Review of scaling effects on physical properties and practicalities of cantilever sensors

Yang, C. K.; Van Der Drift, E. W.J.M.; French, P. J.

DOI
10.1088/1361-6439/ac8559

Publication date
2022

Document Version
Final published version

Published in
Journal of Micromechanics and Microengineering

Review of scaling effects on physical properties and practicalities of cantilever sensors

Yang, C. K.; Van Der Drift, E. W.J.M.; French, P. J.

DOI
10.1088/1361-6439/ac8559

Publication date
2022

Document Version
Final published version

Published in
Journal of Micromechanics and Microengineering

TOPICAL REVIEW

Review of scaling effects on physical properties and practicalities of cantilever sensors

To cite this article: C-K Yang *et al* 2022 *J. Micromech. Microeng.* **32** 103002

IOP Publishing Journal of Micromechanics and Microengineering
J. Micromech. Microeng. **32** (2022) 103002 (15pp) https://doi.org/10.1088/1361-6439/ac8559

Topical Review

Review of scaling effects on physical properties and practicalities of cantilever sensors

C-K Yang[1,*], E W J M van der Drift[2] and P J French[1,*]

[1] TU Delft Faculty EEMCS, Mekelweg 4, 2628CD Delft, Netherlands
[2] TU Delft Faculty of Applied Sciences, Lorentzweg 1, 2628CJ Delft, Netherlands

E-mail: c.k.yang@tudelft.nl and p.j.french@tudelft.nl

Received 25 February 2022, revised 30 June 2022
Accepted for publication 29 July 2022
Published 18 August 2022

Abstract

Reducing sensor dimension is a good way to increase system sensitivity and response. However the advantages gained must be weighed against other effects which also became significant during the scaling process. In this paper, the scaling effect of cantilever sensors from micrometre to nanometre regimes is reviewed. Changes in the physical properties such as Q-factor, Young's modulus, noise and nonlinear deflections, as well as effects on practical sensor applications such as sensor response and sensor readouts, are presented. Since cantilever is an elemental transducer and device building block, its scaling effects can be further extrapolated to other sensing systems and applications.

Keywords: scaling, effects, physical, properties, practicalities, cantilever

(Some figures may appear in colour only in the online journal)

List of abbreviations

S	Scaling factor
k	Flexural stiffness
υ	Deflection
σ	Stress
ω	Frequency
ρ	Density
E	Young's modulus
I	Moment of inertia
N	Section moduli
F	Force acting upon the cantilever
C	Resonant mode coefficient
m	Cantilever mass
L	Cantilever length
w	Cantilever width
t	Cantilever thickness
Q	Quality factor
α_T	Linear thermal expansion
T	Temperature
C	Heat capacitance
D_T	Thermal diffusion coefficient
l_T	Thermal diffusion length
τ_R	Thermal relaxation time constant
τ_σ	Thermal stress relaxation time constants
τ_ε	Thermal strain relaxation time constants
K_B	Boltzmann constant
A_n	Spectra amplitude noise per unit length
E_{th}	Thermal noise energy
E_c	Noiseless carrier energy
Δf	Measurement bandwidth
κ	Exact cantilever curve expression

* Authors to whom any correspondence should be addressed.

1. Introduction

The act to change the size of an object at a fixed geometric ratio, is to scale. Size matters because the physics at large scales are different to that of small scales. Particularly in the scaling from microelectromechanical systems (MEMS) to nanoelectromechanical systems (NEMS), where mass becomes less significant and surface effects become increasingly dominant. As result, NEMS devices tends to behave differently from MEMS and bulk devices.

Early work on scaling related discussions have shown its broadness in multiple study fields. Work such as Trimmer's invention of scaling factors [1], Feynman's speech on miniaturisation [2], Thornell's examples of everyday-life scaling [3], Spearing's observations on MEMS scaling [4], Roukes' introduction to NEMS [5], Bell's summary in sensor selections [6] and Agrawal's study in multiscale experiments [7], to name a few, all agreed on its importance to future technologies. In this paper, we will base our discussion on the scaling of silicon cantilever sensors. This is because cantilever is one of the basic electromechanical device building block, and silicon is one of the most developed processing material in MEMS and NEMS sensors. It is the goal of this paper to present the scaling effect of cantilever sensors and compare its benefits and pitfalls.

Forces typically undergo what is called the 'scaling law' as they act upon a scaled device. The concept of scaling law can be best represented by the Trimmer's notation [1], where an indexed parameter S^F is used to represent the scaling factor of different forces F relative to the change of the device size S. In this paper, Trimmer's notation is extended not only to forces, but also to parameters of a typical cantilever sensor such as resonant frequency, deflection amplitude, stiffness etc.

In this paper, we look at two key aspect changes when scaling a cantilever: the physical aspect and the practical aspects. The physical aspect, concerns the material and intrinsic property changes of the device. It includes sensitivity, energy dissipation, material constants, noise limits and response linearity. The practical aspect on the other hand concerns the actual usage and measurement of the sensor. Discussions include reaction surfaces and readout.

2. Physical aspect of cantilever scaling

2.1. Basic cantilever mechanics and adapted Trimmer's notation

For a cantilever, it is known from basic mechanics [8] that:

$$\text{Flexural stiffness } k = \frac{3EI}{L^3} \tag{1}$$

$$\text{Deflection at tip } \upsilon(L) = \frac{FL^3}{3EI} \tag{2}$$

$$\text{Stress at clamp } \sigma_x(0) = \frac{FL}{N} \tag{3}$$

$$\text{1st resonance } \omega_0 = C_0\sqrt{\frac{3EI}{mL^3}} \tag{4}$$

$$\text{Resonance shift by stiffness } \Delta\omega_{(\Delta m=0)} = \frac{1}{2}\frac{\Delta k}{k}\omega_0 \tag{5}$$

$$\text{Resonance shift by mass } \Delta\omega_{(\Delta k=0)} = -\frac{1}{2}\frac{\Delta m}{m}\omega_0 \tag{6}$$

where E is the Young's modulus of the cantilever, I is the moment of inertia, N is the section moduli, F is the force acting upon the cantilever, C_0 is the fundamental resonant mode coefficient, m is the cantilever mass and L is the length of the cantilever. For a cantilever with rectangular cross-section, $I = wt^3/12$ and $N = wt^2/6$ where w and t are cantilever's width and thickness respectively. The scaling of these parameters can then be derived by expressing them in terms of cantilever length, width and thickness dimensions (ρ is the density):

$$m = \rho wtL \propto SSS = S^3 \tag{7}$$

$$k = \frac{3E}{12}\frac{wt^3}{L^3} \propto \frac{SS^3}{S^3} = S \tag{8}$$

$$\upsilon(L) = \frac{4}{E}\frac{FL^3}{wt^3} \propto \frac{S^F S^3}{SS^3} = S^F S^{-1} \tag{9}$$

$$\sigma_x(0) = 6\frac{FL}{wt^2} \propto \frac{S^F S}{SS^2} = S^F S^{-2} \tag{10}$$

$$\omega_0 = C_0\sqrt{\frac{3E}{(\rho wtL)L^3}}\sqrt{\frac{wt^3}{12}} \propto \left(\frac{SS^3}{SSSS^3}\right)^{1/2} = S^{-1} \tag{11}$$

$$\Delta\omega_{(\Delta m=0)} = \frac{\Delta k}{2}\frac{1}{k}\omega_0 \propto \frac{1}{S}\frac{1}{S} = S^{-2} \tag{12}$$

$$\Delta\omega_{(\Delta k=0)} = -\frac{\Delta m}{2}\frac{1}{m}\omega_0 \propto -\frac{1}{S^3}\frac{1}{S} = -S^{-4}. \tag{13}$$

Hence we can conclude that as a cantilever scales down, its flexural stiffness decrease linearly, its deflection at the tip increases linear, its stress at clamping point increases by power of 2 and its resonant frequency also increases linearly.

Here we also want to comment on scaling of resonance shift by stiffness $\Delta\omega_{(\Delta m=0)}$ and by mass $\Delta\omega_{(\Delta k=0)}$. Suppose a cantilever is designed to detect stiffness changes, assuming mass change of the system is negligible, then scaling the cantilever dimensions results in power of 2 increase in resonance change. On the other hand, if by mass, the change is to the power of 4. This is advantageous for sensor applications because typically the readout frequency analyser of a sensor has finite resolution that decreases at higher resonance frequency. As the resonant frequency ω_0 only scales with factor of S^{-1}, the overall $\Delta\omega/\omega_0$ ratio actually increases, making it possible for the readout electronics to resolve the frequency shifts.

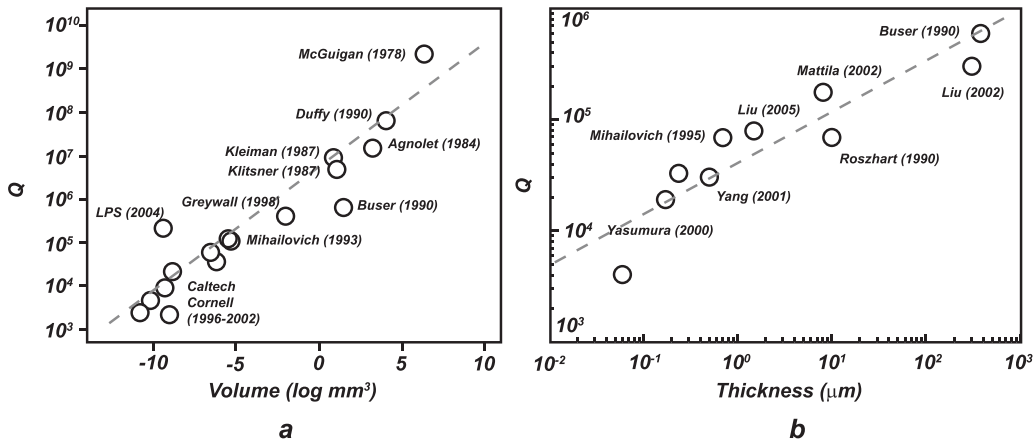

Figure 1. Q-factor size dependence of silicon resonators in vacuum with untreated surfaces. The Q-factors decrease as structure dimension scale down, the dash lines are for eye-guiding. Graph adapted from [13, 14]. Reproduced with permission from [15].

