Dissipation in nanoelectromechanical systems

Matthias Imboden, Pritiraj Mohanty *

Physics Department, Boston University, Boston, MA 02215, United States

A R T I C L E   I N F O

Article history:
Accepted 9 September 2013
Available online 25 September 2013

Editor: G. Bauer

Keywords:
NEMS
Dissipation
Mechanical resonator
Nanotechnology

A B S T R A C T

This article is a review of the dissipation processes in nanoelectromechanical systems (NEMS). As NEMS technology becomes more and more prevalent in research and engineering applications, it is of great importance to understand the dissipative mechanisms that in part define the dynamic response of such devices. The purpose of this work is to understand, sort, and categorize dominant dissipation sources and to determine their significance with respect to physics processes and engineering considerations.

© 2013 Elsevier B.V. All rights reserved.

Contents

1. Introduction .................................................................................................................................................. 90
2. Dissipation effects in NEMS .......................................................................................................................... 90
   2.1. Defining dissipation ................................................................................................................................. 90
   2.2. Resonator elasticity theory ......................................................................................................................... 92
   2.3. The standard linear solid model ............................................................................................................... 93
   2.4. Noise and dissipation ............................................................................................................................... 98
3. Nems devices .................................................................................................................................................. 100
   3.1. Resonator actuation and detection ............................................................................................................. 100
      3.1.1. The magnetomotive technique .......................................................................................................... 100
      3.1.2. The capacitive technique .................................................................................................................. 101
      3.1.3. The piezoelectric technique ............................................................................................................ 101
      3.1.4. Piezoresistive detection ................................................................................................................... 102
      3.1.5. The dielectric technique .................................................................................................................. 102
      3.1.6. Other actuation and detection techniques .......................................................................................... 102
   3.2. Fabrication ................................................................................................................................................ 102
4. Extrinsic sources of dissipation ....................................................................................................................... 103
   4.1. Circuit damping $Q^{-1}$ ............................................................................................................................ 103
      4.1.1. Magnetomotive damping ................................................................................................................... 103
      4.1.2. Capacitive circuit damping .............................................................................................................. 104
      4.1.3. Piezoelectric damping ..................................................................................................................... 105
   4.2. Clamping losses $Q^{-1}$ ............................................................................................................................ 106
   4.3. Dissipation due to multiple materials ....................................................................................................... 109
   4.4. Resonators in a medium ............................................................................................................................ 110
      4.4.1. Low pressure dissipation $Q^{-1}_{dp}$ .................................................................................................. 110
      4.4.2. Viscous dissipation $Q^{-1}_{vd}$ ........................................................................................................... 111
      4.4.3. Squeeze-film damping $Q^{-1}_{sf}$ ....................................................................................................... 111

* Corresponding author. Tel.: +1 617 353 9297; fax: +1 617 353 9393.
E-mail addresses: m.imboden@gmail.com (M. Imboden), mohanty@bu.edu (P. Mohanty).

0370–1573/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.physrep.2013.09.003
5. Intrinsic sources of dissipation

5.1. Thermoelastic dissipation $Q^{-1}_{\text{TED}}$ ................................................................. 112
5.2. Surface losses $Q^{-1}_{\text{SL}}$ ...................................................................................... 116
5.3. Dissipation due to mechanical defects $Q^{-1}_{\text{MD}}$ .................................................... 117
5.4. Quantum dissipation $Q^{-1}_{\text{QD}}$ .................................................................
   5.4.1. Standard tunneling model of two-level systems ............................................. 120
   5.4.2. Variations on the standard tunneling model ............................................. 121
   5.4.3. TLS in NEMS .......................................................................................... 122

6. Dissipation in carbon materials .................................................................................. 124

6.1. Carbon nanotubes ................................................................................................... 124
6.2. Graphene .................................................................................................................. 128

7. Evading dissipation .................................................................................................... 130

7.1. High tensile-strain .................................................................................................. 130
7.2. Surface and heat treatment .................................................................................... 133
7.3. Dissipation control circuits .................................................................................... 134

