



LAWRENCE
LIVERMORE
NATIONAL
LABORATORY

Sensor and actuator considerations for precision, small machines: a review

S. T. Smith, R. M. Seugling

April 12, 2005

Precision Engineering

Disclaimer

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

Sensor and actuator considerations for precision, small machines: a review

*Stuart T. Smith**, *Richard M. Seugling*^{♦♥}

Center for Precision Metrology, UNC Charlotte, Charlotte, NC 28223

[♦] Lawrence Livermore National Laboratory, Livermore, CA 94551

Abstract

This article reviews some design considerations for the scaling down in size of instruments and machines with a primary aim to identify technologies that may provide more optimal performance solutions than those, often established, technologies used at macroscopic, or conventional, scales. Dimensional metrology within emerging applications will be considered for meso- through micro- down to nanometer level systems with particular emphasis on systems for which precision is directly related to function.

In this paper, attention is limited to some of the more fundamental issues associated with scaling. For example, actuator work or power densities or the effect of noise on the sensor signals can be readily evaluated and provide some guidance in the selection for any given size of device. However, with reductions in scale these parameters and/or phenomena that limit performance may change. Within this review, the authors have tried to assess these complex inter-relationships between performance and scale, again from a fundamental perspective. In practice, it is likely that the nuances of implementation and integration of sensor, actuator and/or mechanism designs will determine functionality and commercial viability of any particular system development.

* Corresponding author stusmith@uncc.edu

[♥] This work was performed in part under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

1.0 Introduction

Although, small-scale machine tools and instruments, such as lathes, have been built for many years in the watch-making and other industries, a limited number of these systems have been incorporated into modern manufacturing. Most of the small scale machine tools and instruments are merely a scaled down version of a conventional machine. Typically the accuracy or precision of these small machines is no better, more often worse, than that of high-quality macro- machines/instruments. In an effort to identify areas of future research, fundamental considerations are used to assess the fidelity of possible approaches to common design problems. At extremes of scaling, it is obvious that some requirements of machines and instruments such as primary actuation methods or position sensing, for example, may be best achieved using different technologies. It is important that designers become aware of both the available technologies as well as the relevant performance measures as a function of dimensional scale. With this in mind, it will then be possible to identify transitions in scale at which different technologies will provide a more optimal solution to a design need.

Key features common to the mechanical systems being considered in this review are the ability to provide work in a controlled manner and the use of sensors for feedback of process condition, particularly relative motions. Based on the focus of scaling down, two major machine requirements are discussed, these being; actuators to provide position and/or generate work and sensors for detecting surfaces and measuring relative motions.

The dimensional scales of the systems classifications referred to in this text are

1. Meso-systems, being intermediate between human scale devices and micrometer mechanisms, implies that major dimensions of such a system will typically be measured in millimeters.
2. Micro-engineering is the intermediate region in which micrometers are more sensibly used to specify functional tolerances. This length scale is often considered to incorporate MEMS processing.

3. Nano-engineering is considered to be any process resulting in devices of defined structure or assembly in which the nanometer is the most sensible unit for specifying dimensions or tolerances.

Precision at macroscopic scales is usually considered to start with range to accuracy, or range to resolution performance at around 1 part in 10^5 or better. A similar criterion applied to meso-systems immediately implies that, if precision is to be maintained, total errors should be below 10^{-8} m or so. In practice, economics dictates that such high precision only be maintained when absolutely necessary for the function of the device. Currently, it is our experience that, when assembling small machines using available small parts, tolerance stack-up often leads to unacceptable degradation of performance. In many processes, to control to the relevant precision will require development of metrology and actuation systems of considerably better performance than currently available from macroscopic devices. Consequently, these simple considerations indicate that dimensional metrology and mechanism design for small-scale machines must achieve reliable measurement and positional accuracy at, or approaching, nanometer levels. It is possible that, as the size of systems reduce; major differences are likely to include the following.

1. It may be viable to incorporate metrology systems within manufacture and/or assembly.
2. There are likely to be more metrology systems, possibly one for every device.
3. It must be possible to manufacture these cheaply and utilize automated assembly.
4. Power consumption will be low while, for systems with inherent losses, efficiency will be difficult to maintain or improve.
5. New, or considerably more compact, methods of communication and data handling will be required. Connectors and interfaces often account for a major fraction of the volume of a circuit.

6. It will be necessary to achieve higher bandwidths and resolution than normally required of macro-scale metrology instruments.
7. Alignments of moving axes and subsequent errors will be particularly difficult to adjust and measure.
8. Measurement frames encompassing the whole instrument may not add significant cost and, being small, may be relatively insensitive to environmental variations.
9. Volumetric dimensions for measurement systems and forces imposed by the sensing devices should be low relative to the forces of the process.

A question that has not, in our view, received sufficient attention is the distinction between these different realms of technological endeavor. In practice, it is both logical and, probably, fair to say that we are considering a continuum of component scales incorporated into any modern machine. Currently, it is common to distinguish different scales by the processes used in manufacture. As manufacturing and assembly processes develop and adapt, it is likely that these distinctions will be less clear and designers will be required to incorporate technologies covering all dimensional scales as a matter of course. However, there are important features to be considered as dimensional scales reduce. Primary among these are

1. There is a certain cut-off after which the use of manual assembly, that has been spectacularly useful and essential in the development of technologies so far, is no longer viable.
2. Phenomena within mechanisms that produce significant mechanical forces may change (i.e. contact forces, adhesion forces, meniscus effects, electric dipoles, van der Waals and Casimir forces) thereby requiring additional consideration in the calculation of two-body surface interactions.
3. Scaling can change the relative dynamics of phenomena. For example heat transfer is often considerably slower than mechanical motion at macro-scales

while thermal agitation might be the fastest means of providing motile forces at molecular levels. From another perspective, thermal expansion, which is a major limiting influence on possible precision in a large machine, may be insignificant at smaller dimensions.

4. As the net removal or deposition volumes reduce, other manufacturing processes, reductive or additive, previously considered too slow, may provide economic production rates.

Such considerations may provide a guide to favorable scaling in the selection of processes and components.

Consider a machine or instrument arriving in a box weighing approximately 50 kg or less, able to fit sparingly on a desktop and capable of producing or measuring parts with a precision of 1 part in 10^6 given a work volume of several cm^3 . As already mentioned, given the precision stated, the resolution of the machine or instrument would have to be on the order of tens of nanometers over centimeter ranges. The concept seems very straight forward given the advanced technology of today. However, why don't these types of machines show-up in manufacturing plants around the world? Simply scaling down current technologies used in modern computer controlled machine tools or instruments would initially give the impression of providing for current manufacturing needs. However, this may not be viable because the complex nature of the interrelated physical and chemical interactions at different scales often curtails the designer's ability to shrink a complex machine/instrument without careful consideration of how the reduction-to-practice of such a system would perform under real operating conditions. Such issues are beyond the scope of this review. However, to achieve precision, small actuator and sensor systems are necessary enabling technologies of meso- and smaller scale systems. Fortunately, there are limited numbers of possibilities for these and an assessment of the relative merits of available options in terms of the effects of scaling is presented in this review.

Actuator methods currently considered include; piezoelectric, electrostatic, electromagnetic, magneto-strictive, hydraulic/pneumatic, thermal and shape memory.

One particular measure of the fidelity of these actuators is the theoretical maximum amount of work and/or power that can be produced per unit volume occupied by the actuator (called the work density or, when scaled by the dynamic bandwidth, power density). Considering specific implementations it is found that some of these methods maintain performance (piezo-electric, hydraulic, electrostatic) as scales reduce while electromagnetic actuators tend to exhibit reduced work density and thermal-based actuators appear to improve. Identification of the scales at which different techniques dominate will help the designer to select the appropriate technology for a given application. Such analysis also provides information about other parametric scaling factors such as drive voltage amplitudes, forces etc. A number of implementation issues will also be considered.

A similar approach has been applied to sensors for small-scale systems such as optical (line scale, interferometric, diffractive, position sensitive detectors) and electromagnetic. In this latter category it is necessary to split the categories into transducers and displacement sensor systems. In our study we have attempted to identify the relevant scaling parameters and assess their relative influence as scale reduces.

2.0 Displacement sensors

2.1 General considerations

Displacement sensors may be considered any transduction method that results in a deterministic variation in output for a change in relative position between two objects. Outside of cost, in terms of performance measures signal to noise, repeatability and stability tend to be the most important parameters for the assessment of potential precision. In many cases a smooth and continuous variation of output with displacement will provide adequate means for the determination of relative displacements. Periodic rulings can be used with most of the considered sensing methods to enable an increase in precision by effectively repeating the short-range performance while increasing total range in proportion to the number of periods. However, in this case, accuracy also depends upon the accuracy of the periodic ruling. For example, in optics it is possible to

exploit near field probes to read a grating of sub-wavelength pitch, while using capacitance it is possible to use an almost arbitrarily small tip [1, 2] and again read from a periodic structure. Magnetic sensing from a periodic magnetic structure is commonplace in the magnetic data storage industry and it is possible to discriminate sub-micrometer magnetic features in some cases. It is not always necessary that the rule have exact periodicity and in some cases it is possible to read displacements from a random pattern of known characteristic leading to a pseudo-absolute scale. However, common to most of these methods is the requirement for the sensing element to be in near proximity to the scale and usually with a separation comparable to the finest features on the scale (diffractive sensors that effectively average over many lines of a grating represent an exception). Often times these are referred to as ‘near-field’ probes and represent a generalized approach to small-scale (often highly localized), long-range position sensors. In this case, there will be constraints on the necessary precision of aligning the scale with that of the moving platform to provide sufficient signal contrast during traverse. It is also common to utilize multiple measurements of the periodic grating with a lateral phase separation to enable amplitude independent, directionally sensitive measurement of displacement. Again, the precision of such techniques are predominantly limited by signal contrast.

Other near-field phenomena are adequately represented by the sensing elements commonly used in scanned probe microscopy utilizing contact based or near-surface interactions such as electrostatic, tunneling, van der Waal, meniscus, Casimir, magnetic resonance, optical, ion conductance, thermal, squeeze-film and piezoelectric. Of each of these phenomenological interactions there are a large variety of methods for conditioning the signals to derive different surface interaction information. In many cases, combinations of these interactions are either exploited or impinge upon the interpretation of the measured signals.

2.2 Optical

A comprehensive study of the methods and techniques for optical measurement of displacement could fill libraries. In this section we will restrict our attention to only those factors affected by scaling to small size. In particular, as the mechanisms reduce from

meso- to micro- scale, the size of components will typically be comparable to the wavelength of visible light. Whether or not this imposes practical limitations on optical position measurement methods will be seen in the future.

2.2.1 Laser interferometry

Ignoring technical error sources, the fundamental limits to interferometer resolution are determined by noise in the laser source and the detectors [3, 4]. Matching both noise sources to produce an equal contribution, the bandwidth, B , limited resolution in displacement dx can be shown to be

$$dx \approx \frac{hc\sqrt{B}}{\pi\eta NEP\sqrt{2}}, \quad (1)$$

where h is Plank's constant, c is the velocity of light in a vacuum, η is the quantum efficiency of the detector and NEP is the noise equivalent power of the detector. Using commercially available detectors a noise level of pm over bandwidths of many kHz are theoretically possible. In practice, there are a large number of other factors that limit accuracy and resolution in the implementation of a particular interferometer method. For example, most commercial interferometers utilize frequency splitting with each frequency traversing the two paths within the interferometer. These are recombined and the Doppler shifted beat frequency used as a measure of displacement. The use of polarizing optics and imperfections in the assembly or optical components often lead to errors that are considerably larger than those due to detection and laser noise [5].

At meso- and micro-scales, interferometer packages will require small optics to compactly integrate within a system. As well as the usual environmental effects such as corrections in air for pressure, humidity and temperature, there is also aperture correction given by [6]

$$\Delta s = \left(\frac{R}{2f} \right)^2 s, \quad (2)$$

where R is the radius of the source (aperture stop or core of fiber-optic light guide), s is the measured length change of the optical path based on laser frequency and f the

collimator focal length. Typically, for a meso- system with laser source coming from a fiber feed, $R \approx 10^{-5}$ m, collimated at $f \approx 10^{-3}$ m, it is possible that this particular error term may introduce a limit on precision of the order one part in 10^5 or worse.

Other correction factors depend on the path length in air. Fortunately, while the effects are often more significant for interferometers with large ranges, producing path length variations of the order of parts in 10^7 , this represents typical uncertainties of nanometers for ranges of less than 10 mm.

Another problem is the size of the laser source. Conventional HeNe can provide stabilities of better than 1 part in 10^7 at reasonable cost and with small line width (long coherence length) but the laser tubes are of the order of tens of centimeters in length. Laser diodes with grating stabilization might realize stabilities of the order of parts in 10^7 but are limited in their coherent lengths [7]. Electrical current (or parameter) stabilized diode interferometers with stabilities of the order parts in 10^5 to 10^6 are likely to become available in the near future [8]. Currently, ranges of greater than 100 μm with measurement uncertainties of better than 10 nm are being claimed. In principle, it is possible to enhance coherence (reduce line-width) by filtering through a Fabry-Perot cavity. However, it is not clear whether this can be achieved without resulting in a bulky source.

Interferometers using small optics have been developed by Renishaw, PTB with the frequency stabilized laser diode, the Tropel small bore interferometer [9], Dunn *et al.*, 2002, Swiss-based small optics manufacturers Fisba [10]. Like some of the larger optical interferometers, to keep the bulky laser head out of the machine and away from harsh environmental effects, the laser beams are fed to the optical system through an optical fiber.