2.2. *Quality factors in scaling*

Quality factor (Q-factor) is an important parameter in cantilever sensor. It is defined as the ratio between the energy stored in the system and the energy dissipated by the system per cycle. A high Q-factor resonator typically exhibits a clean frequency signal with very low spectrum spread and low power dissipation, and translates directly to low noise, low energy consumption and high sensitivity [9, 10]. The Q-factor of silicon resonators has been reported to be as high as 10^9 in low temperature vacuum [11]. Yet as shown in figure 1, the values decreases when the size decreases from millimetre down to nanometre scales [9, 12]. This size dependency of the Q-factors degrades device performance and sets a limitation to future NEMS applications.

The decrease of Q-factors in MEMS and NEMS has been treated in many studies as equivalent to the increase of dissipation, expressed as the reciprocal Q^{-1}. The dominant sources of dissipation can be external or internal, which include: air squeeze-film damping [16, 17], surface related loss [18–20], clamping loss [16, 19], impurity and bulk defects [21, 22] and thermoelastic damping [23, 24]. The total dissipation is the sum of each loss mechanisms [9, 13, 24]:

$$\frac{1}{Q_{\text{tot}}} = \frac{1}{Q_{\text{air}}} + \frac{1}{Q_{\text{surface}}} + \frac{1}{Q_{\text{clamp}}} + \frac{1}{Q_{\text{defect}}} + \frac{1}{Q_{\text{thermoelast}}}. \tag{14}$$

Among these losses, the air squeeze-film damping and surface loss are scalable to the surface area S^2, but can be neglected in vacuum measurements and in the discussion of fundamental limits of the Q factor. The clamping and defects are heavily dependent on means of processing and design. The thermoelastic damping however, is intrinsic for a given material and geometry, and sets the fundamental upper limit to the Q-factor [25].

Thermoelastic damping is the result of an energy dissipating interaction between the strain field and the temperature field. The stressed and strained regions of an elastic material in resonance cause local thermal gradients and drive phonon currents inside the material at the resonance frequency. When the phonon flows are damped, the energy is irreversibly dissipated [23]. Thermoelastic dissipation of micromechanical devices is first predicted by Zener, approximated to the first order by Roszhart for silicon microresonators [23], and later investigated comprehensively by Sun *et al* [26]. Lifshitz and Roukes [25] on the other hand, investigated the damping of both MEM and NEM devices and their size-dependency. Following Lifshitz's work, we shall extend his calculations more fully for the scaling of thin beams with high length-to-thickness aspect ratio cantilevers.

According to Lifshitz, Q^{-1} of rectangular cross-sectioned beams can be expressed as:

$$Q^{-1} = \frac{E\alpha_{\text{T}}^2 T}{C}\left(\frac{6}{\xi^2} - \frac{6}{\xi^3}\frac{\sinh\xi + \sin\xi}{\cosh\xi + \cos\xi}\right)$$

where

$$\xi^2 = t^2\frac{\omega_0}{2D_{\text{T}}} = \frac{C_n^2}{4\sqrt{3}}\frac{t^3}{L^2 l_{\text{T}}}. \tag{15}$$

α_{T} is the linear thermal expansion, T is the temperature, C is the heat capacitance, D_{T} is the thermal diffusion coefficient, C_n is the resonance mode coefficient, t is the beam thickness[3] and l_{T} is the thermal diffusion length. Based on the equations above and the constants given in Lifshitz's work ($C_n = 1.875$, $l_{\text{T}}(300\,\text{K}) = 1.257 \times 10^{-2}$ μm and $E\alpha_{\text{T}}^2(300\,\text{K})/C = 7.942 \times 10^{-5}$), the thermoelastic dampings of different beam geometries at 300 K are reproduced in figures 2(a) and (b), both figures corresponds to the calculations of Lifshitz. For a better insight to the scaling effect of the thermoelastic damping, the length and thickness of the beam should be scaled at fixed aspect ratios. Lifshitz calculated Q^{-1} for a beam ratio of 10:1, but for typical MEM and NEM beams following Euler–Bernoulli bending (i.e. negligible shear deformation), aspect ratio higher than 20:1 is needed. We calculated the scaling of Q^{-1} for different aspect ratio shown in figure 2(c), and in figure 2(d) showed the corresponding size-dependency of the Q-factor [25].

[3] The beam thickness t is actually expressed as b and called as 'beam width' in Lifshitz's work.

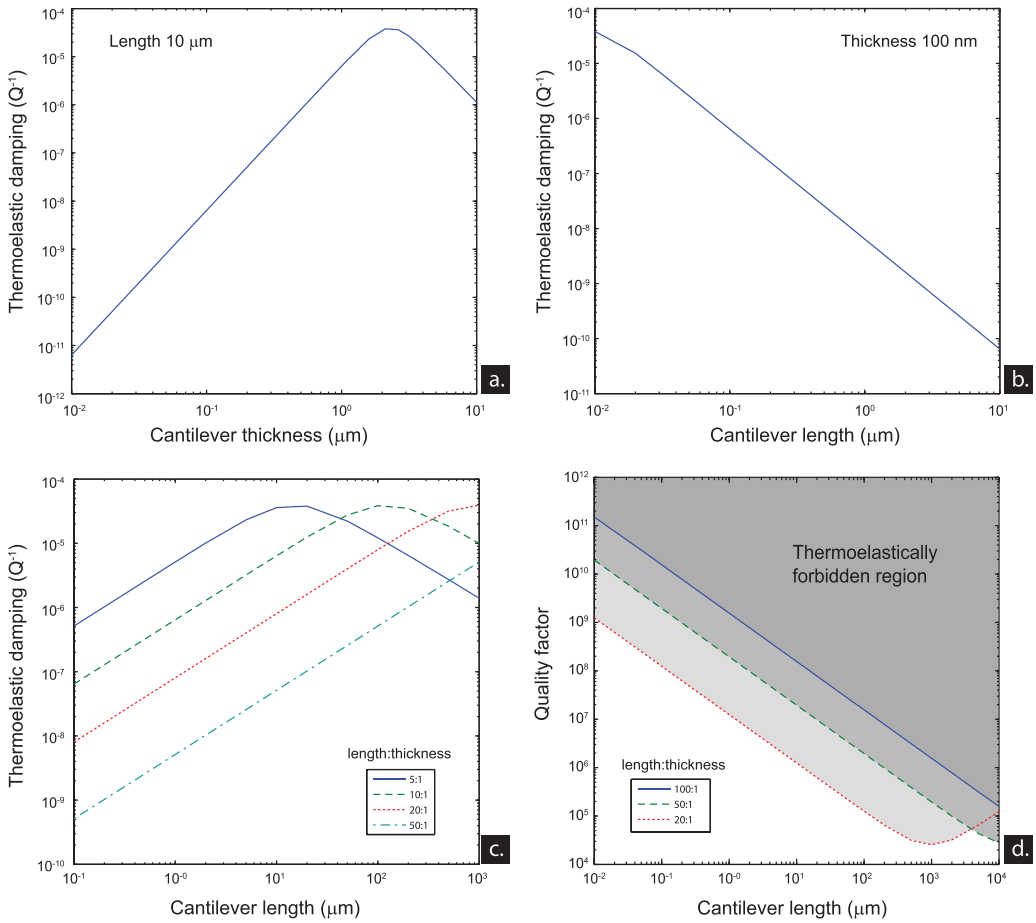

Figure 2. Calculated thermoelastic damping of single clamped silicon beams at 300 K. (a) Damping of fixed beam length 10 μm and varying thickness. (b) Damping of fixed thickness 100 nm and varying length. (c) Damping of several different length-thickness ratios at varying length. (d) Q-factor of the same beams from (c). Reproduced with permission from [15].

According to equation (15), the dissipation is frequency dependent and thus scales with dimension. However instead of a monotonous relation, the dissipation have a maxima at a preferred dimension. This is due to the relaxation time constant τ_{R}, which is related to the dissipation by:

$$\tau_{\mathrm{R}} = \sqrt{\tau_\sigma \tau_\epsilon} = \frac{t^2}{\pi^2 D_{\mathrm{T}}} \tag{16}$$

$$l_{\mathrm{T}} = \sqrt{\frac{\rho}{E}} \frac{t^2}{\pi^2} \frac{1}{\tau_{\mathrm{R}}} \tag{17}$$

where τ_σ and τ_ε are stress and strain relaxation time constants respectively. At low frequency resonance where $\omega_0 \ll \tau_{\mathrm{R}}^{-1}$, the vibrations are isothermal and energy dissipation is little. At high frequency where $\omega_0 \gg \tau_{\mathrm{R}}^{-1}$, the cantilever behave like an adiabatic system, and the dissipation is also little. Only at resonance frequency where $\omega_0 \simeq \tau_{\mathrm{R}}^{-1}$, the stress and the strain become out of phase. This induces a maximum internal friction [23, 27], which results to a peak known as the Debye peak in the dissipation-frequency and dissipation-geometry plots shown in figure 2. Also shown in the figure, is the forbidden region. This is a fundamental limit of elastic energy to which a cantilever can have.

Finally as shown in figures 2(a) and (b), the location of the Debye peak is important to determine the scaling of the dissipation. For a cantilever with a fixed length of 10 μm, dissipation scales-down with thicknesses less than 2 μm. However, for a cantilever with a fixed thickness of 100 nm, then the dissipation increases for shorter cantilevers. This implies the length-thickness ratio of cantilever is relevant and it is advantageous to make cantilever long and slender. In figure 2(c) it is found that the Debye peak shifts to larger scale as the length-thickness ratio increases. For typical cantilever design in MEMS and NEMS applications, the aspect ratio is usually more than 50:1, which means its Debye peak falls in millimetre scales or larger. Therefore in principle for MEMS and NEMS cantilever, the dissipation decreases as the size decreases, and the fundamental Q-factor scales up linearly with decreasing cantilever size. We denote this fundamental Q scaling as S_{FQ}:

$$S_{\mathrm{FQ}} = S^{-1} \equiv S_{\mathrm{FQ}}^{-1}. \tag{18}$$

If the thermoelastic damping decreases with cantilever size, then other dissipation effects from equation (14) must be the causes to the lowering of the Q-factors. Studies have shown that clamping and bulk defect losses can be significant in cantilevers resonance [16, 19, 21, 22]. However they are

highly process dependent, and since both are bulk effects, their significance decreases as surface-to-volume ratio increases. Instead, for nanocantilevers and ultra-thin cantilevers with high surface-to-volume ratio, the surface loss becomes dominant in the decreasing of Q-factor [19, 28]. Surface loss is a collection of several effects which includes adsorptions, surface oxidation, surface defects and surface stress.