8. Nonlinear dissipation ................................................................................................ 135

9. Concluding remarks ................................................................................................ 139

Acknowledgments ........................................................................................................ 140

References .................................................................................................................... 140

1. Introduction

Nanoelectromechanical systems (NEMS) are establishing themselves as viable commercial technologies and are becoming more and more prevalent in research applications. Current uses range from gyroscopes to timing oscillators [1] and accelerometers, among others. Devices are becoming smaller, faster and more sensitive. In addition, functionalization and hybrid structures make this emerging technology suitable for biological applications as well. As tools for basic research, NEMS have demonstrated extraordinary sensitivity to external forces [2], allowing for example, the detection of electron spin flips [3], single molecules [4–8], and even macroscopic quantum states [9]. Quantum computation are now within reach [12], in addition to single spin detection [13], gravitational wave detection [11], and other fundamental phenomena [14,15].

The fundamental characteristic of a mechanical resonator are determined by the resonance frequency $f_0$ and dissipation (inverse quality factor) $Q^{-1}$. In particular, better device response is obtained by high frequency and low dissipation or loss. Reduced dimensions are necessary for achieving high resonance frequencies. However, miniaturization beyond the submicron scale leads to increased dissipation. Dissipation studies of miniaturized resonators have been conducted for decades, starting with tuning forks in the late 1960’s by Newell [16]. Modern nanomachining, such as e-beam lithography and ion milling, have resulted in orders of magnitude reduction in the length scales and increase in resonance frequency of mechanical resonators. Correspondingly, many different dissipation sources have been identified, depending on the operational parameter space of the NEMS device.

Here we study dissipation sources in very basic resonator structures, predominantly cantilevers and doubly-clamped beams, and identify the dominant dissipation mechanisms over a wide parameter space. Typical frequencies range from 1 MHz to 10 GHz and device length ranges from 1–20 µm; typical thickness and width are ≤ 1 µm. The structure materials presented here are metallic and/or dielectric, amorphous, or crystalline, and where instructive, larger MEMS structures or thin films are also discussed. Considerations also apply to torsional oscillators of the equivalent dimensions.

The purpose of this work is to illustrate the dominant dissipation mechanism of state-of-the-art NEMS devices. Dissipation sources are discussed to the depth needed for the reader to understand the basic principles. Those interested in the field can use this as a guide to determine which sources of dissipation contribute to their particular resonator setup. References are given for those interested in further study of the ideas presented here.

2. Dissipation effects in NEMS

Dissipation effects appear in all mechanical systems and help define fundamental dynamical behavior. Understanding their source and relevance is important, not only for design and applications in engineering but also to understand the inner workings of the materials and mechanical structures involved. Low dissipation is generally desirable, which makes a device more efficient, less susceptible to wear and mechanical noise, and more sensitive [17]. NEMS provide an unprecedented opportunity to systematically investigate dissipation effects over a large range of parameter space. Due to their small size, the surface-to-volume ratio increases, and finite volume effects come into play [18]. It is also possible to cool the structures to sub-kelvin temperatures, a regime where quantum effects become relevant [19–22].

Dissipation is a measure of energy lost per oscillation as a ratio of the total energy in the resonator. Just as the relevant forces change on the nanoscale, so do the mechanisms that transfer mechanical energy out of the resonant mode. The dissipative mechanism are depicted in Fig. 1. They are organized in two categories: extrinsic and intrinsic. Specifically, dissipation mechanisms such as clamping loss, thermoelastic loss, and anharmonic mode coupling are compared and
Fig. 1. Extrinsic and intrinsic sources of dissipation commonly found in NEMS devices. The color on the doubly-clamped beam is coded by stress. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Fig. 2. Scaling of the quality factor with miniaturization of the mechanical structures. The resonator categories range from macroscopic to nanoscale devices and are split into Large Scale [18], High Tension [23–25], CNT [26–28], Graphene [29,30], Capacitive Silicon [31,32], Magnetomotive Diamond [33], and Low Temperature Diamond [34].

contrasted with the contributions arising from intrinsic defects in the bulk and the surface. Geometric and external properties are identified at room temperature as the dominant contribution to dissipation in NEMS resonators. These effects are, in first approximation, independent of the material properties. It is discussed how design can minimize these losses. For sub-kelvin temperatures, quantum dissipation due to two-level systems is studied. These studies help develop the next generation of NEMS devices, as well as identify intriguing mechanisms in materials at the nanoscale.