In some systems the fiber itself forms an intrinsic component of the interferometer, thereby further reducing the size and number of optical components. Such systems are often called fiber interferometers [11].

White light probes have been developed to measure local heights on surfaces. In many ways such systems contain similar elements to a Michelson interferometer. In this case the 'white' light is split between a stationary reference and the second is focused onto an area of the surface to be measured. When the two beams are recombined, there

will be interference with the intensity depending on the path lengths of the two beams and wavelength that will generally average out to a mean intensity. In the case where both beams have the same optical path length, all frequencies will constructively interfere with similar phase shifts resulting in an increase in intensity signal. Hence, it is possible, by changing the path lengths to record the corresponding change in intensity at any point and use this information to reconstruct the three dimensional topography of the surface reflections. However, in such a system, local surface height is transferred to a displacement of the stationary mirror, or reference surface, relative to the specimen surface. The point at which this measurement is triggered is complicated by the surface reflective properties of the specimen being measured and the spectral content of the illumination. Given these limitations, it is possible to monitor surface height variations at nanometer levels while scanning surface heights measured in millimeters. Other advances in data processing and system design have enabled dynamic scanning of relatively large areas with movie capabilities providing visualization of motions and distortions in micrometer sized mechanisms (MEMS).

An alternative white light technique is to focus the light to a point on the surface and monitor wavelength maxima upon recombination with a reference. Variants on this include confocal imaging in which the white light is focused through a lens to create a wavelength-dependent focal spot. Again monitoring the wavelength on reflection or chasing the focal point reflected back through a lens enables a measure of the surface height from a reference. This enables fast measurement of surface displacements with nanometer resolution over measurement ranges of hundreds of micrometers (range-to-resolution of better than 10^5) and requiring spot sizes typically less than a millimeter [12].

2.2.2 Linescale interferometry

There are a number of suppliers of linescales for the measurement of both linear and angular motion [13]. Generally, scales are supplied in the form of a bolt-on linear scale or in tape form with self adhesive for ease of installation with a separate read head.

In general, high-precision linescale interferometers utilize change in intensities of the various orders of reflection from a finely scribed diffraction grating. With any grating, one of the limitations is the size of the illumination spot relative to the spacing of the

grating, c , which are, in turn limited by the diffraction angle, ϕ_n , of orders n governed by the relation

$$\sin \phi = \frac{n\lambda}{c}, \quad (3)$$

where λ is the wavelength of illumination on the grating by the incident beam. Clearly this function must be less than 1, thereby setting a limit on the spacing c of the grating for a given wavelength and thereby imposing a limit on scaling. For example a grating of 1 μm illuminated with a wavelength of 0.78 μm will have first order diffraction peaks at angles of $\pm 51^\circ$. Similar limitations apply to transmission gratings. Recent designs have enabled the use of integrated optics for compact optical sensor systems of major linear dimensions of less than 10 mm. An example of this is the commercially available scales of Sony that use 1200 nm grating pitch with quadrature detection yielding a signal pitch of 400 nm [14]. With suitable phase measurements, resolutions of better than 10 nm are possible.

For a suitable geometry of linescale and analyzer gratings at the sensor [15], the phase, Ω , of any individual order of diffraction as a function the relative position of the grating, x , is given by

$$\Omega = \frac{n2\pi x}{c}. \quad (4)$$

The output intensity at the sensor, I_o , is proportional to

$$I_o \propto (1 + \cos(2\Omega)). \quad (5)$$

This corresponds to a periodic variation in intensity at twice that of the grating period, c . Other designs may increase this further to 4 or more variations per period. By monitoring the symmetric orders it is possible to derive a directionally sensitive measurement of motion of the grating relative to the optical probe.

At reduced scales, the effect of the relative spot size is to influence the discrimination of different orders. For a spot size occupying N grating periods, the

intensity of reflection at an angle θ from the point of incidence can be approximated from [16]

$$I_f \propto \left(\frac{\sin\left(\frac{Nkc\theta}{2}\right)}{\sin\left(\frac{kc\theta}{2}\right)} \right)^2 \left(\frac{\sin\left(\frac{kc\theta}{4}\right)}{\frac{kc\theta}{4}} \right)^2, \quad (6)$$

where N is the number of lines in the field of the incident beam and k is the wave number ($=2\pi/\lambda$). It can be seen by plotting the above equation that the discrimination of each diffraction order will diminish with the relative spot size.

In general, mainly because of problems with alignment errors, calibration of scales and fringe interpolation, linescales typically provide 5 nm – 10 nm resolution with some extending down to sub-nanometer. At the extremes of performance, it has been shown that linescales can provide displacement repeatability at better than 0.5 nm [17] and calibration facilities at PTB in Germany and Mitutoyo in Japan have demonstrated uncertainties of better than 30 nm in the calibration of scales longer than 500 mm [18].

Currently, there is a lot of activity in the field of small-scale optical form and displacement measuring devices. Recent developments include pattern recognition of 3-D interferograms, diode laser interferometry [19] and many variants of near-field scanning optical microscopy and related probing techniques [20, 21]. It is clear that efforts are being made to decrease the size of these measurement systems, while maintaining or increasing their resolution capabilities.

2.2.3 X-ray interferometry

X-ray interferometry is often referred to, with some vociferous opponents to this, as the x-ray equivalent of the Mach-Zehnder interferometer. Being based on the lattice spacing of a crystal, it is intrinsically accurate to better than parts in 10^7 , is portable and can be calibrated to two orders of magnitude better than this. Additionally it can operate in air with little influence of refractive index because of the low unit decrement (i.e. near unity refractive index) of materials at x-ray wavelengths.

In principle, it is possible to shrink x-ray interferometers to produce a linescale in which the splitter and mirror need be only large enough to diffract the x-ray beam which, in turn can be micro-focused to sub-millimeter cross section. The analyzer would then be limited only by the length of the available crystal. Currently, crystals of near thermodynamic perfection can be obtained in lengths of up to 1.0 m. In practice, there are considerable technological barriers in implementing this interferometer. Most demanding is alignment with angular deviations of better than a few tens of nano-radians and micrometer separation tolerances between the three crystals being necessary. Another constraint is the speed of the interferometer that is limited by the intensity of available sources and detector quantum efficiency. The expense, size and safety considerations associated with high intensity x-rays has limited this measurement technique to a few research labs scattered around the world.

2.3 Capacitance

Capacitance is commonly utilized for precision measurement of relative displacement. Like any single degree of freedom system the voltage noise will be fundamentally limited by Brownian motion [22] of charge such that

$$\frac{1}{2} C \overline{V^2} = \frac{1}{2} \frac{\overline{q^2}}{C} = \frac{1}{2} kT, \quad (7)$$

where C is the capacitance, V is the applied voltage, q is the stored charge, k is Boltzmann's constant ($1.38 \times 10^{-23} \text{ J}\cdot\text{K}^{-1}$) and T is the absolute temperature.

In practice, such a noise flaw could only be realized if a perfect signal conditioning circuit was to exist. In practice, capacitance is measured by providing an oscillating voltage (or current) across (through) the capacitor and measuring the reactance of the capacitive element. Noise sources in such a measurement include

1. Oscillator noise. This is a function of the oscillators chosen and will manifest as combinations of amplitude and phase components. Typically it is the amplitude that is most difficult to control. However, this situation is changing with digital

oscillators. This will be considered an inextricable noise present in the carrier signal and having both $1/f$ and white components.

2. Amplifier noise.

It is assumed that the output from a Wheatstone bridge is monitored by an instrumentation amplifier, the input being a single harmonic carrier frequency. In such a circuit the amplitude depends on the imbalance of the bridge. Stray capacitance effects will not be included in this analysis.

The output noise per root frequency, e_s , from the amplifier of voltage gain G_V with zero impedance from the source is given by

$$e_s = e_n G_V, \quad (8)$$

where e_n is the input noise typically of the order $1 \text{ nV}\cdot\text{Hz}^{-1/2}$ for a precision instrumentation amplifier at frequencies above 1 kHz. Because we will be using carrier frequencies of 1 kHz or greater the $1/f$ noise component is considered negligible.

Adding a source impedance Z_s the amplifier output noise is given by

$$e_s^2 = \left[e_n^2 + 4kTZ_s + (i_n Z_s)^2 + e_o^2 \right] B G_V^2, \quad (9)$$

where i_n is the input noise figure (often less than $1 \text{ pA}\cdot\text{Hz}^{-1/2}$ at frequencies higher than 1 kHz), B is the bandwidth, and e_o is the oscillator noise per root Hertz (assumed to be zero for the present purposes). The above values have been used to derive Figures 1 and 2 and represent the theoretical noise limits as a function of geometry.

As an example, consider a source of the form of a transformer ratio on one arm and matched capacitances making up the other [23]. This can reasonably be modeled a half bridge input to one terminal and a zero source impedance on the other. The source impedance is

$$Z_s = \frac{Z_1 Z_2}{Z_1 + Z_2} = \frac{1/\omega_c^2 C_o C}{1/j\omega_c C + 1/j\omega_c C_o} \approx \frac{1}{2C_o \omega_c}. \quad (10)$$

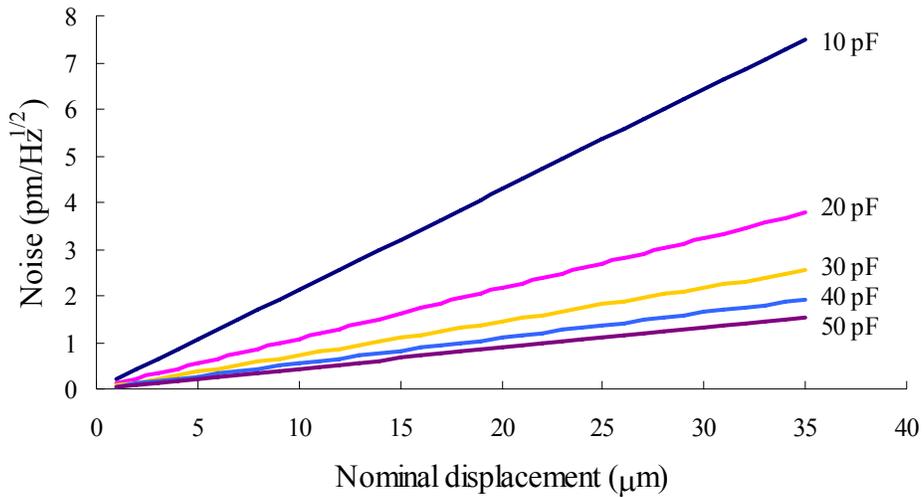


Figure 1: Noise as a function of gap for a number of capacitance values.

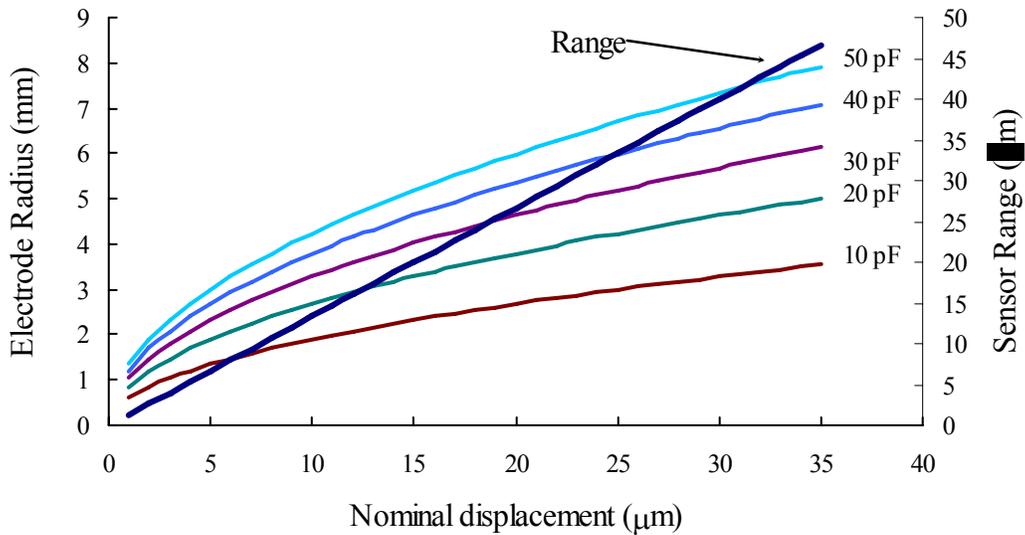


Figure 2: Combined plot of the electrode radius and sensor range as a function of nominal displacement. Illustrates some important performance parameters of typical parallel plate capacitance gages.

For high precision measurements it is most common to utilize the more sensitive variable gap configuration considered below over the variable area. Typically, for a capacitance gage we seek to utilize a range incorporating capacitance in the ratio 1:2:3.

This gives rise to the important relationship that the range of the capacitance gage is 4/3 times the nominal value defined at the center value of capacitance.

In general, the capacitance ratio from a nominal value can be expressed by

$$k_1 C_o : C_o : k_2 C_o . \quad (11)$$

The capacitance range is given by the difference between the maximum and minimum displacement given by

$$\frac{x_o}{k_1} - \frac{x_o}{k_2} = x_o \left(\frac{1}{k_1} - \frac{1}{k_2} \right) . \quad (12)$$

It can be readily verified that the 1:2:3 (k values of 1/2 and 3/2 respectively) rule provides a capacitance range of 4/3 of the nominal electrode separation.

Consequently, at the maximum electrode separation the capacitance will be reduced to a value of 1/2 of the reference. This will increase the source resistance by 17% (1/6). Substituting (10) into (9) for the maximum source resistance yields

$$e_s^2 = \left[e_n^2 + 4kT / 3C_o \omega_c + (2i_n / 3C_o \omega_c)^2 + e_o^2 \right] B G_V^2 . \quad (13)$$

In practice, this appears to be dominated by the amplifier current noise followed by thermal noise.