Experimental data shows the decreasing trend of Q-factor with down-scaling, and has shown significant changes in Q-factor with surface treatments [28, 29]. It is important to note that, the surface induced losses do not inherently increase when down scaling, The scaling of the dominant surface effect needs to be evaluated individually before one can conclude the effective scaling of the Q_{tot}.

2.3. Size dependent Young's modulus

Flexural rigidity EI is one of the most fundamental mechanical properties for structures like cantilevers. The moment of inertial I is a geometric parameter which is known to scale S^2 for rectangular cross-sectioned cantilevers, while the Young's modulus E is a material constant. However since the rise of NEMS, it has been observed that this flexural rigidity does not scale according to I, but is also influenced by changes in E [30]. Figure 3(a) illustrates some reported E size dependencies of crystalline silicon nanobeams. The experimental results show that the modulus stays approximately constant only until around cantilever thickness of 200 nm, where it then starts to decrease abruptly.

Strictly speaking, E is defined as a material elastic constant and should stay intrinsically unaltered, however the experimentally measured E are usually non-ideal and influenced by other effects. When the measured Young's modulus deviates from the ideal value, as is in our case influenced by the scaling effects and the bending measurement, then the resultant Young's modulus is referred to as the *effective* Young's modulus. Hereafter, E will represent the ideal Young's modulus which is the elasticity of perfect crystal lattices, and $\tilde{E}$ will represent the size dependent effective Young's modulus.

The $\tilde{E}$ size dependency is an ongoing research. Despite efforts in finding the cause using atomistic simulations [31, 38], modified continuum models [39–41] and experimental verifications [33, 34, 42, 43], an explanation of the $\tilde{E}$ size dependence is still not well-understood. While some have proposed that the nonlinear effect of the bulk elastic moduli is the main cause for the size dependence [44], most of the work still showed that surface effects are the dominant causes [40–42].

The origin of surface effects comes from the fundamental difference in the coordination numbers of surface atoms and bulk atoms. The surface atoms, having lower coordination number and electron densities, tends to adopt equilibrium lattice spacings differently from the bulk ones. However in order to retain the continuous epitaxy relationship from bulk to surface, bulk atoms strain the surface atoms and create the so called surface stress σ_s [45]. Due to the σ_s, and the fact that semiconductor surface atoms likes to reconstruct and can displace easily from its original places, the elasticity at the surface

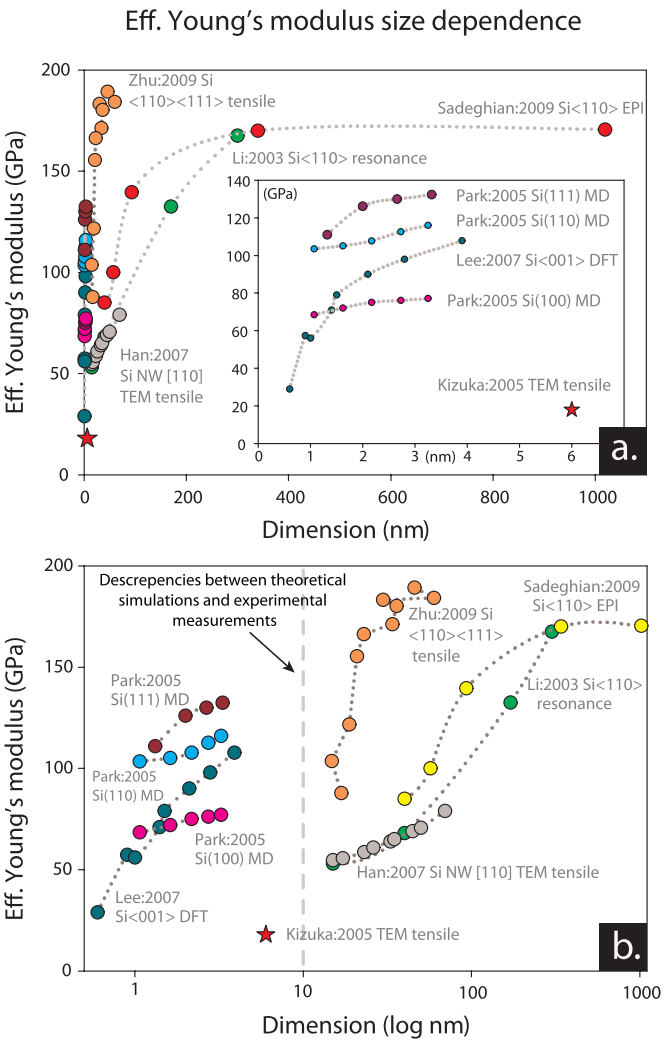

Figure 3. (a) Illustration of the results on the size-dependence of single crystal silicon $\tilde{E}$ from simulation [31, 32] and experiments [33–37]. The inset shows a zoomed-in view of the results under 10 nm. (b) Presents the logarithmic plot of (a), which clearly illustrates the limits and differences between simulation and the experimental results. Reproduced from [33]. © IOP Publishing Ltd. All rights reserved.

E_s (N m^{-1}) is different from the elasticity in the bulk. We can think of the $\tilde{E}$ as the combined effect from both the surface and the bulk elasticity.

On macro scale, most of the atoms in a device dwell within crystal lattices and bulk E dominates the elasticity of the device, but when a device shrinks, the surface-to-volume ratio increases and the influence of E_s on the resulting $\tilde{E}$ increases.

Referring to figure 3(b), we see a significant difference between the model predictions and experimental measurements. The simulation studies are typically limited to structures dimension less than 10 nm due to the computation complexity, while on the other hand, the experimental results are mostly acquired on devices thicker than 10 nm due to the difficulties in measurements. One of the biggest challenge in this field of study then, is the development of multiscale modelling and multiscale experiments [7], where both modelling and measurements need to cross the 10 nm barrier

and cover different scales for a better comparison. Besides the multiscale challenge, current measurement methods also needs to improve for better accuracy, resolution and introduce minimum disturbances to the measured device.

From figure 3, we also observed that σ_s only effects $\tilde{E}$ at characteristic size of less than 10 nm in simulations, while experiments already show 'pre-matured' drops at larger dimensions. The discrepancies between the model and measurements, as well as between measurements and measurements, signal a few possibilities [30]: (a) some extrinsic factors such as clamping, loading modes, nonlinear bulk effect, surface saturation, contamination etc are dominating the experimental results; (b) crystal defects and process induced damages effects becomes significant at sub-micron size; (c) experimental uncertainty, calibration and method (bending, resonance, tension) limitations are influencing the submicron measurements. Therefore, the way how $\tilde{E}$ is experimentally extracted becomes crucial to the research.

In conclusion, we know $\tilde{E}$ scale down with size. However a precise scaling factor could not be determined due to lack of consistent model and experimental results. We therefore denote the scaling effect as:

$$S_{\tilde{E}} = S^{+} \equiv S^{+}_{\tilde{E}}. \tag{19}$$

Following this result, the scaling of the stiffness (equation (8)), deflection (equation (9)), resonance frequency (equation (11)), stiffness resonance shift (equation (12)) and mass resonance shift (equation (13)), should then consider the $\tilde{E}$ size dependence, and be rewritten as:

$$k = \frac{3E}{12}\frac{wt^3}{L^3} \qquad \propto S^{+}_{\tilde{E}} S \tag{20}$$

$$\upsilon(L) = \frac{4}{E}\frac{FL^3}{wt^3} \qquad \propto \frac{S^F}{SS^{+}_{\tilde{E}}} \tag{21}$$

$$\omega_0 = C_0\sqrt{\frac{3E}{(\rho wtL)L^3}}\sqrt{\frac{wt^3}{12}} \qquad \propto \frac{\left(S^{+}_{\tilde{E}}\right)^{1/2}}{S} \tag{22}$$

$$\Delta\omega_{(\Delta m=0)} = \frac{\Delta k}{2}\frac{1}{k}\omega_0 \qquad \propto \frac{1}{S^2\left(S^{+}_{\tilde{E}}\right)^{1/2}} \tag{23}$$

$$\Delta\omega_{(\Delta k=0)} = -\frac{\Delta m}{2}\frac{1}{m}\omega_0 \qquad \propto \frac{\left(S^{+}_{\tilde{E}}\right)^{1/2}}{S^4}. \tag{24}$$

2.4. Noise

In the field of sensor measurements and signal processing, noise reduction is a key subject to achieve successful application. As the devices scale down in dimension, the influences of different noise sources change too. For MEMS and NEMS, noise can be created intrinsically from: thermal-mechanical loss, anelastic defects, or the Nyquist–Johnson noise. Extrinsically, noise can be introduced from: electromagnetic interferences or cross-talk, mechanical-acoustic vibrations, micro temperature fluctuations, and adsorption–desorption molecular interactions [46].

In this work, we will focus on the noise induced by thermal-mechanical-loss, which is often referred to as the thermal noise. This is mainly because among all types of noise, thermal noise is inevitable and sets the fundamental lower limit of the noise level (noise floor). This fundamental noise level is often used in cantilever applications to determine the minimal detectable displacements and frequency shifts [10, 47–50].

Thermal noise is the result of the fluctuation–dissipation process [51, 52]. In thermal equilibrium state, energies are constantly being exchanged between the excitation and its surroundings. The environment drives energies into the system causes microscopically random fluctuations. This driving source however, cannot pump energies into the oscillation system infinitely, it requires instead a dissipation mechanism such as a damper to dissipate and return the extra energies back to the surroundings. Therefore the existence of both the fluctuating source and the dissipating damper cannot be separated, and together they complete the energy path between the oscillator and its surroundings. This is the essence of the fluctuation–dissipation theorem [53, 54], which establishes the quantification of the thermal noise.

From equipartition theorem and simple harmonic oscillation, the spectral noise force per unit length can be derived as [46, 53]:

$$F_n = \frac{2K_BTM\omega_0}{\pi QL^2} \qquad \propto \left(S^{+}_{\tilde{E}}\right)^{1/2} S_{FQ} \tag{25}$$

where K_B is the Boltzmann constant, T is the temperature, ω_0 is the resonant frequency and Q is the quality factor. From previous section we knew that the effective Young's modulus and thermal elastic Q-factor scale $S^{+}_{\tilde{E}}$ and $S^{-1}_{F.Q.}$ respectively. Despite not knowing their true form, we can still derive that the spectra force noise should scale with dimension.

