the needs of nanoscale devices. In this section we will present several emerging sensing technologies specifically designed to work at the nanoscale.

3.1. CNT-based piezoresistors

Carbon nanotube (CNT)-based piezoresistive sensors are able to produce large dynamic ranges for small sensor footprints due to their large gauge factors and inherent nanoscale size. Therefore, CNT-based piezoresistors become the only viable option for high dynamic range sensing as piezoresistive sensors are scaled down to the nanoscale. For example, as we will show in this section, CNTs are the only piezoresistive sensor capable of producing a dynamic range of greater than 60 dB when the sensor footprint is below 1 \( \mu \text{m}^2 \).

3.1.1. Review of CNT-based piezoresistive sensors

The first experiments that showed the potential of using CNTs as high quality strain sensors involved using an AFM tip to push down on a suspended CNT. This method was used by Tombler et al. to show that the conductance of a CNT can be changed by up to two orders of magnitude when strain is applied to the CNT [50]. Similarly, using this method Minot et al. showed that the band structure of a CNT can be altered by the application of strain to the CNT [51]. However, simulations have shown that these results may have been due to local deformations in the CNT structure around the AFM tip, as opposed to uniform strain in the CNT [52]. More recently, experiments that have uniformly strained the entire CNT have shown that the CNT gauge factor could be as high as 2900 [53]. The CNT gauge factor has also been shown to vary widely based on the electrical structure of the CNT [54]. For example, Grow et al. showed that the CNT gauge factor could be either positive or negative based on the CNT structure [55]. This result offers the interesting implication that a full Wheatstone bridge could be formed at one location on a MEMS flexure using CNTs that increase and other CNTs that decrease resistance as the flexure is strained.

Theoretical modeling and simulations have been used to understand the link between CNT structure and gauge factor. These models and simulations typically are used to estimate the change in the band-gap of different types of CNTs. The band-gap changes are then related back to the CNT resistance to obtain an estimate of the gauge factor. For example, Chen et al. showed that a 1% strain of a (12,0) CNT should result in a 6.4% decrease in resistance [56]. Similarly, Yang et al. [57] used tight-binding models to show that CNT band-gap could increase or decrease depending on the chirality of the CNT, confirming the qualitative result given by Grow et al. Unfortunately, there has been little quantitative work on linking specific CNT geometries to specific gauge factors. Also, there has been little theoretical or experimental work done to understand how interactions between the CNT and the substrate/electrodes affect the properties of the CNT-based strain sensor.

Several prototype devices have been fabricated using CNTs as strain sensors. The most common devices use films of randomly oriented films of CNTs as the sensing element. These films are popular due to their ease of assembly and large size. These properties allow these CNT-based film sensors to be integrated into many macroscale sensor systems [58–61]. CNT-based sensors also offer the advantage that they can easily be integrated into rubber and polymer based composites [62–68]. Such composites are commonly used in microfluidic systems for structural health monitoring, artificial skin, and acoustic wave diagnostics. However, these CNT-based piezoresistive films tend to suffer from low gauge factors due to the random orientation of the CNTs in the films and the poor transmittance of the strain in the substrate to the CNTs in the films [69].

There have been several devices fabricated using individual or a small number of CNTs as the sensing elements. For example, CNT-based piezoresistive sensors have been used to measure the force applied to mesoscale beams [71–75] and the strain applied to flexible substrates [76,77]. Strain is applied to the CNTs in these devices through the bending or stretching of the substrate. MEMS pressure sensors have also been fabricated using CNT piezoresistors as the sensing element [78]. These CNT-based pressure sensors are capable of pressure resolutions of about 1 psi. These devices are fabricated by dispersing a random assortment of SWCNTs on a surface and then selecting a few to pattern the MEMS device around, as seen in Fig. 6. When pressure is applied to the substrate, the CNTs are strained and their resistance changes, making it possible to measure the pressure. Force and displacement sensors have also been fabricated by suspending individual CNTs between electrodes and attaching a beam to the center of the CNT [79]. The CNT is stretched when the beam is deflected, causing the resistance to change. Such devices are capable of sub-nano-Newton force resolution. Several CNT-based piezoresistive devices have been demonstrated where compression [80] or torsion [81] is applied to the CNT.