It can readily be shown that the output voltage from the transformer ratio bridge for small changes in one capacitance is given by

$$V_{out} = \left(\frac{V_{bs} G_V}{4x_o} \right) x = S_d x . \quad (14)$$

Additionally the nominal separation is related to this nominal capacitance by

$$x_o = \frac{\epsilon A}{C_o} , \quad (15)$$

where x is the displacement of the capacitance electrode, V_{bs} is the bridge supply voltage and S_d is the sensitivity of the bridge circuit in units of volts per meter. This can be expressed in alternative forms given by

$$S_d = \left(\frac{V_{bs} G_V}{4x_o} \right) = \left(\frac{V_{bs} G_V C_o}{4\epsilon A} \right). \quad (16)$$

From equations (13), (14) and (15), the displacement noise is given by

$$\begin{aligned} \frac{e_x^2}{B} &= \left[e_n^2 + 4kT/3C_o\omega_c + (2i_n/3C_o\omega_c)^2 + e_o^2 \right] G_V^2 \left(\frac{1}{S_d} \right)^2, \\ &= \frac{16x_o^2}{V_{bs}^2} \left[e_n^2 + 4kT/3C_o\omega_c + (2i_n/3C_o\omega_c)^2 + e_o^2 \right], \end{aligned} \quad (17)$$

in units of $[\text{m}^2 \cdot \text{Hz}^{-1}]$.

In general, the amplitude (with sign change across bridge balance) of carrier frequency is provided by the synchronous demodulator. Taking the square root of equation (17) gives the displacement noise of the circuit

$$\frac{e_x}{\sqrt{B}} = \frac{4x_o}{V_{bs}} \sqrt{e_n^2 + 4kT/3C_o\omega_c + (2i_n/3C_o\omega_c)^2 + e_o^2}. \quad (18)$$

In the design of a capacitance gage, there are three considerations; range, electrode geometry and noise floor. Basically it is required that the noise floor of the capacitance gage is less than a design goal. Typically this would be the desired resolution of the measurement system at an acceptable bandwidth. In practice, this could always be achieved with a sufficiently large electrode area. However, this is impractical for reason of electrode alignment requirements (acceptable parallelism) and volume occupied. A comparison of these trade-offs can be seen from the design charts of Figures 1 and 2. From this it is clear that there are limits on size being fundamental and dependent upon the precision of the measurement electronics.

Comparison of these figures with currently state of the art sensors is only possible if the characteristics of the first amplification stage and ratio k are known. Ignoring this and noting that the following two graphs are based on rather conservative estimates for amplifier noise figures, it is found the noise data from commercial suppliers, although close, often surpasses that shown in these plots. Typically, the difference from these plots and available sensors is within factor of 2 and therefore gives a reasonable estimate of achievable precision in term of selecting and appropriate ADC (i.e. chose an ADC of precision better than 1 bit that predicted for the sensor). However, achieved noise levels will depend upon implementation and we consider that these graphs are representative of what might currently be realistic for many applications.

2.4 Inductance

The so-called linear variable differential inductor (LVDI) has been used as the sensing element in the Talystep stylus profilometer for upwards of fifty years and still represents one of the best methods for measurement of nanometer sized step heights for calibration of film thickness monitors in thin film deposition units.

Again, it is common to measure variations in reactance using an AC driven Wheatstone bridge. Assuming equal inductors on each arm of the bridge of nominal value, L_o , the output voltage, v_o , for a change in inductance, $\pm \Delta L$, of the inductors on one arm is

$$v_o = v_s \frac{\Delta L}{2L}. \quad (19)$$

To assess the performance of a simple inductive probe, consider a magnetically permeable core of length l_c and permeability μ_c having a coil of N turns surrounding it. If this core terminates from forming a loop so that there is an air gap of length $2x_a$ with a nearby permeable target that forms a further part of the core, see Figure 3. Upon excitation of the coil with a current I there will be an induced magnetic flux Φ given by

$$\Phi = \frac{NI}{\oint} = \frac{NI}{\frac{2x_a}{\mu_a A_a} + \frac{l_c}{\mu_c A_c}}. \quad (20)$$

where μ_A and A_A represent the permeability and effective area of the gap respectively.

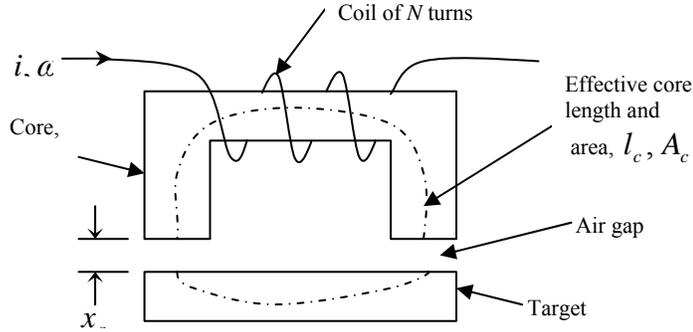


Figure 3: Schematic diagram of an inductive sensor utilizing a permeable target.

The inductance L of the sensor can be approximated from

$$L = \frac{N\Phi}{I} = \frac{\mu_c A_c}{l_c} \frac{N^2}{\frac{2\mu_c A_c}{\mu_a A_a} \left(\frac{x_a}{l_c} \right) + 1} = \frac{L_o}{\frac{2\mu_c A_c}{\mu_a A_a} \left(\frac{x_a}{l_c} \right) + 1} \quad (21)$$

$$\approx L_o \left[1 - \frac{2\mu_c A_c}{\mu_a A_a} \left(\frac{x_a}{l_c} \right) + \left(\frac{2\mu_c A_c}{\mu_a A_a} \left(\frac{x_a}{l_c} \right) \right)^2 \dots \right]$$

For very small air gaps, so that only the second term in parentheses is significant, a small change in displacement will produce a corresponding change in inductance such that

$$L + \Delta L \approx L_o \left[1 - \frac{2\mu_c A_c}{\mu_a A_a} \left(\frac{x_a + \Delta x}{l_c} \right) \right], \quad (22)$$

$$\therefore \Delta L \approx -\frac{2N^2}{\mu_a A_a} \left(\frac{\mu_c A_c}{l_c} \right)^2 \Delta x. \quad (23)$$

As with capacitance, the noise limit of such a sensor can be determined from the amplifier performance and the output impedance of the bridge. With an ideal reference arm of the bridge, the output impedance is

$$Z_0 = \frac{j\omega L}{2}. \quad (24)$$

Using the same procedures as stated previously, the noise power of this sensor is

$$e_s^2 = \left[e_n^2 + 2L\omega kT + (i_n L\omega / 2)^2 + e_o^2 \right] BG_v^2. \quad (25)$$

In this case it can be seen that the noise will increase with the carrier frequency.

A variant on the differential inductor is the linear variable differential transformer (LVDT). LVDT's can consume low power and operate in high vacuum, as a non-contact position sensor that has nanometer resolution and centimeter range [24].

In terms of scaling, for either a transformer or differential inductor it is necessary to incorporate at least two separate loops of conductor with each having as many turns as possible to increase induced voltage. A permeable material must also be introduced into these loops and will experience an induced magnetic flux, the density of which is inversely proportional to area of cross section through which the flux passes. Currently, as far as the authors are aware, such transducers are manufactured and assembled using conventional wires as conductors and ferrite materials to create the permeable paths. Typically, these are of macroscopic scale with most manufacturers providing miniature sensors of the order 6 mm - 8 mm diameter and 10 mm – 20 mm long. Some LVDT's have been manufactured with total dimensions of around 5 mm × 5 mm × 5 mm and able to sense displacements of up to 200 μm and with signal to noise ratio of approximately 1 part in 10,000. By increasing amplifier gains, resolutions of better than 50 pm have been observed with a range of around 50 – 100 nm [25]. At smaller scales, it is difficult to manufacture multi-turn coils having sub-millimeter dimensions and incorporate

permeable materials. While it might be possible to manufacture these components using MEMS or similar processes, it is not clear that coils of adequate performance can be readily produced.

2.5 Resistance

To the author's knowledge, direct methods for using resistance as a measure of displacement have not found widespread application for precision measurement. For limited rotations potentiometers have been used as a method of feedback for spindle rotation. For example, in the auto industry these are sometimes used for throttle position, steering angles etc. However, these have been almost universally superceded by optical or magnetic encoders for more precise sensor applications. For limited range motions, piezo-resistive strain gages have been used for many decades to detect strain of piezoelectric actuators either by attaching the gage directly to the actuator element or indirectly by monitoring strain of the flexure in which the actuator is constrained. Having upwards of 50 times the sensitivity of conventional metal strain gages, relatively high precision measurements are possible. For example, typical commercial actuators having a range of 20 μm will claim resolutions of better than 5 nm with bandwidths of around 1 kHz. Typically, metallic resistors have gage factors of around 2 and maximum elastic strains of the order 0.4 % while voltage noise is often measured in microvolts over bandwidths of around 1 kHz [26]. A simplistic calculation indicates signal to noise ratios of the order $10^3 - 10^4$, a figure comparable to those quoted for commercial devices. Proximity probes are also produced using this sensing method and, suitably designed; demonstrate sub-atomic resolution, although at the expense of limited range.

2.6 Magnetic

There are a number of different sensors that can be used to measure magnetic fields. Particular sensors, such as SQUID's (super-conducting quantum interference detectors) or fluxgate sensors can exhibit very high sensitivities but are currently limited in terms of a relatively large power consumption and size. Promising sensors for the detection of magnetic fields are those based on spin-dependent conduction such as giant magneto-

resistance, (GMR) [27] and spin dependent tunneling (SDT) sensors as well as anisotropic magneto-restrictive (AMR), or Hall and 3-D Hall sensors [28] all of which can be compact, are sensitive to small fields or can be configured as gradiometer sensors. These types of configurations, in general, require low power, operate over a very broad range of frequencies and can be readily manufactured and miniaturized using microelectronic processes.

In fact, it is possible to use GMR sensors of similar type to those used in magnetic disc data storage read heads to obtain positional information with resolution in the region of a few tens of nanometers [29, 30]. In principle, there is no reason why this cannot be used for position measurement if incorporated for the detection of currents in a wire patterned for a specific measurement (i.e. radial lines or a periodic meander for rotational or linear measurements). Alternatively, it might be possible to use the read head directly to detect patterned magnetic domains on a reference. The simplest reference would be a periodic grid that could, for example, be written onto a magnetic tape and then adhered to a surface. However, the ease of manufacturing arrays of such sensors readily enables absolute or incremental encoder configurations [31] for grayscale or quadrature type displacement measurements.

2.7 Eddy current

A simple eddy current probe for monitoring displacement between a surface and a circular coil at a distance z above the surface is given by [32]

$$L = \frac{\pi \mu n^2 \bar{r}}{l^2 t^2} \int_0^\infty \frac{I^2(r_2, r_1)}{\alpha^5} \times \left[2l + \frac{1}{\alpha} \left\{ 2e^{-\alpha l} - 2 + \left[e^{-2\alpha l(l+z)} + e^{-2\alpha z} - 2e^{-\alpha(l+2z)} \right] \left(\frac{\alpha - \alpha_1}{\alpha + \alpha_1} \right) \right\} \right] d\alpha, \quad (26)$$

where

$$I(r_1, r_2) = \int_{\alpha_1}^{\alpha_2} x J_1(x) dx, \quad (27)$$

$$\alpha_1 = (\alpha^2 + j\omega\mu\sigma)^{1/2}, \quad (28)$$

t is the difference between the inner and outer radius, r_1, r_2 , of the excitation coil, l is the length of the coil, μ is the permeability of the specimen surface material, σ is the electrical conductivity of the specimen surface material, ω is the frequency of excitation and α and x are dummy variables of integration. Correspondingly, there will be a repulsive force between the surface and coil of value

$$\begin{aligned} F_z &= \frac{i^2}{2} \frac{\partial L}{\partial z} \\ &= \frac{i^2}{2} \frac{2\pi\mu n^2 \bar{r}}{\ell^2 t^2} \int_0^\infty \frac{I^2(r_1, r_2)}{\alpha^5} (e^{-2\alpha z} (e^{-2\alpha t} - 2e^{-\alpha t} + 1)) \left(\frac{\alpha - \alpha_1}{\alpha + \alpha_1} \right) d\alpha. \end{aligned} \quad (29)$$

These equations can be solved numerically to show the dependence of inductance and force with separation. For a flat coil of 40 turns, an inside diameter of 8.0 mm, an outside diameter of 12.0 mm and a wire diameter of 0.2 mm being driven at a frequency of 20 KHz, 0.35 V at a current of 0.70 A, is shown in Figure 4. In most instances, as with capacitance sensors, the forces might be considered insignificant when compared to the size of the sensor.

In practice, the reference surface should be relatively flat, larger than the coil and extend to a depth considerably greater than the skin depth, δ , given by

$$\delta = (2\omega\mu\sigma)^{-1/2}. \quad (30)$$

A disadvantage is that the sensitivity is dependent upon the materials type and it is possible that the eddy currents can cause local heating of the target surface, which must also be an electrical conductor.

Such eddy current based inductive probes can have high resolution and bandwidth and have the additional advantage that electrical connection to the target surface is not necessary. Typical commercial probes will resolve to within tens of nanometers over millimeter ranges. The limit on range is set by the size of the coil or diameter at which the magnetic field is being generated. For example, if a circular coil is used to generate the oscillating fields, it is found that, as a rule of thumb, the field generated by the coil is

substantially attenuated at a distance approximately equal to its mean radius. Consequently, this type of sensor might be considered as a probe with a measurement zone extending a maximum of the coil radius while also requiring a specimen surface of circular dimension twice that of the coil outer diameter.