For a cantilever, Cleland *et al* derived the spectra amplitude noise per unit length as [46]:

$$A_n = \frac{1}{(1-(\omega/\omega_0)^2)^2 + 1/Q^2}\frac{1}{\omega_0^3}\frac{2K_BT}{\pi mQL^2} \tag{26}$$

$$\propto \frac{S^{-1}_{FQ}}{\left(S^{+}_{\tilde{E}}\right)^{3/2}}S^{-2} = \frac{1}{S_{FQ}\left(S^{+}_{\tilde{E}}\right)^{3/2}S^2}. \tag{27}$$

At resonance frequency $\omega = \omega_0$, thermal noise driven amplitude scales inversely with the cantilever dimension, albeit not knowing the exact relation. Butt and Jaschke, obtained the same result by summing the different cantilever vibration modes under thermal noise [55]. Here we understood that as a cantilever scales down, the noise force acting on it decreases, but the resultant amplitude noise likely increases due to its decrease in stiffness.

Thermal noise not only cause random deflection response of a cantilever in the static mode, but also causes widening (sidebands) of the resonant peak in the dynamic mode [46]. Therefore measurement of cantilever resonance shift can also be vulnerable to thermal noise. It has been shown that $\delta\omega_n$,

defined as the minimum frequency shift that can be resolved from the sideband (noise spread), is expressed as [10, 50]:

$$\delta\omega_{\mathrm{n}} = \left(\frac{E_{\mathrm{th}}}{E_{\mathrm{c}}}\right)^{1/2}\left(\frac{\omega_0\Delta f}{Q}\right)^{1/2} \tag{28}$$

$$\propto \left(\frac{(S_{\tilde{E}}^{+})^{1/2}S^{-1}}{S_{\mathrm{FQ}}^{-1}}\right)^{1/2} = \left(\frac{\left(S_{\tilde{E}}^{+}\right)^{1/2}S_{\mathrm{FQ}}}{S}\right)^{1/2} \tag{29}$$

where E_{th} and E_{c} are the thermal noise energy and a noiseless carrier energy respectively, Δf is the measurement bandwidth. Given that the energy ratio $E_{\mathrm{th}}/E_{\mathrm{C}}$ and the measurement bandwidth are size-independent environment variables, $\delta\omega_{\mathrm{n}}$ has a scaling factor that is very much dependent on the effective Young's modulus and Q-factor.

In real applications such as mass sensing, this noise frequency shift can be used to derive a measurable 'mass-noise' of:

$$\delta m_{\mathrm{n}} = -2m_{\mathrm{eff}}\frac{\delta\omega_{\mathrm{n}}}{\omega_0} = -2M_{\mathrm{eff}}\left(\frac{E_{\mathrm{th}}}{E_{\mathrm{c}}}\right)^{1/2}\left(\frac{\Delta f}{Q\omega_0}\right)^{1/2} \tag{30}$$

$$\propto S^3\left(\frac{1}{S_{\mathrm{FQ}}^{-1}}\frac{S}{\left(S_{\tilde{E}}^{+}\right)^{1/2}}\right)^{1/2} = S^{7/2}\left(\frac{S_{\mathrm{FQ}}}{\left(S_{\tilde{E}}^{+}\right)^{1/2}}\right)^{1/2} \tag{31}$$

which suggests that as device becomes smaller, the derived mass noise caused by cantilever thermal fluctuation will also diminish. In other words, the minimum detectable mass becomes smaller. This is why it is advantageous to scale down the sensor size for resonant mass sensing.

2.5. Elastic nonlinearity

Cantilever sensors can operate in either linear or nonlinear elastic regimes. In the case of linear elastic regime, cantilevers are actuated by small deflections and are assumed to have a linear force–displacement response. As a cantilever scales, the limits of its linear force–displacement response also changes, and therefore is scale dependent.

The geometric nonlinearity of a cantilever origins from its high order bending derivative. Consider small deflection of a long, slender cantilever, then its bending curvature can be approximated $\kappa = \mathrm{d}\theta/\mathrm{d}s \approx \mathrm{d}\theta/\mathrm{d}x \approx \mathrm{d}^2\nu/\mathrm{d}x^2$ as shown in figure 4 (where $\mathrm{d}s \approx \mathrm{d}x$ and $\theta \approx \tan\theta \approx \mathrm{d}\nu/\mathrm{d}x$). In large deformation however, the $\theta \approx \tan\theta$ approximation is no longer valid and the exact curve expression must be used [56]:

$$\kappa = \frac{\frac{\mathrm{d}^2\nu}{\mathrm{d}x^2}}{[1+(\frac{\mathrm{d}\nu}{\mathrm{d}x})^2]^{3/2}}. \tag{32}$$

Figure 5 compares the vertical and horizontal deflection at the cantilever tip of both large and small deflection cases. The y-axis is expressed as a distance ratio in order to compare the amount of displacement over the total cantilever length, while the x-axis is expressed in a dimension-less loading-parameter.

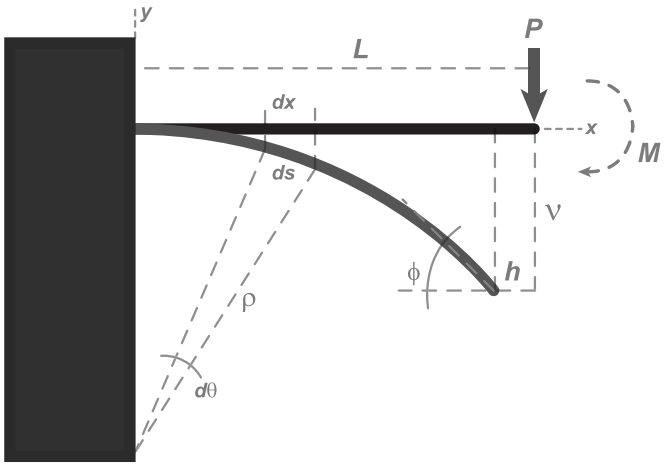

Figure 4. Schematic of a cantilever under a point force P with a large deflection. Assuming the beam is not extensible, then the approximation $\mathrm{d}s \approx \mathrm{d}x$ is still valid. However due to large bending angle, the $\theta \approx \tan\theta$ approximation becomes incorrect. Consequently the approximation will also give wrong estimations of displacement ν and h. Reproduced with permission from [15].

In principle the linear approximation overestimates the vertical deflection ν and neglects totally the horizontal deflection h. The discrepancy between the large and small deflections on ν can be explained physically: initially the system is linear and its displacement is proportional to the applied force. However as the beam bends and locally stretches, an axial tensile force is induced and changes the stress inside the beam, which then results to an overall stiffening effect of the system [57]. The negligence of this stiffening is the main reason for the over-prediction of the displacements by the linear approximation in figure 5(a).

From the solution, it is also observed that when the tip deflection reaches about 30% of the total cantilever length, the linear approximation starts to become inaccurate (10% error). The transition from small to large deflection is a smooth curve and the validity of the linear approximation will depend on the tolerance of the user and the application.

As the structure scales down, its flexural rigidity EI decreases and is easily bent by external forces. Therefore, for a specific force, a smaller cantilever will experience more pronounced large-deflection and geometrical nonlinearity. Hence the maximum allowable force F_{max} for a linear cantilever response decreases with its size by a scaling factor of:

$$F_{\mathrm{max}} = \upsilon_{\mathrm{max}}(L)\frac{3EI}{L^3} \qquad \propto S_{\tilde{E}}^{+}S. \tag{33}$$

3. Practical aspect of cantilever scaling

3.1. Scaling sensitivity up and scaling area down

When a cantilever sensor scales down in size, it becomes more sensitive to external force disturbances and responds with higher deflection, stress and resonant frequency shift. However, as the sensors decrease in size, their surface area, where most of the sensing reactions such as direct force

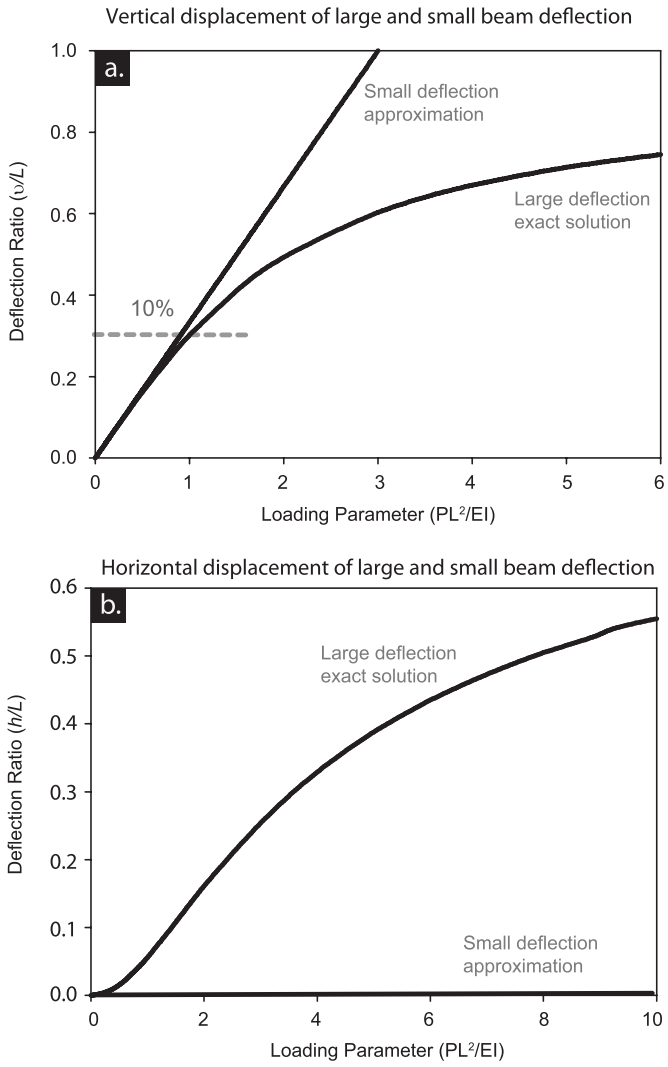

Figure 5. (a) Comparison of the vertical deflection. The two solutions agrees with each others when the loading is low and the displacement ratio is small. However at around 30% ν/L, the error increases to 10%. (b) Comparison of the horizontal displacement. While the large deflection solution calculate considerable displacement, the small deflection solution neglects it completely. Data adapted from [8]. Reproduced with permission from [15].

pressures, chemical bindings and absorption are happening, will also decrease. Therefore the sensitivity of the scaled device is increased but the reaction strength is decreased. The combination of these two opposite effects results to a competition that influences the feasibility of a sensor and its performance in applications. The competition between sensitivity and surface area is most evident in the field of biotechnology sensing, where sensor surfaces are typically coated with receptors to 'trap' targeted molecule in solutions [58]. The probability of a successful trap depends on several factors such as solution concentration, diffusivity, binding or dissociation constants etc. Yet to our interest in scaling, is the total number of receptors available on the surface area of the sensor.