The fundamental characteristics of a mechanical resonator are determined by the resonance frequency $f_0$ and dissipation $Q^{-1}$ (inverse quality factor). In particular, better device response is obtained at high frequency and low dissipation or loss. Reduced dimensions are necessary for achieving high resonance frequencies. However, miniaturization beyond the submicron scale leads to increased dissipation. Dissipation studies of miniaturized resonators have been conducted for decades, starting with tuning forks in the late 1960’s by Newell [16]. Modern nanomachining, such as e-beam lithography and ion milling, have resulted in the reduction of the length scales by many orders of magnitude and, thereby in an increase in resonance frequency. Correspondingly, many different dissipation sources have been identified, depending on the operational parameter space of the NEMS device. Where scaling may include many benefits, as Fig. 2 illustrates, dissipation scales unfavorably. The high dissipation observed on the nanoscale results in a barrier to further device performance and counters some of the benefits of downscaling. While the scaling follows a rough $Q \propto V^{1/3}$ trend over a large size range, arguments for this behavior can only be done within the context of the specific dominant dissipation mechanisms.

Although dissipation is often undesirable, it is fundamentally necessary for any transduction scheme. Dissipation enables coupling to the environment as well as the readout circuit and allows for the transduction of a signal by transferring energy from one system to another. Without dissipation, no actuation or detection would be possible. The response would be limited to an infinitely sharp frequency that could never be accessed. Correspondingly, a high quality factor may help discriminate between different signal sources (through a narrow bandwidth), or inversely, high dissipation may be desired to lower the transient time and widen the bandwidth, thereby allowing a response to a wider range of sources. The coupling to many degrees of freedom and the interaction between these processes is what results in measurable changes with respect to outside influences, including pressure, viscosity, and temperature. In recent years, NEMS resonators have approached the
quantum limit, and it has even been possible to detect macroscopic quantum mechanical states [9]. Dissipative processes contribute to the quantum coherence time and are hence critical in the device performance.

Dissipation is defined and measured in the linear regime. This means that the dissipation is amplitude-independent and manifests itself as a force proportional to the velocity of the structure. Nonlinear dissipation, i.e., energy loss that is a function of device energy or amplitude, is also observed in nanostructures. NEMS provide an ideal framework to study this elusive phenomenon. Currently, no physical mechanism has been identified that can explain this intriguing behavior.

2.1. Defining dissipation

While the calculation and measurement of resonance frequency is straightforward (elasticity theory still applies for extremely small structures [35], and frequencies can be readily measured better than 1 ppm), characterization and understanding of dissipation are much more elusive. Dissipation (Q−1) is a measure of the energy lost per cycle of oscillation (∆W) as a fraction of the total mechanical energy (W0) of the resonator. If the drive force of a resonator abruptly stops, then dissipation will determine for how long the resonator continues to oscillate. Hence, it also defines the transient time of the dynamic system, a measure of how long it takes for the harmonic response of the resonator to be independent or “forget” the initial conditions. The quality factor Q, which is the inverse of the dissipation, along with the frequency f0, are the two most important characteristics of a resonator. In its most general form, dissipation is written as:

$$Q^{-1} = \frac{\Delta W}{2\pi W_0} \approx \frac{\gamma}{\omega_0} \approx \frac{\Delta \omega_0}{\omega_0}.$$  (1)

The first term is the expression for the standard definition of dissipation. The latter two expressions are obtained by solving the standard damped resonator equation; ω0 is the resonance frequency; and γ is the friction, or damping. It appears as the coefficient of the resonator velocity. Δω0 is the full width at half maximum (FWHM) of the Lorentzian fit to the amplitude-drive frequency response. The last term is an approximation, valid for low dissipation and typical for quality factors above 100. The dissipation in the linear response is amplitude- or mechanical-energy-independent. In this framework, dissipation appears as a force term proportional to the resonator’s velocity opposing motion, though this may not always be conceptually the most intuitive picture. It is very difficult to predict exact values of dissipation in NEMS devices. Experiments and theoretical studies generally set out to determine the order of magnitude and dependency of dissipation on a given parameter, such as temperature T, pressure P, frequency f0, or the device dimensions d, w and L.