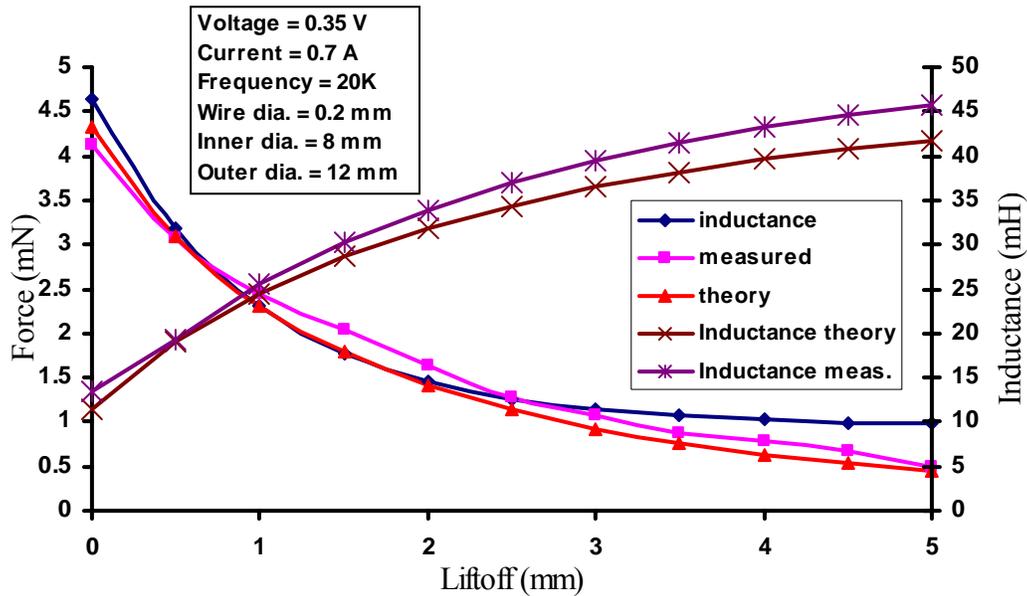


Figure 4: Dependence of inductance and force on separation. Theoretical and measured values are shown.

Alternatives, to inductive coils can be found in the use of sensing elements to detect the magnetic fields produced by the eddy currents. In principle such probes can be made with the sensors suitably placed and biased to measure only the field of the eddy currents independent of the excitation field [33]. Such probes can also be miniaturized to produce localized measurement of surfaces for pattern recognition and, possibly, lateral position measurement [34].

2.8 Proximity probes

A proximity probe is considered to be any sensing method that provides a measurable signal as it is brought into close proximity to a surface. Often proximity probes are used as a means of communicating a translation from a relatively small area on a surface; in a coordinate measuring machine (CMM), stylus profilometer or scanned probe microscope

(SPM) this is the 'probe'. This is then communicated to an actuator that can move the probe and record subsequent displacements. In such systems, it is also apparent that there is still the requirement for a displacement measuring system. One area in which the scaling of proximity sensors has seen extensive development is in the field of SPM sensors. Sensing methods include tunneling, strain gage, near field optical, thermal, ion conductance or capacitance. The relative merits and limitations of these proximity sensors are

1. Small probes can measure surface heights in a small and localized area (sometimes with resolution of better than one atom for the smallest of probe tips).
2. When controlling the probes to maintain a null signal level, as long as the sensing characteristic as a function of tip/sample interaction is stable and repeatable, the accuracy is little influenced by non-linearity.
3. Extremely high resolution and bandwidth measurement of variations in the probe to surface interaction are possible.
4. Most probes provide only proximity information and do not have the intrinsic capability for the measurement of displacement. This is usually transferred to another displacement measurement transducer of the types outlined above.
5. In principle, any of these probes could be used to detect variations of, and therefore represent a sensing method for, a periodic scale.

In the initial scanned probe microscope designs of Young, *et al.* [35, 36], the proximity sensing method was Fowler-Nordheim field emission from a sharp electrode tip and later efforts were attempted to utilize electron tunneling at tip-specimen spacing of less than 2 nm and lower applied voltage levels. Problems with relative vibrations between probe tip and surface prevented scanning until a decade later, Binnig, *et al.* [37, 38], were able to reduce the mechanical noise and achieve atomic scale images on silicon, for a brief review see [39].

A tunneling probe consists of an electrically conductive sharp tip that is brought into close proximity to an electrically conductive surface. With applied voltage differences of a few tenths of a volt, there will be an exponential increase in current as the separation reduces. Typically this will range from picoamps to microamps with separation changes of around one nanometer. Using a logarithmic amplifier it is possible to obtain a sensitivity of many gigavolts per meter over, of course, a very limited range. In practice, such a probe is sensitive to the separation of the outermost asperity of its tip and is therefore extremely localized and, for measuring surfaces that are smoother than the tip geometry, may be considered to measure the proximity at a point on the surface. However, a drawback with tunneling is that it requires electrically conductive surfaces, must maintain nanometer level separations and is susceptible to surface contamination. For these reasons, tunneling probes are not widely used in meso and micro scale applications.

While tunneling probes may not be particularly popular today, because of their spectacular ability to discriminate individual atoms, which in turn, fueled the imaginations of nanotechnology visionaries across a broad range of scientific and commercial disciplines, they were predominantly responsible for the explosive emergence of the scanned probe microscope industry. Among the many proximity-sensing methods that have been developed, by far the most common is the so-called atomic force probe [40]. This also comprises a sharp tip that is brought into close proximity or contact with a point on the surface. However, it is the interaction force that is transformed to a displacement (or, equivalently, strain) and it is this displacement that is then measured to determine proximity. In practice, given sufficient sensitivity it is possible to monitor both non-contact (electric dipole, van der Waals/Casimir, gas squeeze film) and contact (meniscus, elastic) forces with this form of sensor [41]. While, tunneling probes, optical levers, capacitance and (piezo-resistive) strain gaging have been used to determine the force of interaction, in general it is necessary to use a larger surface or volumetric sensing area to determine a measure of the force thereby increasing the footprint of this probe. However, all of these can be readily produced using macro to MEMS and, recently, nano manufacturing processes and therefore such probes can be readily scaled.

3.0 Actuators

There are an almost limitless number of actuator designs that can be employed in a given system using any number of mechanisms. In general, the major figures of merit quantifying the performance of different actuator methods as they apply to a specific machine application include

1. Machine volume/actuator volume.
2. Range/resolution, or precision, requirements.
3. Efficiency, measured as a ratio of power delivered to a process to that supplied to the actuator.
4. Type of coupling between actuator and object to which force is to be applied.
Generally this can be classified as either contact or non-contact.
5. Bandwidth, generally measured by available power-density.

Also at issue is the connection to the motion control system. In particular, there is a concern in design with the removal or reduction of parasitic motion of the driver.

There are relatively few actuators that can be applied to machines at smaller scales. This is rather surprising in view of the scale of biological processes. However, in engineering systems, we are restricted in the selection of power sources and it is considerably more efficient to generate power using macroscopic systems and channel this into the meso-scale machine. Novel power sources include

1. Solar.
2. Electro-rheological fluids [42].
3. RF absorption.
4. Miniature combustion [43, 44].
5. Fuel cells.
6. Biological systems, cellular or life-based.

While interesting, a discussion of these is outside of the scope of this paper and expertise of the authors. However, these are, for the most part, alternative ways of producing electricity that will then be converted back to work through some actuation mechanism. It is these primary actuator mechanisms that form the basis of the discussions in this section.

Assuming that energy sources are available, the main classes of actuator likely to find application in meso- and micro-systems in precision machine design include

1. Piezoelectric
2. Electromagnetic
3. Electrostatic
4. Pneumatic
5. Hydraulic
6. Thermal
7. Shape memory

Thermodynamically, it can be demonstrated that all actuators, in the process of doing work on a system will generate heat. In practice this can severely limit performance of instrument and machine mechanisms. Therefore, actuators such as voice coils and purely electromagnetic or electrostatic actuators can provide highly efficient electromechanical coupling and, therefore, dissipate minimal amounts of heat at the drive. In other cases it may be desirable to locate the actuator remote from the mechanism. In this case it is then necessary to transfer the work mechanically to the machine. This can be very efficiently achieved using push-rods or, possibly, pulleys. In this case, the desirable scaling of the actuator mechanism might also help determine the nature of the work transfer mechanism. For example, because electromagnetic actuators favorably scale to large devices, it might be desirable to use one large motor and use this to drive multiple small-scale machines, just like the power sources in factories in the late nineteenth and early twentieth century!

3.1 Work and power density

The limits on work and power that can be achievable for an actuator constrained within a specific volume (referred to as the *work density* and *power density*), will be considered in terms of what is possible in an ideal system, initially ignoring the issues of implementation. In some cases, particularly aerospace based, the power or work per unit mass may be of interest to the designer. For some of the actuators it will be obvious that other components required for implementation must share the space of the actuator mechanism for it to be exploited, thereby reducing the theoretically achievable densities. Generally the work and power density will be defined as

1. Work density is the maximum amount of work that can be done per unit volume occupied by the actuator*.
2. Power density is the product of maximum work density and bandwidth response.

A very important consideration in the following discussions is that maximization of the actuator work and power density will be *a matter of implementation* and, because simply scaling down from macroscopic or up from microscopic might not be feasible, new approaches may be required. These issues will also be discussed throughout this section with a compilation of work densities for various actuators contained in Table 1.

Table 1: Indicating approximate power densities for a range of actuator types

Actuator type	$\frac{U}{V} \times 10^6$	Eqn.	Comments
Hydraulic	10	P	Pressure
SMA	6	$\sigma\varepsilon$	Cyclic binary
Solid-liquid phase change	5	$\frac{1}{3} \frac{\Delta V}{V} k$	Water-8% acetimide

* Some other researcher's may refer to this as the energy density of an actuator.

Gas expansion (thermal & press.)	1	P	Pressure
Thermal expansion	0.5	$E(\alpha\Delta T)^2 / 4$	200 K temp. change
Electromagnetic	0.4 – 0.02	See text	Variable reluctance motor – 0.25 mm ³
Electrostatic	0.1 - 0.004	See text	Ideal to MEMS comb-drive estimates.
Piezoelectric	0.01 - 0.05	$(d_{33}\mathbf{E}_{\max})^2 E / 2$	PZN
Muscle	0.02	NA	350 kPa at 10%

3.2 Piezoelectric

At macroscopic scales piezoelectric actuators [45, 46] are often used in the form of stacks [47] (to reduce the necessary drive voltage) or with lever mechanisms to increase the range of motion [48, 49]. Single element ceramics have limited range, but losses due to hysteresis can be substantially smaller than stacked orientations. Single element piezoelectric materials have most notably been used in scanned force microscopy [50, 51]. Lever mechanisms can take the form of bimorph type actuators [52] or mechanical linkages. Ignoring such issues for the time being, consider two different actuator implementations; in the first, it will be assumed that the actuator is incorporated into a flexure type mechanism that provides a resistance force proportional to the actuator displacement. In this case, the bandwidth response is going to be limited by the amount of work that can be done by the actuator on the flexure system. Ignoring any preload, it is possible to model the system as an actuator, of stiffness $k_p = EA/L$, that pushes against a flexure of stiffness k_f . In practice, the stiffness of the flexure will reduce the range of motion of the actuator when in its ‘free’ or lightly preloaded state. Denoting the free

motion range of the actuator, x_i and the output motion in the presence of the opposing stiffness x_o it is relatively simple to show that they are related by the equation

$$\frac{x_o}{x_i} = \frac{k_p}{k_f + k_p}. \quad (31)$$

Work done, U , in a linear translation of the flexure system is

$$U = \frac{1}{2}k_f x_o^2 = \frac{1}{2} \frac{k_f k_p^2}{(k_f + k_p)^2} x_i^2. \quad (32)$$

It can be shown that this is a maximum when the two stiffness values are the same, hence

$$U_{\max} = \frac{1}{2}k_f x_o^2 = \frac{k_p}{8} x_i^2 = \frac{EA}{8L} x_i^2. \quad (33)$$

The maximum free strain of the actuator is a product of the maximum electric field, \mathbf{E}_{\max} applied to the material and the coupling constant in the direction of the electrodes usually denoted by the symbol d_{33} in units of $[\text{m}\cdot\text{V}^{-1}]$. Multiplying strain by the length of the actuator, the maximum free translation of the actuator is given by

$$x_l = d_{33} \mathbf{E}_{\max} L. \quad (34)$$

Substituting (34) into (33) and dividing by the volume of the actuator ($\mathcal{V} = AL$), the maximum work density for a flexure driven system is

$$\frac{U}{\mathcal{V}} = \frac{E(d_{33} \mathbf{E}_{\max})^2}{8}. \quad (35)$$

Considering this structure as a positioning stage, effectively doing no work on the outside world, the major performance measure is likely to be response time.

In many applications it might be desired to use the actuator for controlled work on a specimen for machining or controlled distortions. In this case the actuator will be driving against a nominally constant load. Under these circumstances the external work done on

the specimen will be the product of force and extension and these will in turn be limited by the maximum pressure loads, σ_{\max} , that can be sustained while maintaining a finite strain coefficient. For example, the PMN-PT electrostrictive ceramic can reliably achieve strains of up to 0.1% with pressure loads up to 100 MPa and applied fields of $2 \text{ MV}\cdot\text{m}^{-1}$. At this load, the achievable free strain has reduced by around 30 %, while the elastic modulus is a little over 100 GPa. Using similar analysis, it can be shown that the work done per unit actuator volume is given by

$$\frac{U}{V} = \sigma_{\max} d_{33} \mathbf{E}_{\max} \approx \frac{E d_{33} \mathbf{E}_{\max}}{1000}. \quad (36)$$

It should be mentioned in passing that larger strains could be obtained using an electrode pattern corresponding to an excitation field perpendicular to the polarization axis that results in a shear mode deformation. The drawback with this is the additional cost associated with poling of the ceramics.