Experimental results have shown inconsistent effects in low concentration bio-detections where high sensor sensitivity is most needed. Sheehan and Whitman showed in their experiments and theoretical work, that scaling sensors to nano will result to an overall degrading sensor performance [59]. In their case, the decrease in the binding events greatly dominates the increase of the sensor sensitivity, and ultimately results to an impractical sensor response time of few days to weeks. On the other hand, similar nanowire sensors have been experimentally tested by various researchers to have not only high sensitivity, but also fast response time in low concentration solutions [60–62]. The discrepancies and the controversial measurement results can be explained by the work of Squires *et al* [63] in which the authors identified crucial parameters that can affect the time response of surface-reaction biosensors in nanochannels. It concludes that the availability of sensor surface for reaction is significant, but other parameters such as target concentration, diffusivity, channel size and aspect ratio etc will also play decisive roles in the overall sensor responsiveness, regardless of the sensor sensitivity.

Returning to cantilever sensors and its fundamentals, it is not difficult to see that, for a surface dependent sensing, scaling the sensors size can reduce its performance. For a typical surface reaction sensor, consider the definition of sensitivity (coefficient) and two parameters defined here as the *time per stimulus* and *time per response*:

$$\text{sensitivity} = \frac{\text{output}}{\text{input}} = \frac{\text{response}}{\text{stimulus}}$$

$$\text{time per stimulus} = \frac{\text{time}}{\text{stimulus}}$$

$$\text{time per response} = \frac{\text{time}}{\text{response}}.$$

Stimulus is the measurement input of the sensor, it can be a target DNA binding to a receptor, a molecule absorbed or some deposition mass applied onto the cantilever surface. *Response* is defined here as the detectable reaction of the sensor to the stimuli; it is typically a deflection, a stress accumulation or a resonance frequency shift. Therefore, *time per stimulus* means the time needed for a unit binding or adsorption to happen, and *time per response* means the time needed for a sensor to accumulate enough input for a measurable response[4].

It has been shown that the sensitivity of deflection sensing scales with S^{-1} (equation (9)), and the sensitivity of frequency sensing scales with S^{-2} (equation (12)) for stress-changes and S^{-4} for mass-changes (equation (13)). We can then write the cantilever sensitivity scaling in Trimmer's matrix notation as:

$$\begin{bmatrix} \text{Sensitivity static deflection} \\ \text{Sensitivity resonant stress} \\ \text{Sensitivity resonant mass} \end{bmatrix} = \begin{bmatrix} S^{-1} \\ S^{-2} \\ S^{-4} \end{bmatrix}. \qquad (34)$$

[4] This is not to be confused with the sensor response time, which is typically the time needed for a sensor to respond to a stimulus change.

By definition, *time per stimulus* should be inverse proportional to the probability of an stimulus event, which is related to the reaction surface area. Therefore:

$$\text{time per stimulus} \qquad \propto S^{-2}. \tag{35}$$

Then it is possible to derive *time per response* scaling from its definition and combining equations (34) and (35):

$$\text{time per response} = \frac{\frac{\text{time}}{\text{stimulus}}}{\frac{\text{response}}{\text{stimulus}}} = \frac{\text{time per stimulus}}{\text{sensitivity}} \propto S^{-2}\begin{bmatrix} S^{-1} \\ S^{-2} \\ S^{-4} \end{bmatrix}^{-1} = \begin{bmatrix} S^{-1} \\ S^{0} \\ S^{2} \end{bmatrix}. \tag{36}$$

From the above we see that the time it takes to acquire measurable deflection response increases as the cantilever decreases. While for resonant stress sensing response stays constant. For resonant mass sensing the response decreases quadratic with size.

Finally, equation (36) is about accumulation of signals over a period of time, but similar scenario can be applied to sensing forces on cantilever surface, such as force from electrostatic, magnetic or gravitation forces etc. In this case, the important aspect is on the amount of total force strength acting on the device at any given moment. Scaling of such forces is related to the surface (S^2) or the volume (S^3) of the device, and the response is simply:

$$\begin{aligned} &\text{response} = \text{stimulus force} \times \text{sensitivity} \\ &\propto S^2\begin{bmatrix} S^{-1} \\ S^{-2} \\ S^{-4} \end{bmatrix} = \begin{bmatrix} S^{1} \\ S^{0} \\ S^{-2} \end{bmatrix} \quad \text{for area } S^{\mathrm{F}} = S^2, \text{ and} \\ &\propto S^3\begin{bmatrix} S^{-1} \\ S^{-2} \\ S^{-4} \end{bmatrix} = \begin{bmatrix} S^{2} \\ S^{1} \\ S^{-1} \end{bmatrix} \quad \text{for volume } S^{\mathrm{F}} = S^3 \end{aligned} \tag{37}$$

for deflection, the response decreases with size because it receives less total input. The increase of sensitivity is counter-weighted by the decrease of force acting on the device. However for resonant mass sensing, the increase of the sensitivity still over-weights the loss of the stimulus as the cantilever scales down.

In conclusion, scaling of cantilever sensors for higher sensitivity is not always beneficial to the overall sensor performance. In order to keep the surface reaction sensing under a reasonable sampling time, the loss of surface area can be compensated by implementing arrays of nanosensors so that each one has higher sensitivity while the total surface of the array stays comparable to that of a microsensor. Other options are by optimising the measurement parameters and sensor design to maximise sample collections, such as directing targets into focused flows, changing the concentrations or diffusivity of the fluids and enhancing surface reactions etc. For force sensing however, the array approach will have no effect since what counts is the area or volume per cantilever, and not the total area of the array. The solution instead lies fundamentally at asking which force scales better. As an example, electrostatic force scales with area, but it is also inversely proportional to the gap distance, therefore by considering the gap into the scaling design may result to a better sensor performance. Nevertheless, in principle, distributed forces that scales equal or more than S^2 is not suitable for scaled, high sensitivity cantilever sensing.

3.2. Scaling of readout

Cantilever is a mechanical transducer, it transforms external disturbances into bending and resonance shift. In order to completely interpret the cantilever responses, a mechanism is needed to 'read-out' and turn them into electrical signals. When the cantilever scales, the applicability of the readout transducers are also affected.

A few common cantilever detection methods are listed in table 1. The list provides a general overview of the popular methods and shows the wide possibilities available. It is however not exhaustive, as different and new varieties of techniques also exist. For example in optical techniques, existing variations include lever deflection, interferometry [85], scattering [86] and other detections that monitor laser light's intensity, phase, wavelength, position, frequency or polarisation changes [87]. Yet not all these optical methods are easily implemented. Hitherto, the lever deflection technique is the most successful and widely implemented optical technique for cantilever sensing.

In the following we will look at the scaling of these read methods. One particular parameter of interest is the backaction. Backaction is the disturbance on cantilever's original response due to interactions between the cantilever and the detection system; in other words, it is the unwanted effects imposed on the cantilever by the detection system during the detection.

3.2.1. Optical lever. The technique is based on light deflection from the cantilever [64]. The deflection principle itself is not cantilever size dependent. However the decrease of cantilever surface area leads to decrease of total laser light reflected on the photo-detectors, which results to decrease in signal to noise ratio (SNR). The low SNR signals means the need of low-noise amplifiers which typically have limited bandwidth and response time. The back-action of laser light on the cantilever can result to joule heating and light pressure effects. Although the increasing surface-to-volume ratio of a scaled cantilever can reduce the effect of the joule heating by its higher dissipation rate, the increasing mechanical sensitivity of the cantilever means even small temperature and force differences from heating and light pressure, can induce significant changes [88]. Finally, the integration of optical lever technique is fundamentally challenging. Integrated optical technologies such as waveguides and photonic crystals do not have the means to focus light onto cantilevers for detections; instead, these technologies sense cantilever displacements by mode-coupling of lights [89, 90].

Table 1. Different MEMS and NEMS cantilever detection techniques. Some of the techniques are developed and mainly used in double clamped beams, but are also applicable to cantilevers. Reproduced with permission from [15].

Techniques	Description	Reference
Optical lever	Focus laser beam applied onto an optical reflective resonator, the laser light is then deflected onto a set of photodiodes. Cantilever motion is sensed by detecting the displacement of the deflected laser light. This method is one of the most popular method in microcantilevers, and widely used in AFM systems.	[64, 65]
Capacitive	Capacitive cantilever sensing involves a moving cantilever and a fixed counter-electrode. Cantilever motion is sensed by monitoring capacitance change due to displacement in the gap. This method is very popular in MEMS devices in the form of interdigitated structures. When used in cantilever sensing, good controls in thermal compensation and parasitic capacitance are needed.	[66]
Single-electron-transistors (SETs)	Single electron transistor is a quantum effect device involving tunnelling and Coulomb blockade on an isolated Coulomb island. The sensing technique is based on using SET as a high sensitive electrometer/amplifier, for monitoring the capacitance or voltage changes induced by a displacement transducer. This sense-and-amplify configuration increases SNR by direct amplification of the charge signals near the source, and with high sensitivity, high response. Typically the SET is linked to a cantilever with a capacitive readout, yet other readout transducers such as piezoelectric also exists.	[67–70]
Piezoresisitive	Piezoresisitive sensing involves electric resistance change due deformation. The resistance change is a combinational effect of both material property (gauge factor) and geometric change, and therefore exists in all kinds of material. Cantilever motion is senses by monitoring the amount of deformation near the base/anchoring point. This is usually done by depositing an extra layer on one side of the cantilever.	[71–74]
Piezoelectric	Piezoelectric sensing involves special material which produces electric charges when subjected to force deformation. Cantilever motion is sensed by monitoring the total charge accumulated along the deformed cantilever. This method requires the piezoelectric material to be sandwiched between electrodes, and this usually results to fabrication complications and multi-layered cantilevers. A particular advantage of integrating piezoelectric element on cantilever is that it can provide both sensing and actuation.	[75, 76]
Magnetomotive	Magnetomotif involves running an AC electric current path though the cantilever in a magnetic field. The electromagnetic field creates a Lorentz force that actuate the cantilever into motion. Yet in return, the displacement of the conducting current in the magnetic field induces an electric potential on the cantilever that can be measured and referred back to the motion. Similar to piezoelectric technique, magnetomotif can be used as a combined cantilever sensing and actuation.	[77]
Hard contact	The hard contact technique is based on the on-off switch mechanism, where a biased cantilever 'turns-on' the electrical path when it bends and touches the counter electrode, and 'turns-off' when it returns to the original position. The technique can generate large electrical signals comparing to most of other readouts, and therefore potentially simplifies the electronics design. Due to the on-off nature of the technique, it cannot be used for quasi-static bending measurements. However, it can produce 'digitized' output that enables dynamic monitoring of the cantilever by counting switching cycles. Since the technique requires contact, stiction prevention, current induced micro-welding, contact wear, shock and overall mechanical reliability is significant in its application.	[78]
Direct tunnelling	Direct tunnelling sensing is a solid-solid quantum tunnelling that involves a counter electrode at the proximity of the moving cantilever. Under slight electric potential difference, electrons will jump from one electrode to the other at small distance. The probability, or the rate, of jumping is dependent on the distance. Cantilever motion is therefore sensed by monitoring the tunnelling rate/current.	[79, 80]
Field emission	Field emission sensing is a solid-surface quantum tunnelling that involves a counter electrode, typically pointed shape, under high electric tension. The sharp electrode is used as a cathode that shoots electrons and the cantilever collects the electrons. The emission current is a function of the distance, therefore the cantilever motion is sensed by monitoring the current changes.	[81–84]