3.1.2. Scaling of CNT-based piezoresistors

Flicker noise becomes the dominate noise source as piezoresistive sensors are scaled down to the nanoscale. This is because the conductance fluctuations of the piezoresistor are dependent on the number of charge carriers in the resistor. The number of charge carriers in the resistor is a function of the carrier concentration and the volume of the piezoresistor. Most classic piezoresistor materials such as silicon, polysilicon, and metals show a square root relation between resistor volume and flicker noise as shown in Eq. (12), where \( C_C \) is the carrier concentration and \( \Omega \) is the resistor volume. Dynamic range will likewise scale with \( \Omega^{-1/2} \) since it inversely scales with sensor noise.

\[
\sigma_V = \sqrt{\frac{dV^2_S}{\Omega C_C} \ln \left( \frac{f_{max}}{f_{min}} \right)}
\]

The noise in carbon nanotube-based piezoresistive sensors, however, scales with the number of CNTs in the sensor [82,83] as well as with the geometry of the CNTs [84,85]. Therefore, the dynamic range of CNT-based piezoresistive sensors will not scale with \( \Omega^{-1/2} \). The optimal size of the piezoresistor will tend to scale with the flexure size when flicker noise is the dominant noise source in the piezoresistive sensor system. The optimal length of the silicon, polysilicon, or metal based piezoresistor is 1/3 of the beam length for a cantilever beam [13] in these cases. Therefore, the volume of the piezoresistor scales in direct proportion to the length of the flexure beam. The thickness of the piezoresistor is similarly set by the
Fig. 6. (a) CNT film strain gauge [69], (b) single suspended CNT displacement sensor [70], and (c) pressure sensor with CNT piezoresistors [70].

The optimal maximum ion implantation depth for a silicon-based piezoresistive sensor is 1/3 of the beam thickness [13]. For piezoresistors that sit on top of the flexure beam, such as metal and polysilicon piezoresistors, the maximum flexure thickness is generally set to be less than 1/10th of the flexure thickness. This thickness ensures that the performance of the flexure is not affected significantly by the presence of the piezoresistor. The volume of the piezoresistor therefore scales linearly with the thickness of the flexure. The maximum width of the piezoresistor is set by the width of the flexure beam, since the piezoresistor cannot exist outside the confines of the flexure. These scaling rules mean that when flicker noise is the dominant noise source, the volume of the piezoresistor scales with the cube of the flexure volume. Noise in these sensors then scales with the flexure volume to the 3/2 power.

Carbon nanotube-based sensors are not subject to the same scaling laws. The size of high performance CNT-based piezoresistors is generally set by the inherent size of the CNTs. CNT-based piezoresistors are a single monolayer of CNTs, so their thickness is constant no matter the beam thickness. The CNT-based piezoresistive sensors are also generally less than 1 μm in length, in order to ensure that each CNT is connected between two electrodes. This eliminates the need for CNT–CNT connections, which can significantly decrease the sensor’s strain sensitivity [69], and allows the CNTs to behave as ballistic conductors [86,87]. The size of the CNT-based sensor is therefore not affected as the flexure length is scaled down until the flexure beam is a few microns in length. The noise in the sensor should not be affected even when the length of the CNT-based sensors is reduced. This is because flicker noise in ballistic conductors is not affected by the number of charge carriers in the resistor [88]. The number of charge carriers only plays a central role in the noise of diffusive conductors where the scattering of different carriers is uncorrelated. However, this scattering is not relevant in devices where the length of the device is below the mean free path length of the electrons. The noise in the ballistically conducting CNT-based piezoresistive sensors should therefore not be dependent on the sensor length.

The width of the flexure will have an effect on the noise in the CNT-based piezoresistor. This is because the width of the flexure controls the width of the sensor, which in turn sets the number of CNTs which can fit in parallel in the sensor. Each CNT acts as a conduction channel through which electrons can flow. Decreasing the sensor width reduces the number of conduction channels, which increases the noise in the sensor system [82]. This means that the noise in CNT-based piezoresistive sensors scales with the square root of the sensor volume. This makes CNT-based piezoresistive sensor systems much less sensitive to sensor size than traditional silicon, polysilicon, and metal thin film piezoresistors.

If we assume a constant 10:10:1 length-to-width-to-thickness ratio flexure system, it is possible to plot the optimal performance of each type of piezoresistive sensor. The ideal CNT-based piezoresistors have the potential to outperform traditional piezoresistive sensors at most size scales, as shown in Fig. 7. This is only true if their strain sensitivity can be optimized by selecting the CNT chiralities with the highest strain sensitivities for the

Fig. 7. Sensor dynamic ranges versus size for various piezoresistive sensing materials.
piezoresistive sensor [72]. The difference between ideal CNTs and traditional piezoresistors is magnified when the flexure is less than about 10 μm in length, due to the differences in sensor scaling. CNT-based piezoresistors are currently capable of outperforming silicon based sensors when the flexure length is less than about 1 μm, and polysilicon sensors when the flexure length is less than about 100 μm.