To develop long-range systems with a power density on the order of the piezoelectric stack, accumulation type actuators have been constructed [53, 54]. Common forms of this type of system can be found in Picomotors™ and Inchworm™ drives where a ceramic actuator is used as a means of stepping or walking along a guide rail. Other useful applications for piezoelectric actuation include multiple degree-of-freedom structures [55, 56] and closed-loop controlled error compensation systems [57, 58].

3.3 Electromagnetic

Electromagnetic actuators are a little more difficult to quantify [59] in terms of power density. In general, there are two designs for force actuation involving the magnetization of ferromagnetic materials. The first uses electrical currents to produce attractive forces between two solids and the second uses permanent magnets to produce a magnetic flux through current carrying wires. Magnetic field driven actuators can be classified as electromagnetic, electrodynamic, magnetostrictive and magnetorheological as discussed in more detail by Janocha [60]. In general, the force, \mathbf{F} , on the moving charge of strength q moving with a velocity, \mathbf{v} , can be obtained from the Lorentz equation

$$\mathbf{F} = q(\mathbf{E} + \mathbf{B} \times \mathbf{v}), \quad (37)$$

where \mathbf{E} is the electric field and \mathbf{B} is the magnetic field. For most systems this equation must be integrated over all interactions to determine the force available to do work. One of the most common methods for deriving known forces is to arrange for the currents to pass through wires in a fixed magnetic field often created by a permanent magnet. Such devices are often lumped into the generalized category known as voice coils. In such systems, in which the charge distribution and velocity is common throughout the length of the wire in the field, the integral over all charges can be expressed in the scalar form

$$F = Bli \sin(\theta). \quad (38)$$

In equation (38) the product Bl represents the integral of the field values over the length of the wire of the voice coil and i is the applied current. This force is clearly maximized when the permanent field and the currents are perpendicular. In principle, the force could be expressed as the flux from the permanent magnet multiplied by the maximum current density, j_{\max} transmitted through a given area over a given length. Considering that it is possible to derive sufficient currents, it should be possible to utilize the maximum energy stored within a given magnet material. In general this is given by the maximum product of flux density and applied magnetic field which is a direct measure of the energy storage per unit volume given by

$$\frac{U}{\mathcal{V}} = \frac{(BH)_{\max}}{2}. \quad (39)$$

Unfortunately, a method for implementation of such a device is not readily apparent and in practice it is desired to create forces using a current carrying wire occupying a gap of part of a magnetic circuit. Again, an ideal system may be envisaged in which the gap flux, B_g is related to the energy density of the permanent magnet by the relationship

$$B_g = \left(\frac{(BH)_{\max} \mathcal{V}_m}{\mathcal{V}_g} \right)^{1/2}. \quad (40)$$

The volumes in this equation represent the magnet and gap respectively. Generally, this gap flux could be used to provide the magnetic force on a current carrying wire. From (38), the force on a wire entirely immersed in the field can be approximated by

$$F = B_g li = B_g l A_g j_{\max} = \left(\frac{(BH)_{\max} \mathcal{V}_m}{\mathcal{V}_g} \right)^{1/2} l A_g j_{\max} \quad (41)$$

Dividing this expression by the total volume ($\mathcal{V} = \mathcal{V}_g + \mathcal{V}_m$) yields the maximum force

$$\frac{F_{\max}}{\mathcal{V}} = \left(\frac{(BH)_{\max} \mathcal{V}_m \mathcal{V}_g}{(\mathcal{V}_g + \mathcal{V}_m)^2} \right)^{1/2} j_{\max} = \left(\frac{(BH)_{\max} k}{(1+k)^2} \right)^{1/2} j_{\max}, \quad (42)$$

where

$$k = \frac{\mathcal{V}_m}{\mathcal{V}_g}. \quad (43)$$

Noting that this is a maximum when $k = 1$ and assuming that the force on the wire will reduce linearly as it moves out of the volume at a displacement l , the work done per unit volume may be expressed in the form

$$\frac{U_{\max}}{\mathcal{V}} = \frac{j_{\max} ((BH)_{\max})^{1/2}}{4} l. \quad (44)$$

This rather over simplified view of a voice coil drive ignores a number of complicating issues such as recoil permeability, temperature and field interactions among many others. In many voice coil actuator designs, the flux is often concentrated using soft iron cores. While this will increase force, it will correspondingly limit the range and, ultimately, the two effects tend to cancel in the work density picture. However, the key feature of this equation is the linear scaling of work density with dimension. This is illustrated by the graph in Figure 5, in which the work (product of force and stroke) that can be obtained from commercial voice coil actuators divided by their volume is plotted as a function of the actuator length in this case considered to indicate size of actuator.

Voice coil actuators have been successfully designed and utilized in a number of

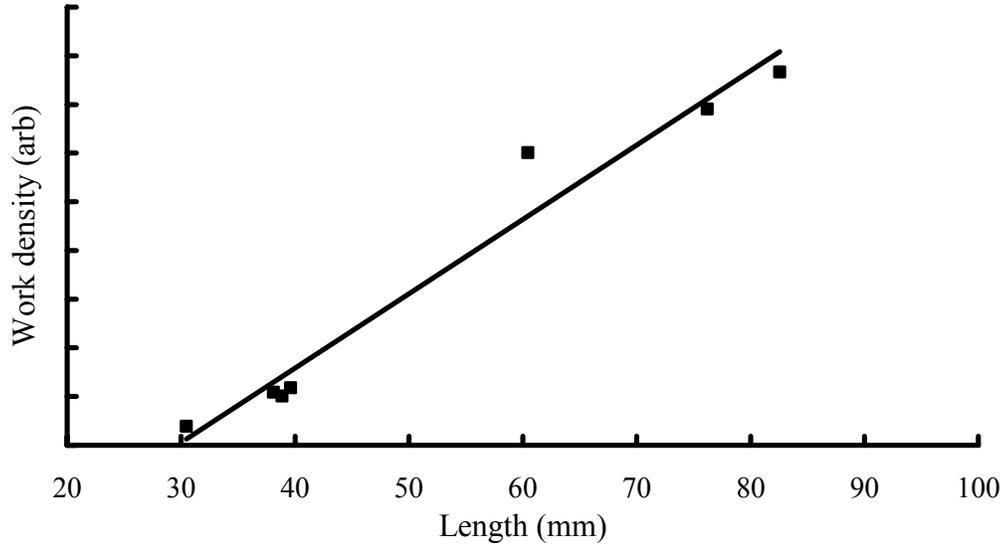


Figure 5: Work density plotted against length of commercial voice coil actuators (details available on request)

precision applications including machine tools [61] and displacement instruments [62].

The solenoid actuator design is not dissimilar to the voice coil already considered (in fact most voice coil actuators combine permanent magnets and soft magnet materials to guide and concentrate the flux). In this system, the field is generated by the coil and flows along the permeable path of a core material with the moving component changing the reluctance, \wp , of the loop of magnetic flux. In general, a solenoid of length, l_s provides the field resulting in a flux loop passing through a total of K components of the magnetic circuit and having a magnitude given by

$$\Phi = \frac{NI}{\sum_k \wp_k} = \frac{NI}{\sum_{k=1}^K \left(\frac{l_k}{\mu_k A_k} \right)}. \quad (45)$$

By definition, the inductance is the flux passing through the N turns divided by the current required for its generation i.e.

$$L = \frac{N\Phi}{I} = \frac{N^2}{\sum_k \left(\frac{l_k}{\mu_k A_k} \right)}, \quad (46)$$

where l_k , A_k and μ_k correspond respectively to the effective length, effective area and permeability of the k^{th} component in the flux loop. Again, a simple system will be considered consisting of a permeable core denoted using a subscript c (usually ‘E’ shaped) with a moving component constructed of the same material. These two components are considered to be separated by an air gap where the magnetic and geometric properties of which are given a subscript a . In general the magnetic potential, U , of this system is given by

$$U = L \frac{i^2}{2} = \frac{N^2}{\left(\frac{l_E}{\mu_c A_E} + \frac{l_m(x, y)}{\mu_c A_m(x, y)} + \frac{l_g(x, y)}{\mu_a A_g(x, y)} \right)} \frac{i^2}{2}, \quad (47)$$

where the subscripts E , m and g represent the stationary core, moving component and gap between them, respectively.

Being a general expression for the potential of the systems, the vector components of force in the x y plane can be obtained from

$$\mathbf{F} = -\text{grad}(U) = -\left(\frac{\partial U}{\partial x} \mathbf{i} + \frac{\partial U}{\partial y} \mathbf{j} \right). \quad (48)$$

Usually, by design, it is arranged for the moving component of the actuator to operate with a changing separation at constant area, x motion, or vice versa for y motion.

Substituting (47) into (48), each of these actuators produces a force given by

$$F_x = -\frac{\partial U}{\partial x} \approx -\frac{\partial U}{\partial l_g} = \frac{N^2}{\mu_a A_g \left(\frac{l_E}{\mu_c A_E} + \frac{l_m}{\mu_c A_m} + \frac{l_g}{\mu_a A_g} \right)^2} \frac{i^2}{2}, \quad (49)$$

$$F_y = -\frac{\partial U}{\partial y} \approx -b \frac{\partial U}{\partial A_g} = \frac{bN^2 l_g}{\mu_a \left(\frac{l_E}{\mu_c A_E} + \frac{l_m}{\mu_c A_m} + \frac{l_g}{\mu_a A_g^2} \right)^2} \frac{i^2}{2}. \quad (50)$$

Again, considerable license with assumption of the parameters has been used to derive these expressions and there are other more precise modeling approaches that may be applied for specific applications and materials. However, a few general points become clear from these equations. Being proportional to the square of the current, the force between the two bodies is always attractive independent of the current polarity. Also the force varies approximately with the inverse square of the motion. To assess the effect of scaling, consider equation (49) where the geometry of the actuator is fixed with a reasonably large gap and in the case when the permeability of the core is considerably higher than that of air while the geometric factors are of comparable value. Under these circumstances equation (49) reduces to

$$F_x = -\frac{\partial U}{\partial x} \approx -\frac{\partial U}{\partial l_g} = \frac{\mu_a A_g (Ni)^2}{l_g^2} \frac{1}{2}. \quad (51)$$

Again, using similar arguments as before, the product of turns and current represents current density. This will scale with area of the winding. Additionally, it has rashly been assumed that the permeability is a constant independent of the field. In practice, this will be limited by the saturation field that has, in turn, been shown to have a maximum potential corresponding to a high $(BH)_{\max}$ on the $B-H$ curve for many of the core materials commonly selected for solenoid applications [63]. Simple dimensional analysis on equation (51) produces a work density scaling of the form

$$\frac{U_{\max}}{\forall} = \frac{\mu_a A_g (Ni)^2}{l_g^2} \frac{l}{2} \frac{1}{\forall} \propto (BH)_{\max} \frac{(j_{\max} l^2)^2}{2} \frac{1}{l^2} \propto (BH)_{\max} \frac{j_{\max}^2}{2} l^2. \quad (52)$$

Comparing equation (44) with the work density proportional to the square of the dimensions of a device, it is suggested that such systems scale even less desirably than voice coil designs.

In both of the above approaches, it is apparent that, based on the simplified model, the force scales with maximum current density. In practice, it is resistive heating and the subsequent heat dissipation (usually being transferred by conduction from the surface of the outside of the coil) within the actuator that imposes a limit on the current density and, therefore, the work available to the system. In general, because the ratio of surface area to volume increases with reduced size, it is possible that the limiting current density will be increased with reduced size, thereby offsetting, to some extent, the unfavorable scaling indicated by the above analyses.

In general, based on the above considerations, it is not surprising that electromagnetic actuators have found few applications in meso and micro-scale devices. However, this is not to say that applications do not exist [64, 65]. One has only to disassemble digital watches with mechanical ‘hands’ to find meso-scale examples of both types of actuators.

3.4 Electrostatic

Electrostatic actuators have been used extensively for providing forces in MEMS systems [66,67]. This, in part, provides a clue as to the favorable, or, at least, competitive, scaling of such a mechanism. One particular advantage of electrostatic over piezoelectric actuators is that electrostatic forces can be non-contact, utilizing air or fluids as the dielectric material. When they are capable of providing sufficient force to produce the desirable displacements, this makes them favorable for continuous motion systems and, for positioning systems. In general, the force is derived from the gradients of electrostatic potential between two or more electrodes. Taking the simplest instance of two parallel plate electrodes of square shape of length L and separation distance x , the potential energy, U , as a function of applied voltage V is approximately given by

$$U = \frac{1}{2} CV^2 = \frac{1}{2} \frac{\epsilon(L-y)(L-z)}{x} V^2, \quad (53)$$

where x , y and z represent the possible linear translations of one of the electrodes.