3.2.2. Capacitive. The capacitive sensing is achieved through monitoring of the capacitance changes on the cantilever and its counter-electrode. Under a constant initial separation gap, its SNR decreases with the size of cantilever due to smaller capacitance resulted from a S^2 reduction of area. Similarly, the sensitivity also suffers as the capacitance decreases. The frequency response on the other hand, increases due to smaller capacitance, which results to less RC time constant and faster response. Nevertheless, low SNR leads to more stringent electronic requirements as in the case of the optical lever, and the overall system bandwidth will be limited by the electronic readout. The back-action of the technique is the bias voltage applied to create charge accumulations for the sensing, this voltage can exert electrostatic force on the cantilever and induce stiffening effects [91]. Finally, the capacitive technique is highly integrable.

3.2.3. Single electron transistor (SET). The SET displacement sensing is a combination of a displacement transducer and an SET electrometer amplifier based on Coulomb blockade. Since it is a combined sense-amplify technique, its scaling effect is highly dependent on the choice of the transducer. Despite the use of SET in conjunction with piezoelectric transducers [67], SET displacement sensing nowadays typically refers to cantilever monitoring with capacitive readout [68–70, 92]. When the cantilever is scaled down, the capacitance charge signal of the displacement decreases similarly to the capacitive technique. Yet despite the decreasing nature, the remarkable charge sensitivity of SET devices offsets the disadvantages of the normal capacitance sensing, and enables a high SNR in nano-displacement sensing [69]. The frequency response of SET is very high [93], since there is no apparent limitation to the frequency response of electron tunnelling and Coulomb blockade. Back-action of the technique is mainly due to the coupling between tunnelling events and the nanocantilever [94], as well as due to charge fluctuation at the centre island of the SET device [95].

3.2.4. Piezoresistive. Piezoresistive sensing is based on the changes in geometrical and material resistive properties due to external force. In principle, the geometric factor dominates the scaling effect, results to dimension-related decreases in SNR and sensitivity. However, material factor such as the giant piezoresistance effect [96], can play an important role and increase both the SNR and the sensitivity greatly. The frequency response is not fundamentally limited by the piezoresistive effect, however since the method requires running electric current through the device, the response is affected by the increasing resistance due to decreasing cross-section area. This increase of the overall device resistance can result in increased parasitic RC time constant, and challenging impedance matching. Using thin metal film as the piezoresistive layer can solve the problem [97]. Back-action of the technique come from the heating effect by the electric current, but can be minimised by decreasing the device resistance (which also increases the response at the cost of sensitivity). In most cases, piezoresistive sensing is integrable with silicon devices.

3.2.5. Piezoelectric. Piezoelectric cantilever sensing is based on a layer of piezoelectric material sandwiched between two electrodes. The stress-induced voltage differences is proportional to the thickness of the piezoelectric layer [98], therefore both SNR and sensitivity of the technique decreases with down-scaling. Frequency response increases due to the decreasing scale-dependent parasitic capacitance, however the upper bandwidth is still limited due to the low SNR similar to capacitive technique. In quasi-static measurements, charge accumulated across the device leaks through the piezoelectric element or readout electronics, therefore the technique has some limitations on low-frequency applications. In terms of back-action, since piezoelectric element serves both sensing and actuation purpose, the act of sensing can easily cause actuation of the cantilever, such as voltage noise at the readout circuits acting on the piezoelectric element and cause additional deflection [67]. Finally, integration of piezoelectric element on cantilever has been achieved both in micron and nano scales [67, 99], but the requirement for piezoelectric layer and sandwich contacts typically complicates the device fabrication comparing to other techniques.

3.2.6. Magnetomotive. Magnetomotive sensing is based on measuring electromotive force induced by the conductive beam in the presence of magnetic field. Based on the theory of electromotive force, the induced electric force (voltage) is proportional to length and the displacement of the beam [92]. Therefore both SNR and sensitivity scales with the beam dimension. Having said that, due to the low output impedance nature of the device, its electric noise can be minimised by impedance matching at the readout circuits, and therefore achieving overall high system SNR. The sensitivity does not scale well, but the high SNR results to high resolution and easy signal amplification, enabling nano-scale measurements of this technique. The frequency response is related to the impedance of the beam, therefore size dependent and decreases with the dimensions. Yet in applications where the beams are made highly conductive, frequency response of the technique can achieve GHz range [100]. In the low frequency range however, due to the motion dependent nature of the technique, quasi-static deflections of beams are difficult to measure. Similar to piezoelectric technique, the magnetomotif technique also serves sensing-actuation purposes. This advantage becomes disadvantage when the back-action and cross-coupling of the signals occurs. Nevertheless new techniques such as balanced detection [101], have been developed to minimise the effects, making magnetomotif a suitable NEMS sensing technique. The integration of the device is difficult due to the need of strong magnetic field [92], and the smaller the device, the stronger the magnetic field is required.

3.2.7. Tunnelling. The electron tunnelling effect is strictly dependent on the gap between electrodes and not on their dimensions, therefore the current SNR and sensitivity of the technique is scale independent. However, because a current path is established on the cantilever, its resistance at nanometre dimensions can reduce the current signal, increase the

Table 2. Scaling effects of cantilever discussed in this chapter.

Parameter	Scaling factor	Notes on general interpretation, assumptions and limitations	Reference
Flexural stiffness k	$S_{\tilde{E}}^{+}S$	Linear deflection, long and thin cantilevers (Euler–Bernoulli theorem)	Equation (20)
Cantilever deflection $\upsilon(L)$	$\frac{S^F}{SS_{\tilde{E}}^{+}}$	Linear deflection, long and thin cantilevers (Euler–Bernoulli theorem)	Equation (21)
Cantilever stress $\sigma_x(0)$	$S^F S^{-2}$	Small deflection	Equation (10)
Resonant frequency ω_0	$\frac{(S_{\tilde{E}}^{+})^{1/2}}{S}$	Small deflection	Equation (22)
Resonance shift by stiffness $\Delta\omega_{(\Delta m=0)}$	$\frac{1}{S^2(S_{\tilde{E}}^{+})^{1/2}}$	Assuming decoupled from mass	Equation (23)
Resonance shift by mass $\Delta\omega_{(\Delta k=0)}$	$\frac{(S_{\tilde{E}}^{+})^{1/2}}{S^4}$	Assuming decoupled from stiffness	Equation (24)
Fundamental Q-factor Q_{F}	S_{FQ}^{-1}	Thermoelastic limited. Less dissipation for smaller cantilevers with high length-thickness aspect ratio at 300 K. Also depends on the critical dimension where Debye peak is located.	Equation (18)
Effective Young's modulus $\tilde{E}$	$S_{\tilde{E}}^{+}$	Silicon. Decrease in smaller devices. Can also be generalised to other modulus. Experimentally concluded.	Equation (19)
Minimum detectable force F_{n}	$(S_{\tilde{E}}^{+})^{1/2}S_{\mathrm{FQ}}$	Spectra noise force per unit length. Less thermal force acting at high frequency, which means smaller device that operate at high resonance will experience less noise.	Equation (25)
Thermal noise driven deflection A_{n}	$\frac{S_{\mathrm{F.Q.}}^{-1}}{(S_{\tilde{E}}^{+})^{3/2}}S^{-2}$	The cantilever deflection response of total thermal noise scales inversely but weakly.	Equation (27)
Minimum resonance shift	$\left(\frac{(S_{\tilde{E}}^{+})^{1/2}S_{\mathrm{FQ}}}{S}\right)^{1/2}$	Thermal noise driven resonance shift decreases for smaller devices.	Equation (29)
Thermal noise driven equivalent mass shift δm_{n}	$S^{7/2}\left(\frac{S_{\mathrm{FQ}}}{(S_{\tilde{E}}^{+})^{1/2}}\right)^{1/2}$	Thermal noise induced mass error decreases for smaller devices.	Equation (31)
Large deflection nonlinearity $F_{\max}$	$S_{\tilde{E}}^{+}S$	The maximum force allow before the large deflection effect becomes significant, scales with dimension.	Equation (33)
Time per response	S^{-1}, S^0, S^2	S^{-1} for deflection response, S^0 for resonant stress response and S^2 for resonant mass response	Equation (36)
Response to area-related sensing	S, S^0, S^{-2}	S for deflection, S^0 for resonant stress response and S^{-2} for resonant mass response	Equation (37)
Response to volume-related sensing	S^2, S, S^{-1}	S^2 for deflection response, S for resonant stress response and S^{-1} for resonant mass response	Equation (37)

thermal noise, and ultimately decrease the SNR. The frequency response of the technique is not affected by the scale. The back-action is mainly caused by the voltage biasing of the cantilever and the counter electrode, which induces electrostatic force on the cantilever. Integration of the technique with the cantilever on single substrate has been routinely achieved.

3.2.8. Hard contact. When the hard contact is configured in current mode, the running current SNR scales with both the cantilever resistance as well as the contact resistance. Yet since the cantilevers are either made of, or coated with low-resistance metals, the scaling effect on SNR is not obvious [78]. The back-action of the technique is dominated by the contact of the cantilever tip on the counter electrode, affecting the free-vibration. The potential difference between the cantilever and the electrode can also induce electrostatic force which further impede free-vibration. Finally, the technique can be integrated.