The total mechanical energy in a mode can be calculated from the volumetric strain $\epsilon(\bar{x}, t) = \epsilon_{\max}(\bar{x}) \sin(\omega t)$ (assuming sinusoidal oscillation) [36]

$$W_0 = \int_V \int_{\epsilon_{\max}} \sigma d\epsilon dV = \frac{1}{2} \int_V E\nu \epsilon_{\max}^2(\bar{x}) dV,$$  (2)

where $\sigma$ is the longitudinal stress and the integral is over the volume of the resonator. The maximal strain energy must be equal to the maximum kinetic energy

$$W_0 = \frac{1}{2} \int_L d\lambda \rho \dot{\lambda}^2 X(x)^2,$$  (3)

where it is assumed that displacement X(x), perpendicular to the beam length (defined in Section 2.2), is small and the beam length $L \gg d, w$ is much greater than beam thickness and width. Following Unterreithmeier et al. [37], the general expression for the strain of a resonator is given by

$$\epsilon(x, z, t) = \frac{1}{2} \left( \frac{\partial_x X(x) \hbar(\exp(i\omega t))}{\partial_x X(x)} \right)^2 + z \frac{\partial_z X(x) \hbar(\exp(i\omega t))}{\partial_z X(x)},$$  (4)

where x, measured from the neutral plane, is the position along the length of the beam, and z is the position along the thickness. It is assumed that the stress is constant over the width. For flexural modes with no intrinsic strain, the strain due to bending dominates the strain caused by elongation. Other forms must be considered for breathing modes. The resulting strain energy in a beam takes the form:

$$W_0 \approx \int_L dx \frac{E\nu w d}{8} \left[ (\partial_x X)^4 + \frac{d^2}{3} (\partial_z X)^2 \right]$$

$$\approx \int_L dx \frac{E\nu w d^3}{24} (\partial_z X)^2,$$  (5)

where it is assumed that the low strain limit applies and higher order terms in $\epsilon$ and $\partial_x \epsilon$ can be neglected. The first term in the integral is the elongation strain energy that vanishes for diminishing intrinsic strain, and the second term is the bending strain energy. To complete the calculation of the dissipation defined in Eq. (1), an expression for $\Delta W$ is needed.
The dissipation can be introduced though a complex Young’s modulus $E_I \rightarrow E_0 + iE_{Im}$, as is discussed in more detail in Section 2.3. The energy loss is then given by integrating the complex component of Eq. (5) over one oscillation period. Assuming bending energies dominate, one finds:

$$\Delta W \approx \pi E_{Im} \frac{w_0^3}{12} \int_L dx (\partial_x^2 X(x))^2.$$  \hfill (6)

Combining expressions (5) and (6), dissipation can now be expressed as

$$Q^{-1} = \frac{E_{Im}}{E_0},$$  \hfill (7)

where $E_0$ here is the relaxed Young’s modulus valid in the low frequency limit (where the susceptibility function defined below vanishes).

The origin of this complex term and how it is measured, how it is calculated and how to determine its significance is the topic of this paper.