Considering the case of variable separation (i.e. $y = z = 0$), the force on the electrodes in the direction normal to the plane of the electrodes is

$$F_x = \frac{\partial U}{\partial x} = \frac{-\varepsilon L^2 V^2}{2x^2}. \quad (54)$$

The maximum force is clearly a function of the maximum voltage that can be applied to the electrodes and this will, in turn, be limited by the maximum electric field, \mathbf{E}_{\max} that can be sustained by the dielectric. Consequently, the maximum voltage that can be utilized will vary with x according to

$$V_{\max} = \mathbf{E}_{\max} x. \quad (55)$$

Assuming that the actuator operates at this voltage, the maximum work that can be provided by this actuator is given by

$$U_{\max} = \int_{x_{\min}}^{x_{\max}} F_x dx = -\varepsilon \int_{x_{\min}}^{x_{\max}} \frac{L^2 V_{\max}^2}{2x^2} dx. \quad (56)$$

In the above equation, it is assumed that the dielectric constant and field strength are both constant. However, it is known that the dielectric strength can increase at reduced separation. For example, in air the field strength is known to increase dramatically as separation between electrodes reduces to below a few micrometers (a plot of field strength as a function of the separation is sometimes referred to as the Paschen curve). Substituting (55) into (56) and rearranging, the maximum work is given by

$$U_{\max} = \frac{\varepsilon L^2 \mathbf{E}_{\max}^2 (x_{\min} - x_{\max})}{2}. \quad (57)$$

The maximum volume of the actuator is

$$\mathcal{V} = L^2 x_{\max}, \quad (58)$$

from which the maximum work density is given by

$$\frac{U_{\max}}{\mathcal{V}} = \frac{\varepsilon \mathbf{E}_{\max}^2}{2} \left(1 - \frac{x_{\min}}{x_{\max}} \right). \quad (59)$$

Clearly, the work density is limited by the translation range which will, in practice, be limited at one end by the maximum voltages that can be generated and at the other by the ability to measure and control small separations and/or actuator geometry.

There is another limitation imposed on the minimal separation that occurs in many applications due to the non-linear stiffness of the actuator at a given applied voltage. In some cases, the actuator will be used to drive a flexible mechanism that itself will have a finite and, generally, linear stiffness k_f . However, the stiffness of the electrostatic actuator will be a function of the separation, x , given by

$$k_E = \frac{\partial F_x}{\partial x} = \frac{\epsilon L^2 V^2}{x^3}. \quad (60)$$

Subtracting the work done by the actuator from the work done on the flexure mechanism gives the work available to drive external processes i.e.

$$U_{ext} = U_f - U_E = k_f(x - x_o) - k_E x. \quad (61)$$

Differentiating (61) yields

$$F_{ext} = \frac{\partial}{\partial x} U_{ext} = (k_f - k_E)x = \left(k_f - \frac{\epsilon L^2 V^2}{x^3} \right) x. \quad (62)$$

Noticeably, there will be a point in x after which the force term of equation (62) changes sign. At this point, any small perturbation in the separation is likely to result in an unstable, and sometimes catastrophic motion of the electrodes towards each other. In the absence of mechanical stops, this motion will continue unabated until the electrodes come into contact. This is often referred to as the 'jump-to-contact' condition at a separation, representing a theoretical minimum separation, given by

$$x_{jump} = \left(\frac{\epsilon L^2 V^2}{k_f} \right)^{1/3}. \quad (63)$$

If there is no feedback to determine the maximum possible voltages that can be applied, assuming that this is greater than the 'jump-to-contact condition', it is common to use a

maximum supply voltage that can be tolerated at the minimum mechanical separations encountered in practice. Under this condition the maximum applied voltage is fixed and the work done by the actuator is given by

$$U_{\max} = \int_{x_{\min}}^{x_{\max}} F_x dx = -\epsilon L^2 V_{\max}^2 \int_{x_{\min}}^{x_{\max}} \frac{1}{2x^2} dx = \frac{\epsilon L^2 V_{\max}^2}{2} \left[\frac{1}{x_{\max}} - \frac{1}{x_{\min}} \right]. \quad (64)$$

Assuming that the minimum separation (or maximum voltage) is field strength limited, we may assume

$$\frac{x_{\max}}{x_{\min}} = \text{const} = c, \quad (65)$$

where

$$x_{\min} = \frac{V_{\max}}{\mathbf{E}_{\max}} \quad (66)$$

Substituting equations (65 and 66) into (64), dividing by the maximum actuator volume and rearranging gives

$$\frac{U_{\max}}{\mathbb{V}} = \frac{\epsilon L^2 V_{\max}^2}{2x_{\min} c} \left[\frac{1}{c} - 1 \right]. \quad (67)$$

The maximum of (67) is given when

$$c = 2. \quad (68)$$

After which the work density can be expressed by the equation

$$\frac{U_{\max}}{\mathbb{V}} = \frac{-\epsilon \mathbf{E}_{\max}^2}{8}. \quad (69)$$

This is also independent of scaling.

Consider now the alternative mode of operation in which the electrode is constrained to move in a direction parallel to the plane of the electrodes. In this case, only translations

in the y direction will be considered (i.e. $z = 0$) so that x is a constant, given as x_o . An interesting consideration with such an approach is that, in theory, it should be possible to overcome the jump to contact by using a ‘comb’ of electrodes. In this the forces and successive derivatives, in the direction perpendicular to the plane will compensate, thereby eliminating the problem of jump to contact. In practice this problem will still exist in the presence of small perturbations and/or asymmetries due to manufacturing tolerances and other sources. However, because the electrodes are constrained in this direction by design, the flexure stiffness is likely to be considerably higher, thereby reducing the possible separations for a given applied voltage. In this mode of operation, the force on the electrodes is given by

$$F_y = \frac{\partial U}{\partial y} = -\frac{\epsilon L}{x_o} V^2. \quad (70)$$

Again substituting for the maximum voltage as determined from the field strength of the dielectric gives

$$F_y = \frac{\partial U}{\partial y} = -\epsilon L \mathbf{E}_{\max}^2 x_o. \quad (71)$$

It is readily apparent that this is independent of displacement and therefore such an actuation mode contributes zero stiffness in the drive direction. Such a simplistic model enables the force to be applied over a complete traverse of the electrode over a distance $2L$. Hence, ignoring edge effects and other complications, the work density is given by

$$\frac{U_{\max}}{\forall} = \frac{2\epsilon L^2 \mathbf{E}_{\max}^2 x_o}{L^2 x_o} = 2\epsilon \mathbf{E}_{\max}^2. \quad (72)$$

In this case it is apparent that the work density is, yet again, independent of scaling.

Recently, there has been a growing interest in the production of polymeric layered electrostatic actuators that can have enhanced dielectric coefficients and therefore increase the achievable work density. The relatively large strains that are possible with these actuators have lead to considerable research towards the manufacture of artificial muscle and these integrated force arrays produce comparable strains of approximately

1/10th of the work density of human muscle [68]. More recently, some ionic polymer-metal composites have demonstrated work densities comparable to human muscle [69].

3.5 Electrostrictive

Electrostriction is a phenomena, similar to the piezoelectric effect, which gives rise to a strain proportional to the square of the applied electric field [70] as seen in lead magnesium niobate (PMN)-based relaxer ferroelectric materials [71]. A comparison of piezoelectric, electrostrictive and phase-change materials has been given by Uchino [72] and Cross [73]. Electrostriction polymers [74] have been shown to produce actuation with strains near 4.0%. These actuators tend to be highly non-linear due to their dependence on the square of the applied voltage, but may be less hysteretic at a given temperature than the more common lead-zirconate-nitrate PZT materials. Electrostrictive materials also tend to have 3 to 4 times the capacitance of piezoelectric materials, which may reduce their effectiveness for dynamic applications.

3.6 Magnetostrictive

Magnetostrictive actuators found a resurgence with the discovery of Terbium-Iron-Dysprosium alloys commonly called terfenolTM that exhibit a relatively large magnetostrictive effect. For this material, application of a magnetic field causes the internal magnetic domains to rotate resulting in an expansion strain that can be up to around 0.1% at magnetic fields of around 40,000 amp turns per meter. This strain is significantly affected by mechanical pre-stresses that tend to align the internal magnetic domains along the crystal. Maximum energy densities of around 25 kJ·m³ are possible with this material and this appears to remain approximately constant with scale. In practice actuators typically comprise a plate or rod of the metallic crystal surrounded by a solenoidal coil. Because of the similar strains produced by this actuator, macroscopic devices tend to have a performance and footprint comparable to piezoelectric actuators.

Scaling of these actuators is difficult to assess. One clue to this can be obtained by examining commercially available systems [75]. In these, the winding tends to occupy a considerable proportion of the actuator volume. In principle this should lead to a constant

work density. In practice, as coils get progressively smaller, there are substantial problems associated with the volume occupied by insulation and air gaps (referred to as the packing factor of a coil) and this will ultimately limit the fields that can be generated as scale reduces. Such considerations also apply to electromagnetic actuators.

3.7 Hydraulic

This and the following category, represent a form of fluid power. One key attribute of hydraulic systems is that the work can be arbitrarily distributed through a system given appropriate piping and control valves. This appears to be the major obstacle. In the absence of such concerns, the maximum work density can be considered to be that provided by filling a fixed volume with a fluid of pressure P . From basic thermodynamic considerations, the maximum work density is

$$\frac{U}{V} = \frac{\int PdV}{V} = P. \quad (73)$$

One of the problems associated with such a simplistic model for possible power density is that of controlling the work delivered by the hydraulic actuator. In macroscopic systems, this is achieved via discrete valves that are used to direct the pressurized fluid to either side of a hydraulic piston. Typically, spool valves are used and these rely on precision manufacture to provide the sealing action often with accuracies of the order of tens of micrometers or better. In turn, the spools of these valves are actuated either by solenoids or the moving components of the hydraulic actuators themselves. Normally, the rate of extension or retraction of the piston is controlled by a restrictor at the outlet port of the actuator. When the piston is reversed the role of the ports either side of the piston are switched and a non-return valve is necessary to enable free flow of high-pressure fluid bypassing the restrictor. Although possible, it is unlikely that manufacturing processes will be developed to scale these systems to mesoscopic dimensions and it might become inefficient to use solenoids as the main actuator.

In all cases, the components for control of hydraulic systems can be thought of as providing the functions of restriction or flow redirection. In essence a restrictor is a two-port device while a redirection device requires a minimum of three and more typically

four ports (one for the drive (inlet), one for the exhaust (outlet) and one (two) to which the flow is to be directed). In many cases the outlet port represents a return pipe to the fluid reservoir. All of these functions can be achieved through suitably controlled arrays of restrictor mechanisms capable of continuous operation from full flow to complete restriction.

While macroscopic systems have been evolving for over a hundred years and have been incorporated for precision applications [76], it is difficult to predict what devices will be best suited for flow control as dimensional scales reduce. At meso-scales, it is still likely that the fluid power will be supplied from some central source where elaborate plumbing systems will be used to route the fluid to a specific location, not unlike biological organisms.

A brief mention regarding microfluidic devices; microfluidic systems have been developed for use as pumps, valves and flow sensors [77]. At this time, it appears that the majority of these devices are being applied in chemistry and micro-biology studies as illustrated by Lab-on-Chip (LOC) devices. There have been two major focus areas of microfluidic design as mentioned by Andersson and van den Berg [78]. The first looks at applying fluidic components (pumps, valve, etc.) into fluid based systems, while the second looks at miniaturizing analytical chemical methods, such as the ability to sort and manipulate single cells. Pumps and valves are currently being used to mix fluids, pump fluids and sort different micro-structures [79, 80] at the microscopic scale. The functionality of these types of systems at the cellular level has and will continue to promote single cell manipulation and treatments. It is clear that these kinds of applications will continue as micro- and/or nanotechnology applications become more prevalent.

3.8 Pneumatic

Pneumatic based actuators have the advantage of a readily available source (air) and the exhaust can be freely vented. In many aspects the relative advantages mirror those of hydraulic based actuators with only one major exception being that the pressures attainable with gases are typically less than those achievable with a fluid. Another

limitation is the compressibility that leads to thermal changes, and therefore losses, upon compression. This is commonly used to explain the Carnot cycle in introductory thermodynamics, although, in practice all actuators can be modeled by the Carnot cycle [81].

Typically actuators might be in the form of a piston, diaphragm or bellows type of mechanism. In all cases, the method of constraining the gas and enabling its expansion with pressure will occupy space thereby reducing the work density of the actuator.

One of the major problems is that of redirecting the flows of pressurized air. To do this requires valves and restrictors for full motion control. At the meso-scale, it has been possible to use fluidic devices to control airflow at moderate pressure differences (around 0.25 bar). These use the boundary layer effect to guide air flows along different paths within a valve enabling common logic operations such as ‘and’, ‘or’ and flip-flop control of pressure signals.

As the scale reduces it is necessary to correspondingly reduce the scale of the valves and flow control device. There has been considerable research and development of thermo-pneumatically controlled valves [82, 83], which may be used at the meso-, micro-scales since thermal time constants scale favorably thus increasing operational bandwidths at small-scales. Typically, the valve is opened or closed by deflecting a diaphragm against an orifice using the thermo-pneumatic actuator. This actuator consists of a fluid (Fourier) that is contained in an enclosure with one of its walls being a flexible diaphragm. Also within this enclosure is a heating element that is used to heat and therefore expand the fluid. Movement of this diaphragm can then be used to open and close a valve or can be used to operate a peristaltic type pump, in the reference cited to generate pressures of up to $30,000 \text{ N}\cdot\text{m}^{-2}$, for examples see [84, 85].

3.9 Thermal

At macroscopic scales, thermal actuators tend to be too slow and of limited displacement range for most applications. However, when bandwidth is not important, such actuators can provide a smooth and continuous action that has been used for scanning mirrors in

optical interferometers and other fine motion applications. Typically, ignoring the thermo-pneumatic actuators mentioned above, the actuator comprises a simple monolithic element heated either by surrounding it with a heating wire, induction coil or, if it conducts electricity, directly by passing a current through it. While other sources of heat are possible (including chemical potential from interactions or burning of fuels) electrical heating provides more controlled power. One of the main advantages of a thermal actuator is its simplicity sometimes requiring little more than a pair of electrical connections.