3.2.9. Field emission. Field emission technique bares much resemblance to tunnelling in the scaling effects of SNR, sensitivity and frequency response. The back-action of the

Table 3. Comparisons between Si cantilever detection techniques for NEM cantilever. The indications are represented by positive (+), negative (−) and cross (×) to respectively mean advantageous, disadvantageous and not-applicable (or independent) to sensing scaled cantilevers. Reproduced with permission from [15].

	SNR	Sensitivity	Frequency response	Backaction	Integration	Static or dynamic	Conditions
Optical lever	− −	×	×	−	− −	Both	Assume high bandwidth readout circuits
Capacitive	− −	−	+	−	++	Both	Assume high bandwidth readout circuits
SET	++	−	+	−	++	Both	
Piezoresistive	++	++	×	−	++	Both	With thin-film metals or giant piezoresistive effect
Piezoelectric	−	−	+	− −	−	Frequency	Quasi-static limited due to leakage
Magnetomotive	++	−	+	−	− −	Frequency	
Tunnelling	×	×	×	−	++	Both	Highly conductive cantilever
Hard contact	−	×	+	− −	+	Frequency	Highly conductive cantilever
Field emission	+	×	+	− −	+	Both	vacuum

field emission however is typically stronger, as the voltage bias required in field emission is much higher. The integration is also more complicated than the tunnelling technique, as sharp-tip geometry is needed in creating the emission, and a shielding gate structure is needed to shield the imposed strong electrostatic force.

4. Conclusion

The scaling of micro cantilever sensor to nano dimension was discussed in both physical and practical aspects. Table 2 presents a summary of the scaling effects which can be used as guideline of how scaling can affect a device (cantilever) performances. Some of the effects either cannot be expressed in simple Trimmer's notation, or do not have analytical solutions. This is perhaps the biggest challenges ahead as the physics models, in particular the surface interactions resulting to changes of Young's modulus and Q-factor, are not well known. The determination of scaling is therefore often under assumptions and dependent on the applications.

Apart from the scaling of the cantilever itself, the readout of the cantilever also scales. Table 3 summarises the scaling of the discussed readout methods and their sensor parameters. The scalabilities are represented by positive (+), negative (−) and cross (×) to mean advantageous, disadvantageous and not-applicable (or independent) to sensing applications. It is worth to note that, although some technique displays more advantages than others in scaling, the scalability of a technique also depends on the applications. The table does not conclude with recommendation to any particular technique, rather it simply shows the scaling advantages and disadvantages of each aspect of different techniques.

Data availability statement

No new data were created or analysed in this study.

Acknowledgments

The author would like to thank Dr Hamed Sadeghian, Dr Khashayaer Babaei Gavan, Prof. Herre van der Zand and Prof. Fred van Keulen for the collaborations, support and discussions during the research.

ORCID iD

P J French https://orcid.org/0000-0003-3260-4645

References

[1] Trimmer W 1989 Microrobots and micromechanical systems *Sens. Actuators* **19** 267–87
[2] Feynman R 1992 There's plenty of room at the bottom [data storage] *J. Microelectromech. Syst.* **1** 60–66
[3] Thornell G 2000 A nurse less likely *Digest of Workshop of MicroMechanical Europe (Uppsala, Sweden)*
[4] Spearing S M 2000 Materials issues in microelectromechanical systems (MEMS) *Acta. Mater.* **48** 179–96
[5] Roukes M 2001 Plenty of room, indeed *Sci. Am.* **285** 48–57
[6] Bell D, Lu T, Fleck N and Spearing S 2005 MEMS actuators and sensors: observations on their performance and selection for purpose *J. Micromech. Microeng.* **15** S153–64
[7] Agrawal R and Espinosa H 2009 Multiscale experiments: state of the art and remaining challenges *J. Eng. Mater. Technol.* **131** 041208
[8] Gere J and Timoshenko S 1991 *Mechanics of Materials* 3rd edn (London: Chapman & Hall)
[9] Yasumura K, Stowe T, Chow E, Pfafman T, Kenny T, Stip B and Rugar D 2000 Quality factors in micron- and submicron-thick cantilevers *J. Microelectromech. Syst.* **9** 117–25
[10] Albrecht T, Grutter P, Horne D and Rugar D 1991 Frequency modulation detection using high-*Q* cantilevers for enhanced force microscope sensitivity *J. Appl. Phys.* **69** 668–73
[11] McGuigan D, Lam C, Gram R, Hoffman A, Douglass D and Gutche H 1978 Measurements of the mechanical *Q* of single-crystal silicon at low temperatures *J. Low Temp. Phys.* **30** 621–9
[12] Carr D, Evoy S, Sekaric L, Craighead H and Parpia J 1999 Measurement of mechanical resonance and losses in nanometer scale silicon wires *Appl. Phys. Lett.* **75** 920–2
[13] Liu X, Vignola J, Simpson H, Lemon B, Houston B and Photiadis D 2005 A loss mechanism study of a very high *Q* silicon micromechanical oscillator *J. Appl. Phys.* **97** 023524

[14] Ekinci K and Roukes M 2005 Nanoelectromechanical systems *Rev. Sci. Instrum.* **76** 061101
[15] Yang C K 2012 From MEMS to NEMS: scaling cantilever sensors *Doctoral Thesis* TU Delft
[16] Hosaka H, Itao K and Kuroda S 1995 Damping characteristics of beam-shaped micro-oscillators *Sens. Actuators* A **49** 87–95
[17] Lee J, Lee S, Yao C and Fang W 2007 Comments on the size effect on the microcantilever quality factor in free air space *J. Micromech. Microeng.* **17** 139–46
[18] Blom F, Bouwstra S, Elwenspoek M and Fluitman J 1992 Dependence of the quality factor of micromachined silicon beam resonators on pressure and geometry *J. Vac. Sci. Technol.* B **10** 19–26
[19] Yang J, Ono T and Esashi M 2002 Energy dissipation in submicrometer thick single-crystal silicon cantilevers *J. Microelectromech. Syst.* **11** 775–83
[20] Wang D, Ono T and Esashi M 2003 Crystallographic influence on nanomechanics of ultra-thin silicon resonators *12th Int. Conf. on Solid State Sensors, Actuators and Microsystems, Proc. Transducers* pp 336–9
[21] Mihailovich R and MacDonald N 1995 Dissipation measurements of vacuum-operated single-crystal silicon microresonators *Sens. Actuators* A **50** 199–207
[22] Liu X, Haucke H, Vignola J, Simpson H, Baldwin J, Houston B and Photiadis D 2009 Understanding the internal friction of a silicon micro-mechanical oscillator *Mater. Sci. Eng.* A **521–522** 389–92
[23] Roszhart T 1990 The effect of thermoelastic internal friction of the *Q* of micromachined silicon resonators *Solid-State Sensor and Actuator Workshop, 1990. 4th Technical Digest* (IEEE) pp 13–16
[24] Vignola J *et al* 2002 Loss mechanisms in MEMS-oscillators *Proc. SPIE* **4827** 466–77
[25] Lifshitz R and Roukes M 2000 Thermoelastic damping in micro- and nanomechanical systems *Phys. Rev.* B **61** 5600–9
[26] Sun Y, Fang D and Soh A 2006 Thermoelastic damping in micro-beam resonators *Int. J. Solids Struct.* **43** 3213–29
[27] Gysin U, Rast S, Ruff P, Meyer E, Lee D W, Vettiger P and Gerber C 2004 Temperature dependence of the force sensitivity of silicon cantilevers *Phys. Rev.* B **69** 045403
[28] Ono T and Esashi M 2005 Effect of ion attachment on mechanical dissipation of a resonator *Appl. Phys. Lett.* **87** 044105
[29] Yang J, Ono T and Esashi M 2001 Investigating surface stress: surface loss in ultrathin single-crystal silicon cantilevers *J. Vac. Sci. Technol.* B **19** 551–6
[30] Park H, Cai W, Espinosa H and Huang H 2009 Mechanics of crystalline nanowires *MRS Bull.* **34** 178–83
[31] Lee B and Rudd R 2007 First-principles study of the Young's modulus of Si <001> nanowires *Phys. Rev.* B **75** 041305
[32] Park S, Kim J, Park J, Lee J, Choi Y and Kwon O 2005 Molecular dynamics study on size-dependent elastic properties of silicon nanocantilevers *Thin Solid Films* **492** 285–9
[33] Sadeghian H, Yang C, Goosen J, van der Drift E, Bossche A, French P and van Keulen F 2009 Characterizing size-dependent effective elastic modulus of silicon nanocantilevers using electrostatic pull-in instability *Appl. Phys. Lett.* **94** 221903
[34] Kizuka T, Takatani Y, Asaka K and Yoshizaki R 2005 Measurements of the atomistic mechanics of single crystalline silicon wires of nanometer width *Phys. Rev.* B **72** 035333
[35] Li X, Ono T, Wang Y and Esashi M 2003 Ultrathin single-crystalline-silicon cantilever resonators: fabrication technology and significant specimen size effect on Young's modulus *Appl. Phys. Lett.* **83** 3081–3
[36] Han X, Zheng K, Zhang Y, Zhang X, Zhang Z and Wang Z 2007 Low-temperature *in situ* large-strain plasticity of silicon nanowires *Adv. Mater.* **19** 2112–8
[37] Zhu Y, Xu F, Qin Q, Fung W and Lu W 2009 Mechanical properties of vapor–liquid–solid synthesized silicon nanowires *Nano Lett.* **9** 3934–9
[38] McDowell M, Leach A and Gall K 2008 On the elastic modulus of metallic nanowires *Nano Lett.* **8** 3613–8
[39] Cuenot S, Frétigny C, Demoustier-Champagne S and Nysten B 2004 Surface tension effect on the mechanical properties of nanomaterials measured by atomic force microscopy *Phys. Rev.* B **69** 165410
[40] Miller R and Shenoy V 2000 Size-dependent elastic properties of nanosized structural elements *Nanotechnology* **11** 139–47
[41] He J and Lilley C 2008 Surface effect on the elastic behavior of static bending nanowires *Nano Lett.* **8** 1798–802
[42] Chen C, Shi Y, Zhang Y, Zhu J and Yan Y 2006 Size dependence of Young's modulus in ZnO nanowires *Phys. Rev. Lett.* **96** 075505
[43] Gavan K B, Westra H, van der Drift E, Venstra W and van der Zant H 2009 Size-dependent effective Young's modulus of silicon nitride cantilevers *Appl. Phys. Lett.* **94** 233108
[44] Liang H, Upmanyu M and Huang H 2005 Size-dependent elasticity of nanowires: nonlinear effects *Phys. Rev.* B **71** 241403
[45] Cammarata R and Sieradzki K 1994 Surface and interface stresses *Annu. Rev. Mater. Sci.* **24** 215–34
[46] Cleland A and Roukes M 2002 Noise processes in nanomechanical resonators *J. Appl. Phys.* **92** 2758–69
[47] Martin Y, Williams C and Wickramasinghe H K 1987 Atomic force microscope–force mapping and profiling on a sub 100-Å scale *J. Appl. Phys.* **61** 4723–9
[48] Rugar D, Stipe B, Mamin H, Yannoni C, Stowe T, Yasumura K and Kenny T 2001 Adventures in attonewton force detection *Appl. Phys.* A **72** S3–S10
[49] Kenny T 2001 Nanometer-scale force sensing with MEMS devices *IEEE Sens. J.* **1** 148–57
[50] Ekinci K, Yang Y and Roukes M 2004 Ultimate limits to inertial mass sensing based upon nanoelectromechanical systems *J. Appl. Phys.* **95** 2682–9
[51] Johnson J B 1928 Thermal agitation of electricity in conductors *Phys. Rev.* **32** 97–109
[52] Nyquist H 1928 Thermal agitation of electric charge in conductors *Phys. Rev.* **32** 110–3
[53] Gabrielson T 1993 Mechanical-thermal noise in micromachined acoustic and vibration sensors *IEEE. Trans. Elect. Dev.* **40** 903–9
[54] Callen H and Welton T 1951 Irreversibility and generalized noise *Phys. Rev.* **83** 34–40
[55] Butt H and Jaschke M 1995 Calculation of thermal noise in atomic force microscopy *Nanotechnology* **6** 1–7
[56] Kopmaz O and Gundogdu O 2003 On the curvature of an Euler–Bernoulli beam *Int. J. Mech. Eng. Educ.* **31** 132–42
[57] Heidelberg A, Ngo L, Wu B, Phillips M A, Sharma S, Kamins T, Sader J and Boland J J 2006 A generalized description of the elastic properties of nanowires *Nano Lett.* **6** 1101–6
[58] Arlett J, Myers E and Roukes M 2011 Review article—comparative advantages of mechanical biosensors *Nat. Nanotechnol.* **6** 203–15
[59] Sheehan P E and Whitman L J 2005 Detection limits for nanoscale biosensors *Nano Lett.* **5** 803–7
[60] Patolsky F, Zheng G and Lieber C 2006 Nanowire sensors for medicine and the life sciences *Nanomedicine* **1** 51–65
[61] Stern E, Klemic J, Routenberg D, Wyrembak P, Turner-Evans D, Hamilton A, LaVan D, Fahmy T and