2.2. Resonator elasticity theory

A detailed description of elasticity theory is given by Weaver et al. [38]. Specifically, for MEMS and NEMS, elasticity is comprehensively introduced in Ref. [39]. Here, we include a brief overview and build a framework from which dissipation can be interpreted. The Euler–Bernoulli model describes the deflection of a solid due to an applied load. With appropriate simplifications for the thin-beam approximation and no damping, the model takes the following form:

$$EI \frac{\partial^4 D(x, t)}{\partial x^4} + \rho A \frac{\partial^2 D(x, t)}{\partial t^2} = F_{dr},$$  \hfill (8)

where $E$ is the Young’s modulus (or tensile modulus) and $I$ the second moment of inertia ($I = \frac{w_0 l}{12}$ for a rectangular beam cross section, actuated in the thickness direction). $\rho$ is the mass density and $A = w t$ the cross-sectional area, where $w$ and $t$ are the width and thickness, respectively. $x$ is the distance along the length of the beam $L$, and $D$ is the displacement from equilibrium due to the drive force $F_{dr}$. This equation, fourth order in space and second in time, can be solved by separation of variables using $D(x, t) = X(x)T(t)$. We are interested in the homogeneous solution, which can be added to the particular solution if the transient behavior is of interest. Here, we consider the solution after the transient time has lapsed:

$$- \frac{EI}{\rho A} \frac{\partial^2 T}{\partial t^2} \frac{1}{\partial x^4} X + \frac{\partial^2 T}{\partial t^2} = -\omega^2.$$  \hfill (9)

The time component is of second order and is solved by

$$T(t) = a \sin(\omega t + \phi),$$  \hfill (10)

where $a$ and $\phi$ are determined by the boundary conditions. For the spacial term $X(x)$, the most general solution takes the form:

$$X(x) = A \sin(\beta x) + B \cos(\beta x) + C \sinh(\beta a) + D \cosh(\beta x),$$  \hfill (11)

where we define $\beta = \left(\frac{\omega_0^2}{EI}\right)^{\frac{1}{2}}$. $A, B, C, D$ are given by the boundary conditions. For example, a doubly-clamped beam has no displacement at either end and the slope at both clamping points vanishes as well (see Fig. 3). Hence, the boundary conditions take the form:

$$D(0, t) = D(L, t) = 0$$

$$\frac{dD}{dx}(0, t) = \frac{dD}{dx}(L, t) = 0.$$  \hfill (12)

For the cantilever case, one uses the zero transverse force, $\frac{d^2 D}{dx^2}(L, t) = 0$ and zero torque $\frac{d^2 D}{dx^2}(L, t) = 0$ boundary conditions. One quickly finds that $A = -C$ and $B = -D$. To obtain a non-trivial solution, restrictions are imposed on $\beta$ by setting the determinant of the set of linear equations to zero. For the beam structure this results in the condition that

$$\cos(\beta L) \cosh(\beta L) - 1 = 0.$$  \hfill (13)

This equation can only be solved numerically. It results in a discrete set of solutions for $\gamma_n = \beta_n L$, where $n$ corresponds to the discrete harmonics of the resonant beam. The allowed frequencies are given by

$$\omega_n = \frac{\gamma_n^2}{2\pi} \sqrt{\frac{E}{12\rho l^2}} = 1.028 \sqrt{\frac{E}{\rho l^2}},$$  \hfill (14)

where the final expression holds for the fundamental mode. Eq. (14) is of great importance, as it describes the resonance response for a beam oscillator. It is composed of a geometric term $\frac{1}{l^2}$ that illustrates how the frequency increases with
Table 1
Material properties used in NEMS resonators. Values are given for room temperature. For some materials, considerable ranges have been experimentally reported. It is common for material properties to vary for thin films from which NEMS are manufactured. (Sources: a www.ceramics.nist.gov, b [63–65], c [64], d [66,67], e www.matweb.com [68,69], f [70–73], g [37,74–77], h www.memsnet.org, i www.accuratus.com.) j [78], k [79]. Values for NCD, UNCD, graphene, and SWCNT are approximate; large variations in the literature can be found.