In its simplest form, a thermal actuator can be made from a cube of material of length L . While the cube will typically expand linearly with an increase in temperature, it will generally be used to provide a linear motion in a defined direction. For relatively small changes in temperature, it may be assumed that the volume expansivity is constant resulting in a dimensional change

$$\delta x = \alpha L \Delta T, \quad (74)$$

where α is the coefficient of thermal expansion and ΔT is the change in temperature (usually taken from a nominal 20°C). An estimate of work that can be obtained from a thermal actuator is complicated by a corresponding variation in the elastic modulus with temperature as well as variation in the expansion coefficient with temperature [86]. The change in energy of the actuator system can be expressed by

$$dU = mC_p \Delta T - \int_V p dV, \quad (75)$$

where m is the mass, C_p is the specific heat, p is the pressure and V is the volume. More generally, the first term in this equation represents the entropy while the second term represents work done by the actuator. Again, for small temperature changes it is reasonable to assume that the actuator will operate at a nominal elastic modulus and thermal expansion coefficient given by the average between the start and finish value. Again, assuming that the actuator drives a stiff load, the strain energy stored is given by

$$U = \frac{1}{2} k_f x^2 = \frac{1}{2} \frac{EA}{2L} (\alpha L \Delta T)^2 = \frac{EAL}{4} (\alpha \Delta T)^2. \quad (76)$$

Hence the energy density is given as

$$\frac{U}{V} = \frac{1}{4} \frac{EA}{L^2} (\alpha\Delta T)^2 = \frac{E}{4} (\alpha\Delta T)^2. \quad (77)$$

The speed of response of the thermal actuator can be represented by the diffusivity. In general, the time constant, τ , of a thermal system is given by

$$\tau = \frac{k}{L^2 \rho C_p}, \quad (78)$$

where k is the conductivity and ρ is the density. Clearly, this increases with the reciprocal square of the linear dimension, thereby indicating a very favorable dynamic benefit for small systems.

3.10 Shape memory

In practice, shape memory actuators are really a subset of thermal actuators. Shape memory alloys (SMA) were first developed in the early 1950s and have been used for a myriad of mechanical devices, such as couplings, sensors, springs and actuators [87]. Because of the vast array of processing techniques available, SMA actuators can take on any number of forms. Research has been done using SMAs as biomimetics [88] and in medical devices [89] where fluid environments increase heat transfer rates due to convection and stimulate actuation. They have also been used as actuation mechanisms in robotics [90] and been combined with other materials such as elastomers [91] to increase response time and controllability. Thin film actuators for MEMS type applications provide high work to volume ratios and can have bandwidths up to 100 Hz [92] given that the rate of heat transfer is increased due to the large surface area to volume ratio. SMA materials can also be utilized for damping in structures because of relatively high internal friction. There are a number of parameters that could affect the SMAs damping capacity. These include material type, grain size, martensite density and defect structure. SMA materials such as Cu-based alloys and Ni-Ti have been shown to have damping capacities an order of magnitude larger than common steels, which may prove to be critical in some precision applications [93].

The shape memory effect (SME) is a property of certain materials whereby deformations attained at low temperatures can be removed by heating the material. For some materials strains of up to nearly 10% can be completely removed upon heating above the transformation temperature. This type of transition is called one-way shape memory effect. Two-way shape memory effect allows the SMA to return to a preset shape upon heating and then return back to its original shape upon cooling. The SME process is fundamentally related to phase transformation of the material, for most shape memory alloys, from austenite to martensite and vice versa. These materials also have another interesting and potentially useful property of superelasticity. Superelastic SMAs can sustain strains of 8-10 % at a reasonably constant stress. This makes SMAs useful as constant force springs, for applications such as preload mechanisms for precision actuators [94], provided that the strain levels stay within the superelastic limit. There are three primary variables that affect the stress behavior of the SMA, ξ the fraction of martensite, T the operating temperature and ε the strain, which give $\bar{\sigma} = \bar{\sigma}(\bar{\varepsilon}, T, \xi)$, where $\bar{\sigma}$ is the Piola-Kirchhoff stress and $\bar{\varepsilon}$ is the Green strain. A common form of the constitutive equation is given by the following;

$$\dot{\bar{\sigma}} = D\dot{\bar{\varepsilon}} + \theta\dot{T} + \Omega\dot{\xi}. \quad (79)$$

Taking the integral of equation (79) with respect to time, yield the following relation;

$$\bar{\sigma} - \bar{\sigma}_o = D(\bar{\varepsilon} - \bar{\varepsilon}_o) + \theta(T - T_o) + \Omega(\xi - \xi_o), \quad (80)$$

where D is Young's modulus, θ is the thermoelastic tensor and Ω is the transformation tensor as given by Liang and Rogers [95]. A more complete description of the details of shape memory effects (SMEs) and superelasticity are presented by Otsuka and Wayman [96] and Srinivasan and McFarland [97]. To develop the amount of work available per unit volume of SMA actuator, the strain energy of the system can be calculated using the straightforward relationship;

$$U = \int \sigma d\varepsilon, \quad (81)$$

in units of $[J/m^3]$. Considering a stress of 300 MPa and a 2.0% strain operating in the superelastic region where the stress is nearly constant over the strain range, the work density of a SMA actuator is on the order of $6.0 \times 10^6 J \cdot m^{-3}$. This simplified representation of the work density is a useful tool for describing the material behavior and possible application of such a material as an actuation mechanism.

Concluding remarks

The actuator and sensor methods outlined above represent a large number of design options. It is hoped that by highlighting the relative merits and limitations of these devices, future designers might be able to narrow their choices more rapidly. As functional tolerances and overall dimensions become smaller and more attainable, scientists and engineers will undoubtedly be expanding and reinventing some, if not all, of the topics discussed throughout this paper. It is the wish of the authors that this somewhat limited review provides a starting point for the plethora of research and development required before fundamental physical limits are reached for a large number of systems.

Clearly, as we enter into the design of ever shrinking mechanical systems it will be the job of the designer to make choices based on performance specifications **and** their scale dependence. It is hoped that such data will be collected in a systematic and unified format, and any gaps in knowledge ‘filled in’ by future researchers. As stated earlier, we believe that it will be the implementation of these technologies that will determine their relative success in the marketplace.

We are also aware that there can be other significant factors, particularly economic, that might drive the widespread use of one technology over another. As smaller scale systems are developed so too will manufacturing processes. Subsequently, with growing markets there will be accompanying economies of scale. For this reason, our attention is limited to the physics of the devices rather than technology barriers that also have an influence on implementation.

Sensors are continuously shrinking with the development of new manufacturing processes. While there are no clear winners, the highest precisions are more often than not achieved with periodic scales having some form of quadrature detection.

One pattern becoming apparent is the relative footprint of actuators in comparison to the processes they drive. This is readily seen with the electromagnetic motors of mesoscale lathes being comparable to that of the lathe itself and the electrostatic comb drives of MEMS devices typically around 1- 2 or more times the size of the micro-mirrors and gears that they drive. Solution of these types of problems might drive new paradigms in system design. One possibility, for example, would be to select across these scales and use a macro-scale motor to drive many mesoscale machines, much like the old 19th century factories.

Perhaps this, and other, broader views may help identify patterns of change in scale. From this, optimal system designs will emerge and change as scales are further reduced, ultimately to the fundamental physical limits imposed at atomic scales.

Acknowledgements

The authors would like to thank Dr's. Lowell Howard and Jon Pratt for their interest and thought-provoking conversation regarding the topics expressed in this paper. Thanks are also due to Dr. Steve Ludwick for his insightful comments. This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under contract No. W-7405-Eng-48.

References

- [1] Matey J.R. and Blanc J., 1985, Scanning capacitance microscopy, *J. Appl. Phys.*, **57**(5), 1437 – 1444.
- [2] Bugg C.J and King P.J., 1988, Scanning capacitance microscopy, *J. Phys. E: Sci. Instrum.*, **21**, 147 – 151.
- [3] Lawall J. and Kessler E., 2000, Michelson interferometry with 10 pm accuracy, *Rev. Sci. Instrum.*, **71**(7), 2669-2676.
- [4] Badami V.K. and Patterson S.K., 2002, A frequency domain method for the measurement of non-linearity in heterodyne interferometry, *Precision Engineering*, **24**(1), 41-49.
- [5] Bobroff N., 1993, Recent advances in displacement measuring interferometry, *Meas. Sci Technol.*, **4**, 907-926.
- [6] Schott, Walter and Jager, Gerd, 2002, Miniature interferometers for precise distance measurements, *Proc. ASPE*, **27**, 67 – 72.
- [7] www.toptica.com.
- [8] Abou-Zeid A. and Wiese P., 1998, Interferometer with a wavelength tuned diode laser for surface profilometry, *Meas. Sci. Technol.*, **9**, 1105 – 1110.
- [9] Dunn T.J., Michaels R., Lee S., Tronolone M. and Kulawiec A, 2002, Fast optical form measurements of rough cylindrical and conical surfaces in diesel fuel injection components, *Proc. ASPE*, **27**, 89 – 93.
- [10] www.fisba.ch.

-
- [11] Young, M., 2000, *Optics and lasers, Including Fibers and Optical Waveguides*, Springer-Verlag, New York.
- [12] See for example www.optipro.com/Stil/index.html, www.zygo.com, www.veeco.com, www.phase-shift.com.
- [13] See for example www.renishaw.caom, www.micr-e.com, www.heidenhian.com,
- [14] Kuroda A., Optical displacement measurement system for detecting the relative movement of a machine part, US Patent # 6,166,817 and references therein.
- [15] Teimel A., 1991, Technology and application of grating interferometers in high-precision measurement, in *Progress in Precision Engineering*, eds. Seyfried P. *et al.*, Springer-Verlag, Berlin, ISBN 0-387-53986-7, 15 – 30.
- [16] Born M. and Wolf E., 1980, *Principles of Optics*, 6th edition, chapter VIII, Cambridge University Press, the equation presented in this text is an approximation to the more complete analysis presented in this book.
- [17] Yacoot A. and Cross N., 2003, Measurement of picometer non-linearity in an optical grating encoder using x-ray interferometry, *Meas. Sci. Technol.*, **14**, 148 – 152.
- [18] Sawabe M., Maeda F., Yamaryo Y., Shimomura T., Saruki Y., Sakai H. and Aoyagi S., 2002, Development of the vacuum interferometric comparator for calibrating the fine linear encoders and scales, presented at the 7th International Symposium on Laser Metrology, *IMEKO*, September 10th.
- [19] Howard, L., Stone, J. and Fu, J., 2001, Real-time displacement measurements with a Fabry-Perot cavity and a diode laser, *Precision Engineering*, **25** (4), pg. 321-335.

-
- [20] Froehlich, F. F. and Milster, T. D., 1994, Minimum detectable displacement in near-field scanning optical microscopy, *Applied Physics Letters*, **65** (18), pg. 2254-2256.
- [21] Paesler M.A. and Moyer P.J., 1996, *Near-field optics: Theory, instrumentation, and applications*, J. Wiley and sons, NY, ISBN 0471043117.
- [22] Bowling-Barnes R. and Silverman S., 1934, Brownian motion as a natural limit to all measuring processes, *Rev. Mod. Phys.*, **6**, 162–192.
- [23] Jones R.V. and Richards J.C.S., 1973, The design and some applications of sensitive capacitance micrometers, *J. Phys. E.*, **6**, 589 – 600.
- [24] Tariq, H., et al., 2002, The linear variable differential transformer (LVDT) position sensor for gravitational wave interferometer low-frequency controls, *Nuclear Instruments and Methods in Physics Research, Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, **489** (1-3), pg. 570-576.
- [25] Bowen D.K., Chetwynd D.G. and Schwartzenberger D.R., 1990, Sub-nanometer displacement calibration using x-ray interferometry, *Meas. Sci. Technol.*, **1**(2), 107 – 119.
- [26] Beckwith T.G., Buck N.L. and Marangoni R.D., 1982, *Mechanical measurements*, 3rd ed., AddisonWesley, Ma, chapter 12.
- [27] Miller, M. M., et. al., 1997, Development of a high precision absolute linear displacement sensor utilizing GMR spin-valves, *IEEE Transactions on Magnetics*, **33** (5), pg. 3388-3390.
- [28] Popovic R.S., Flanagan J.A. and Besse P.A., 1996, The future of magnetic sensors, *Sensors and Actuators A*, **56**, 39 – 55.

-
- [29] Tsang C., 1984, Magnetics of small magnetoresistive sensors, *J. Appl. Phys.*, **55**(6), 2226 – 2231.
- [30] Scrag B.D. and Xiao G, 2003, Submicron electrical current density imaging of embedded microstructures, *Appl. Phys. Letts.*, **82**(19), 3272 - 3274
- [31] Smith C.H., Schneider R.W., Dogaru T., and Smith S.T., 2002, GMR magnetic sensor arrays for NDE eddy current sensing, *Rev. Prog. In Quantitative NDE*, **22**, to be published
- [32] Dodd C.V. and Deeds W.E., 1968, Analytical solutions to eddy-current probe-coil problems, *J. Appl. Phys.*, **39**(6), 2829 – 2838.
- [33] Smith C.H., Schneider R.W., Dogaru T., and Smith S.T., 2002, GMR magnetic sensor arrays for NDE eddy current sensing, *Rev. Prog. In Quantitative NDE*, **22**, 419 - 426
- [34] Dogaru T., Smith C.S., Schneider R.W. and Smith S.T., 2001, New directions in eddy current sensing technology, *Sensors Magazine*, **18**(6), 52-56.
- [35] Yound R.D., 1966, Field emission ultramicrometer, *Rev. Sci. Instrum.*, **37**(3), 275 – 278.
- [36] Young R.D., Ward J. and Scire F., 1972, The topografiner, An instrument for measuring surface microtopography, *Rev. Sci. Instrum.*, **40**(12), 999 – 1011.
- [37] Binnig G., Rohrer H., Gerber Ch. And Weibel E., 1982, Tunneling through a controllable vacuum gap, *Appl. Phys. Letts.*, **40**(12), 178 – 180.
- [38] Binnig G. and Rohrer H., 1982, Scanning tunneling microscopy, *Helv. Phys. Acta*, **55**, 726 – 735.