Reed M 2007 Label-free immunodetection with CMOS-compatible semiconducting nanowires *Nature* **445** 519–22
[62] Kim A, Ah C, Yu H, Yang J-H, Baek I-B, Ahn C-G, Park C, Jun M and Lee S 2007 Ultrasensitive, label-free and real-time immunodetection using silicon field-effect transistors *Appl. Phys. Lett.* **91** 103901
[63] Squires T, Messinger R and Manalis S 2008 Making it stick: convection, reaction and diffusion in surface-based biosensors *Nat. Biotechnol.* **26** 417–26
[64] Beaulieu L, Godin M, Laroche O, Tabard-Cossa V and Grutter P 2007 A complete analysis of the laser beam deflection systems used in cantilever-based systems *Ultramicroscopy* **107** 422–30
[65] Meyer G and Amer N 1988 Novel optical approach to atomic force microscopy *Appl. Phys. Lett.* **53** 1045–7
[66] Neubauer G, Cohen S, McClelland G, Horne D and Mate C 1990 Force microscopy with a bidirectional capacitance sensor *Rev. Sci. Instrum.* **61** 2296–308
[67] Knobel R and Cleland A 2002 Piezoelectric displacement sensing with a single-electron transistor *Appl. Phys. Lett.* **81** 2258–60
[68] White J 1993 An ultra high resolution displacement transducer using the Coulomb blockade electrometer *Jpn. J. Appl. Phys.* **32** L1571–3
[69] Blencowe M and Wybourne M 2000 Sensitivity of a micromechanical displacement detector based on the radio-frequency single-electron transistor *Appl. Phys. Lett.* **77** 3845–7
[70] Knobel R and Cleland A 2003 Nanometre-scale displacement sensing using a single electron transistor *Nature* **424** 291–3
[71] Arlett J, Maloney J, Gudlewski B, Muluneh M and Roukes M 2006 Self-sensing micro- and nanocantilevers with attonewton-scale force resolution *Nano Lett.* **6** 1000–6
[72] Barlian A, Park W, Mallon J R, Rastegar A and Pruitt B 2009 Review: semiconductor piezoresistance for microsystems *Proc. IEEE* **97** 513–52
[73] Harley J and Kenny T 1999 High-sensitivity piezoresistive cantilevers under 1000 Å thick *Appl. Phys. Lett.* **75** 289–91
[74] Kim Y-S, Nam H-J, Cho S-M, Hong J-W, Kim D-C and Bu J 2003 PZT cantilever array integrated with piezoresistor sensor for high speed parallel operation of AFM *Sens. Actuators* A **103** 122–9
[75] Itoh T and Sugar T 1994 Scanning force microscope using a piezoelectric microcantilever *J. Vac. Sci. Technol.* B **12** 1581–5
[76] Minne S C, Manalis S R, Atalar A and Quate C F 1996 Contact imaging in the atomic force microscope using a higher order flexural mode combined with a new sensor *Appl. Phys. Lett.* **68** 1427–9
[77] Cleland A and Roukes L 1996 Fabrication of high frequency nanometer scale mechanical resonators from bulk Si crystals *Appl. Phys. Lett.* **69** 2653–5
[78] Dohn S, Hansen O and Boisen A 2006 Cantilever based mass sensor with hard contact readout *Appl. Phys. Lett.* **88** 264104
[79] Bocko M 1990 The scanning tunneling microscope as a high-gain, low-noise displacement sensor *Rev. Sci. Instrum.* **61** 3763–8
[80] Kenny T, Kaiser W, Podosek J, Rockstad H, Reynolds J and Vote E 1993 Micromachined tunneling displacement transducers for physical sensors *J. Vac. Sci. Technol.* A **11** 797–802
[81] Busta H, Pogemiller J and Zimmerman B 1993 The field emitter triode as a displacement/pressure sensor *J. Micromech. Microeng.* **3** 49–56
[82] Yamashita K, Sun W, Kakushima K, Fujita H and Toshiyoshi H 2006 rf microelectromechanical system device with a lateral field-emission detector *J. Vac. Sci. Technol.* B **24** 927
[83] Axelsson S, Campbell E, Jonsson L, Kinaret J, Lee S, Park Y and Sveningsson M 2005 Theoretical and experimental investigations of three-terminal carbon nanotube relays *New J. Phys.* **7** 245–62
[84] le Fébre A, Abelmann L and Lodder J 2008 Field emission at nanometer distances for high-resolution positioning *J. Vac. Sci. Technol.* B **26** 724–9
[85] Kouh T, Karabacak D, Kim D and Ekinci K 2005 Diffraction effects in optical interferometric displacement detection in nanoelectromechanical systems *Appl. Phys. Lett.* **86** 013106
[86] Sanii B and Ashby P 2010 High sensitivity deflection detection of nanowires *Phys. Rev. Lett.* **104** 147203
[87] Beeby S, Ensell G, Kraft M and White N 2004 *MEMS Mechanical Sensors* (Boston, MA: Artech House Inc.) pp 94–97
[88] Vassalli M, Pini V and Tiribilli B 2010 Role of the driving laser position on atomic force microscopy cantilevers excited by photothermal and radiation pressure effects *Appl. Phys. Lett.* **97** 143105
[89] Nordström M, Zauner D, Calleja M, Hübner J and Boisen A 2007 Integrated optical readout for miniaturization of cantilever-based sensor system *Appl. Phys. Lett.* **91** 103512
[90] Levy O, Steinberg B Z, Nathan M and Boag A 2005 Ultrasensitive displacement sensing using photonic crystal waveguides *Appl. Phys. Lett.* **86** 104102
[91] Kozinsky I, Postma H, Bargatin I and Roukes M 2006 Tuning nonlinearity, dynamic range and frequency of nanomechanical resonators *Appl. Phys. Lett.* **88** 253101
[92] Ekinci K 2005 Electromechanical transducers at the nanoscale: actuation and sensing of motion in nanoelectromechanical systems (NEMS) *Small* **1** 786–97
[93] Schoelkopf R J, Wahlgren P, Kozhevnikov A A, Delsing P and Prober D E 1998 The radio-frequency single-electron transistor (RF-SET): a fast and ultrasensitive electrometer *Science* **280** 1238–42
[94] Steele G, Huttel A, Witkamp B, Poot M, Meerwaldt H, Kouwenhoven L and van der Zant H 2009 Strong coupling between single-electron tunneling and nanomechanical motion *Science* **325** 1103–7
[95] Aldridge J, Cleland A, Knobel R, Schmidt D and Yung C 2001 Nanoelectronic and nanomechanical systems *Proc. SPIE* **4591** 11–21
[96] He R and Yang P 2006 Giant piezoresistance effect in silicon nanowires *Nat. Nanotechnol.* **1** 42–46
[97] Li M, Tang H and Roukes M 2007 Ultra-sensitive NEMS-based cantilevers for sensing, scanned probe and very high-frequency applications *Nat. Nanotechnol.* **2** 114–20
[98] Sirohi J and Chopra I 2000 Fundamental understanding of piezoelectric strain sensors *J. Intell. Mater. Sys. Struct.* **11** 247–57
[99] Watanabe S and Fujii T 1996 Micro-fabricated piezoelectric cantilever for atomic force microscopy *Rev. Sci. Instrum.* **67** 3898–903
[100] Yang Y, Callegari C, Feng X, Ekinci K L and Roukes M L 2006 Zeptogram-scale nanomechanical mass sensing *Nano Lett.* **6** 583–6
[101] Ekinci K, Yang Y, Huang X and Roukes M 2002 Balanced electronic detection of displacement in nanoelectromechanical systems *Appl. Phys. Lett.* **81** 2253–5