Silicon piezoresistors are competitive with ideal CNT-based piezoresistors at flexure footprints of about 10,000 μm², as shown in Fig. 7. For large flexures, silicon piezoresistors are limited by Johnson noise, and their dynamic range stops increasing. For small flexures, the silicon-based piezoresistor’s performance drops off significantly as a result of the sensor scaling rules discussed previously. The polysilicon sensors perform similarly to the silicon piezoresistors except that they have a lower gauge factor and larger Hooge constant. This means that the performance of the polysilicon sensors starts to drop off when the flexure length decreases below about 1 mm. The metal thin-film piezoresistors have a dynamic range limited to about 80 dB due to their low gauge factor. They, however, do not start to see a drop off in performance in dynamic range until the flexure length decreases below about 10 μm, due to their high carrier concentration.

As piezoresistive sensors are scaled down to the nanoscale, CNT-based piezoresistors become the only viable option for high dynamic range sensing. For example, for footprints of less than 1 μm², CNTs are the only piezoresistive sensor capable of producing a dynamic range of greater than 60 dB. However, more work needs to be done to improve the performance of the CNT-based sensor systems before the CNT-based piezoresistive sensors can start to approach their theoretical performance limit and be incorporated into real high dynamic range nanoscale sensor systems. For example, it is not currently possible to produce sensors with only the highest gauge factor CNTs, since CNT sorting is limited to a few specific chiralities [89,90]. Therefore, it is necessary to improve CNT sorting technology so that the highest gauge factor CNTs can be sorted out and incorporated into nanoscale mechanical sensors. Also, improved manufacturing techniques are needed in order to reduce the noise in the sensors and reduce sensor-to-sensor variations in strain sensitivities. With these improvements, it should be possible for CNT-based piezoresistive sensors to approach their ideal theoretical performance limits.

3.2. Resonance sensors

3.2.1. Introduction to resonant sensors

Resonant NEMS sensors are a promising new class of sensors that have been made possible at the nanoscale due to the small mass, high natural frequency and high quality factor of nanoscale resonators. The small mass of NEMS devices allows such devices to operate at high frequencies. The natural frequency of the device is inversely proportional to the square root of the mass, as is seen in Eq. (13).

\[ f = \frac{1}{2\pi} \sqrt{\frac{k}{m}} \]  

(13)

Because of their reduced mass, NEMS devices may be run at much higher speeds for a given stiffness. For example, NEMS devices have been demonstrated that can run at frequencies of greater than 1 GHz [91]. This increased natural frequency can improve signal processing speeds and increase the sensitivity of nanoscale sensors and transducers that rely on the resonance of a beam.

The amount of power dissipated by mechanical systems such as MEMS and NEMS devices is also small, giving these mechanical devices high quality or Q factors compared to electrical resonators. The Q factor is defined in Equation (14), where \( \omega \) is the angular frequency, \( m \) is the mass, \( K \) is the stiffness, and \( R \) is the mechanical resistance or mechanical damping factor.

\[ Q = \frac{\omega \times \text{Energy Stored}}{\text{Power Loss}} = \frac{\sqrt{mK}}{R} \]  

(14)

The Q factor of NEMS devices may be hundreds of times better than similar high frequency electrical resonators [92].

3.2.2. Resonance-based mass sensors

Resonance-based mass sensors work by creating a frequency shift, \( \Delta f \), that is directly proportional to the change in mass. When mass is added to a resonating beam the natural frequency of the beam changes as described by Eq. (15) [93].

\[ \Delta f = -\frac{1}{2} \frac{\Delta m}{m} \]  

(15)

Nanoscale resonators are ideal for these types of sensors due to their extremely low mass and high natural frequency. The sensitivity of these sensors is generally limited by thermomechanical noise, which is caused by energy exchange between the flexure and environment [94]. These thermomechanical vibrations generally limit the dynamic range of resonant mass sensors to about 60 dB [95].