-
- [39] Chetwynd D.G. and Smith S.T., 1991, High precision surface profilometry: From stylus to STM, in *From Instruments to Nanotechnology*, ed. Gardner J.W. and Hingle H.T., Gordon and Breach, London, 273-300.
- [40] Binnig G., Quate C.F. and Gerber Ch., 1986, Atomic force microscope, *Phys. Rev. Letts.*, **56**(9), 930 – 933.
- [41] See for example Matthais Scherge and Stanislav N. Gorb, 2001, *Biological micro- and nanotribology: Nature's solutions*, Springer-Verlag, Berlin, ISBN 1434-4904.
- Sarid D., 1994, *Scanning force microscopy: With applications to electric, magnetic and atomic forces*, Oxford University Press.
- [42] Mavroidis C., Pfeiffer C. and Mosley M., 1999, Conventional actuators, shape memory alloys, and electrorheological fluids, in *Automation, miniature robotics and sensors for non-destructive testing and evaluation*, ed. Bar-Cohen Y., chapter 5.
- [43] Spadaccini, C., Zhang, X., Cadou, C., Miki, N., and Waitz, N., 2003, Preliminary Development of a Hydrocarbon Fueled Catalytic Micro Combustor, *Sensors and Actuators*, **103**, pg. 219-224.
- [44] Whalen, S., Thompson, M., Bahr, D., Richards, C. and Richards, R., 2003, Design, Fabrication and Testing of the P3 Micro Heat Engine, *Sensors and Actuators*, **104**, pg. 290-298.
- [45] Anderson, B., 2002, Piezoelectric actuators move precisely and powerfully, *Laser Focus World*, **38** (9), pg. 119-123.
- [46] Bittle, S. D. and Kurfess, T. R., 1997, An Active Piezoelectric Probe for Precision Measurement on a CMM, *Mechatronics*, **7** (4), pg. 337-354.

-
- [47] Mitrovic, M., Carman, G. P. and Straub, F. K., 2001, Response of piezoelectric stack actuators under combined electro-mechanical loading, *International journal of solids and structures*, **38** (24-25), pg. 4357-4374.
- [48] Harb, S. Smith, S. T. and Chetwynd, D. G., 1992, Subnanometer behavior of a capacitive feedback, piezoelectric displacement actuator, *Rev. Sci. Instrum.*, **63** (2), pg. 1680-1689.
- [49] Samuelson, M., et al., 1997, Actuator systems for precision motion control, *Advances in the Astronautical Sciences*, **94**, Proceedings of the 1997 Annual AAS Rocky Mountain Guidance and Control Conference, Feb 5-9 1997, Breckenridge, CO, USA, pg. 215-232.
- [50] Binnig, G. and Smith, D.P.E., 1986, Single-tube-3-dimensional scanner for scanning tunneling microscopy, *Rev. Sci. Instrum.*, **57** (8), pg. 1688-1689.
- [51] Chen, C. J., 1992, Electromechanical deflections of piezoelectric tubes with quartered electrodes, *Appl. Phys. Lett.*, **60** (1), pg. 132-134.
- [52] Wang, Z., Jouaneh, M. K. and Dornfeld, D. A., 1990, Design and characterization of a linear motion piezoelectric micropositioner, *IEEE Control Systems Magazine*, **10** (2), pg. 10-15.
- [53] Park, J., et al., 2001, Development of a compact displacement accumulation actuator device for both large force and large displacement, *Sensors and Actuators, A: Physical*, **A 90**, pg. 191-202.
- [54] Kim, S-C. and Kim, S. H., 2001, A precision linear actuator using piezoelectrically driven friction force, *Mechatronics*, **11** (8), pg. 969-985.

-
- [55] Kim, J-D. and Nam, S., 1997, Development of a micro-depth control system for an ultr-precision lathe using a piezoelectric actuator, *Int. J. Mach. Tools Manufact.*, **37** (4), pg. 495-509.
- [56] Gao, P. and Swei, S., 1999, A six-degree-of-freedom micro-manipulator based on piezoelectric translators, *Nanotechnology*, **10** (4), pg. 447-452.
- [57] Gao, P., Swei, S-M. and Yuan, Z., 1999, A new piezodriven precision micropositioning stage utilizing flexure hinges, *Nanotechnology*, **10** (4), pg. 394-398.
- [58] Janocha, H. and Kuhnen, K., 2000, Real-time compensation of hysteresis and creep in piezoelectric actuators, *Sensors and Actuators, A: Physical*, **79** (2), pg. 83-89.
- [59] Xu, Y. and Jones, B., 1997, A simple means of predicting the dynamic response of electromagnetic actuators, *Mechatronics*, **7** (7), pg. 589-598.
- [60] Janocha, H., 2001, Application potential of magnetic field driven new actuators, *Sensors and Actuators, A: Physical*, **91** (1-2), pg. 126-132.
- [61] Awabdy, B. A., Shih, W-C. and Auslander, D. M., 1998, Nanometer Positioning of a Linear Motion Stage Under Static Loads, *IEEE/ASME Transactions on Mechatronics*, **3** (2), pg. 113-119.
- [62] Smith, S. T. and Chetwynd, D. G., 1990, An optimized magnet-coil force actuator and its application to precision elastic mechanisms, *Proceedings of the Institution of Mechanical Engineers, Part C: Mechanical Engineering Science*, **204** (4), pg. 243-253.
- [63] Hadfield D., 1962, *Permanent magnets and magnetism*, John Wiley and Sons, Inc., London.

-
- [64] Kiang, M.-H., Solgaard, O., Lau, K. Y. and Muller, R. S., 1998, Electrostatic Combdrive-Actuated Micromirrors for Laser-Beam Scanning and Positioning, *Journal of Micromechanical Systems*, 7 (1), pg. 27-37.
- [65] Kallenbach, E., et al., 1999, New polarized electromagnetic actuators as integrated mechatronic components – design and application, *Mechatronics*, **9** (7), pg. 769-784.
- [66] Chen, F., Xie, H. and Fedder, G.K., 2001, A MEMS-Based Monolithic Electrostatic Microactuator for Ultra-Low Magnetic Disk Head Fly Height Control, *IEEE Transactions on Magnetics*, **37** (4), pg. 1915-1918.
- [67] Feddeme J.T., Warne L.K., Johnson W.A., Ogden A.J. and Armour D.L., 2002, Assembly of LIGA using electric fields, *Sand report*, sand2002-1084, see also <http://www.ntis.gov/ordering.htm>
- [68] Bobbio S., Goodwin-Johansson J., DuBois T., Tranjan F., Smith S., Fair R., Ball C., Jacobson J., Bartlett C., Eleyan N., Makki H. and Gupta R., 1996, Integrated force array: Positioning drive applications, *Proc SPIE*, **2722**, 123 – 135.
- the higher work densities quoted in this paper are based on a private communication and represent more recent work carried out in 2002.
- [69] Kim K.J. and Shahinpoor M., 2002, Applications of polyelectrolytes in ionic polymeric sensors and artificial muscles, in *Handbook of Polyelectrolytes and their Applications*, ed. Tripathy S.K., Kumar J. and Nalwa H.S., vol. 3: *Applications of Polyelectrolytes and Theoretical Models*, American Scientific Publishers, ISBN 1-58883-016-0.

-
- [70] Taylor, G. W., et. al., 1985, *Piezoelectricity, Ferroelectricity and Related Phenomena Vol. 4*, Gordon and Breach Science Publishers, New York, NY, ISBN: 0677166605.
- [71] Zhuang, Z. and Shi, H., 1993, Temperature-stable high electrostrictive strain relaxor ferroelectric ceramics for servo-actuator applications, *Sensors and Actuators, A: Physical*, **35** (3), pg. 279-282.
- [72] Uchino, K., 1995, Advances in ceramic materials, *Materials Letters*, **22** (1-2), pg. 1-4.
- [73] Cross, E. L., 1996, Ferroelectric materials for electromechanical transducer applications, *Materials Chemistry and Physics*, **43** (2), pg. 108-115.
- [74] Lu, T. J. and Evans, A. G., 2002, Design of a high authority flexural actuator using an electro-strictive polymer, *Sensors and Actuators, A: Physical*, **99** (3), pg. 290-296.
- [75] See for example Etrema Products Inc., 2500 North Loop Drive, Ames, Iowa, IO 50010
- [76] Portman V., Sandler B.Z., 1999, High-stiffness precision actuator for small displacements, *Int. J. Machine Tools and Manuf.*, **39**, 823 – 837.
- [77] Gravesen, P., 1993, Microfluidics – a review, *Journal of Micromechanics and Microengineering*, **3** (4), pgs. 168-182.
- [78] Andersson, H. and van der Berg, A., 2003, Microfluidic devices for cellomics: a review, *Sensors and Actuators, B: Chemical*, **92**(3), 315-325.
- [79] Harmon, M. E., Tang, M., and Frank, C. F., 2003, A microfluidic actuator based on thermosponsive hydrogels, *Polymer*, **44**(16), 4547-4556.

-
- [80] Chong, J. M. and MacDonald, N.C., 1998, Suspended moving channels and channel actuators for microfluidic applications, *American Society of Mechanical Engineers, Dynamic Systems and Control Division*, **66**, Micro-Electro-Mechanical Systems (MEMS), 433-438.
- [81] See for example Zemansky M.W., 1957, *Heat and Thermodynamics*, McGraw-Hill, NY, chapter 9.
- [82] Henning, A.K., 1998, Microfluidic MEMS, In Proceedings, IEEE Aerospace Conference, Paper 4.906, IEEE Press, Piscataway, NJ, 1998,
<http://www.redwoodmicro.com/Papers/Aerospace.pdf>.
- [83] Butefisch, S., and Buttgenbach, S., A New Pneumatically Actuated Miniature Gripper for Micro Assembly,. <http://www.tu-bs.de/institute/imt/mitarbeiter/buetefisch/pdf/bostpneum.pdf>.
- [84] Grosjean C. and Tai Y.A., Thermopneumatic Peristaltic Micropump, Caltech Micromachining Lab, Caltech EE 136-93, Pasadena, CA 91125.
<http://touch.caltech.edu/publications/charlesg/trans99/trans99.pdf>.
- [85] Meng E., Wang X., Mak H., and Tai Y., “A Check-Valved Silicone Diaphragm Pump” Caltech Micromachining Lab, 136-93, Caltech, Pasadena, CA 921126.
<http://touch.caltech.edu/publications/ellis/mems00/mems00.pdf>.
- [86] Bridgeman P.W., 1949, *The physics of high pressure*, G. Bell and Sons, Ltd, London, chapter 6, compressibility of solids.

-
- [87] Mihalcz, I., 2001, Fundamental Characteristics and Design Method for Nickel-Titanium Shape Memory Alloy, *Periodica Polytechnica, Mechanical Engineering*, **45** (1), pgs. 75-86.
- [88] Webb, G., Wilson, L., Lagoudas, D. and Rediniotis, O., 2000, Adaptive Control of Shape Memory Alloy Actuators for Underwater Biomimetic Applications, *AIAA Journal*, **38** (2), pgs. 325-334.
- [89] Otsuka, K. and Kakeshita, T., 2002, Science and Technology of Shape-Memory Alloys; New Developments, *MRS Bulletin*, **27** (2), pg. 91-98.
- [90] Bergamasco, M., Dario, P. and Salsedo, F., 1990, Shape Memory Alloy Microactuators, *Sensors and Actuators, A: Physical*, **21** (1-3), pgs. 253-257.
- [91] Gordaninejad, F. and Wu, W., 2001, A two-dimensional shape memory alloy/elastomer actuator, *International Journal of Solids and Structures*, **38** (19), pgs. 3393-3409.
- [92] Krulevitch, P., et al., 1996, Thin Film Shape Memory Alloy Microactuators, *Journal of Microelectromechanical Systems*, **5** (4), pg. 270-282.
- [93] Rivin E.I., Sayal G., Johal P.R.S. and Singh R.P.G., 2004, Structural applications of SMA/superelastic materials, *Proc. 2004 NSF Design, Service and Manufacturing Grantees and Research Conference*, Dallas, TX.
- [94] Samuelson, M., et al., 1997, Actuator systems for precision motion control, *Advances in the Astronautical Sciences*, **94**, Proceedings of the 1997 Annual AAS Rocky Mountain Guidance and Control Conference, Feb 5-9 1997, Breckenridge, CO, USA, pg. 215-232.

-
- [95] Liang, C and Rogers, C.A., 1992, Design of Shape Memory Alloy Actuators, *Journal of Mechanical Design*, **114** (2), pg. 223-230.
- [96] Otsuka, K. and Wayman, C.M., 1998, *Shape Memory Materials*, Cambridge University Press, Cambridge, UK, 284 Pgs., ISBN: 052144487.
- [97] Srinivasan, A.V. and McFarland, D.M., 2001, *Smart Structures: Analysis and Design*, Cambridge University Press, Cambridge, UK, 228 Pgs., ISBN: 0521650267.