3.2.3. Resonance-based strain sensors

Resonance-based strain sensors are based upon the principle of strain stiffening. The pseudo-rigid-body large-deflection beam-bending model given by Eq. (16) can be used to predict the strain stiffening behavior of a clamped-clamped flexure beam [98]. Eq. (16) relates the applied force, \( F \), to the change in length of the beam, \( \Delta L \), via the initial clamped-clamped beam length, \( L_0 \), the position along the beam, \( x \), the stiffness coefficient \( K_0 \), the characteristic radius factor, \( \gamma \), flexural rigidity, \( (EI)_{\text{bend}} \), and \( (EA)_{\text{axial}} \) [99].

\[ F = \frac{2\Delta L(EI)_{\text{axial}}}{L_0 + \Delta L} \sin^{-1} \left( \frac{x}{\sqrt{x^2 + \gamma^2 L_0^2}} \right) \]

\[ + \frac{8\gamma^2 K_0(EI)_{\text{bend}} \tan^{-1}(x/yL_0)}{x^2 + \gamma^2 L_0^2} \]  

(16)

When tension is applied to the beam, the beam elongates and \( \Delta L \) increases, which causes the stiffness to increase. This increase in stiffness then results in a higher natural frequency of the beam. This increased natural frequency can be described by Eq. (17), where \( \mu \) is the mass per unit length of the flexure and \( \varepsilon \) is the axial strain applied to the flexure [100].

\[ f = \frac{4\gamma^2}{2\pi EI} \sqrt{\frac{EI_{\text{bend}}}{\mu} + \frac{(EA)_{\text{axial}} L_0}{4\pi^2 \mu}} \]  

(17)

Therefore, based on the change in natural frequency of the resonating beam it is possible to measure the strain applied to the beam.

Resonant strain sensors are not commonly used as macroscale strain sensors due to the range limitations created by the low failure strains of most macroscale materials. However, CNT-based resonant strain sensors have recently been shown to be able to overcome some of these limitations [100–103]. A schematic of a CNT-based resonant strain sensor is given in Fig. 8.

Carbon nanotubes are well suited for resonance-based strain sensors due to their high elastic modulus [104], high strength [105], and failure strain of greater than 20% [106]. These properties give
CNT-based resonant strain sensors have a high natural frequency and large range. For example, based on molecular dynamics simulations of a 15.28 nm long, (5,5) single walled carbon nanotube, CNT-based resonant sensors can have a strain sensitivity of about 1 picostrain per hertz, as seen in Fig. 9. For the 15.28 nm long, (5,5) CNT this translates to a force sensitivity of approximately 1 pN/MHz and a displacement sensitivity of approximately 10 pm/GHz with a range of approximately 3.7 nm. These sensors are therefore good at measuring small strains over nanoscale displacement ranges.

Graphene-based nanoelectromechanical resonators are also promising due to their high elastic modulus (≈1 TPa [107]), high yield strain (up to 20% [107]), and atomic thickness. These properties allow graphene-based resonators to have natural frequencies in the 100’s of MHz range [108–110] while maintaining a small mass. This is important because it means that even a small change in force or mass will result in a large change in natural frequency. Therefore, it is possible to get very sensitive force and mass measurements by measuring the change in natural frequency of graphene-based resonators.

4. Conclusions

A large variety of nanoscale sensors will become necessary as new types of NEMS devices are invented. Resonance sensors, tunneling sensors, and CNT-based piezoresistive sensors offer the potential to overcome some of the fundamental limitations of nanoscale sensing. However, more work needs to be done on these types of sensors in order to make many nanoscale devices practical for real world applications. For example, it is necessary to improve the sorting of CNTs by chirality in order to maximize the strain sensitivity of CNT-based piezoresistive sensors. Similarly, the design and environmental isolation of nanoscale tunneling sensors must be improved in order to reduce both the noise in these sensors and their sensitivity to changes in the operating environment.

NEMS nanomanufacturing techniques must also be improved to allow reliable, repeatable NEMS to be produced at the high volumes and at low costs necessary for many medical, chemical and electronics applications. For example, currently the fabrication of nanotube-based resonators often involves placing many CNTs on a substrate, then using electron beam lithography to define the rest of the electronics structure around the CNTs. This type of fabrication procedure is time consuming and costly, which makes it impractical for large-scale fabrication of resonant sensors. In addition, this type of fabrication procedure can result in wide variations in the initial strain applied to the CNT. This causes the natural frequency of the NEMS resonator to be inconsistent from device to device. Therefore, new sensor designs and fabrication techniques are required in order to improve the repeatability of NEMS resonant sensors.

Despite these limitations, there are many important applications for NEMS devices that should drive the development of new types of nanoscale sensors and create interesting new research opportunities.

References