<table>
<thead>
<tr>
<th>Material</th>
<th>$E$ (GPa)</th>
<th>$\nu$</th>
<th>$\rho$ (kg/m$^3$)</th>
<th>$C_p$ (J/(kgK))</th>
<th>$\alpha$ ($10^{-6}$/K)</th>
<th>$\kappa$ (W/(mK))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si</td>
<td>124</td>
<td>0.28</td>
<td>2329</td>
<td>702</td>
<td>2.49</td>
<td>124</td>
</tr>
<tr>
<td>SiC$^a$</td>
<td>415</td>
<td>0.16</td>
<td>3160</td>
<td>670</td>
<td>4.6</td>
<td>114</td>
</tr>
<tr>
<td>Diamond SCD</td>
<td>1200</td>
<td>0.2</td>
<td>3500</td>
<td>2760–3490</td>
<td>1.5–3.8</td>
<td>2000</td>
</tr>
<tr>
<td>NCD$^b$</td>
<td>800–1100</td>
<td>0.069</td>
<td>3500</td>
<td>$\approx$1</td>
<td>$\approx$1000</td>
<td></td>
</tr>
<tr>
<td>UNCD$^c$</td>
<td>450–1000</td>
<td>0.057</td>
<td>3500</td>
<td>1.5</td>
<td>1–12</td>
<td></td>
</tr>
<tr>
<td>C amorphous$^d$</td>
<td>631</td>
<td>0.17</td>
<td>3050</td>
<td>600</td>
<td>1.67</td>
<td>3</td>
</tr>
<tr>
<td>Graphite$^e$</td>
<td>4.8</td>
<td>0.126</td>
<td>2250</td>
<td>708</td>
<td>0.6–4.3</td>
<td>24</td>
</tr>
<tr>
<td>SWCNT$^f$</td>
<td>1000</td>
<td>0.3</td>
<td>1350</td>
<td>600</td>
<td>$\approx$10.5</td>
<td>500–2500</td>
</tr>
<tr>
<td>Graphene$^{g,i}$</td>
<td>1000</td>
<td>0.165</td>
<td>2250</td>
<td>700</td>
<td>$\approx$7.4</td>
<td>500–2500</td>
</tr>
<tr>
<td>GaAs$^{e,g}$</td>
<td>85.5</td>
<td>0.31</td>
<td>5316</td>
<td>325</td>
<td>5.4</td>
<td>50</td>
</tr>
<tr>
<td>AlN$^{e,h}$</td>
<td>350</td>
<td>0.24</td>
<td>3260</td>
<td>819</td>
<td>4.5</td>
<td>180</td>
</tr>
<tr>
<td>SiN$^{e,h}$</td>
<td>160–230</td>
<td>0.27</td>
<td>2800–3300</td>
<td>170</td>
<td>3.3</td>
<td>30</td>
</tr>
<tr>
<td>Au$^a$</td>
<td>77.2</td>
<td>0.42</td>
<td>19320</td>
<td>128</td>
<td>14.4</td>
<td>300</td>
</tr>
<tr>
<td>Al$^f$</td>
<td>68</td>
<td>0.35</td>
<td>2699</td>
<td>900</td>
<td>24</td>
<td>210</td>
</tr>
<tr>
<td>Ti$^a$</td>
<td>110</td>
<td>0.34</td>
<td>4500</td>
<td>518</td>
<td>8.9</td>
<td>17</td>
</tr>
</tbody>
</table>

![Fig. 3.](image)
(a) Doubly-clamped beam and support pads (purple) including undercut and pedestal, often silicon oxide (red) and base (gray). $D_0$ and $D_o$ refer to the actuated displacement in and out of plane, respectively. (b) Harp structure consisting of three doubly-clamped beams of various lengths. The polycrystalline diamond used here for the beam material shows significant surface roughness. The structures thicknesses is $\approx$1 µm. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The resulting beam equation takes the form:

$$X(x) = A_n \left[ \sin \left( \frac{\gamma_n x}{L} \right) - \sinh \left( \frac{\gamma_n x}{L} \right) \right] + B_n \left[ \cos \left( \frac{\gamma_n x}{L} \right) - \cosh \left( \frac{\gamma_n x}{L} \right) \right].$$

Finally, the relation between $A_n$ and $B_n$ is given by

$$\frac{A_n}{B_n} = \frac{\cos(\gamma_n) - \cosh(\gamma_n)}{\sin(\gamma_n) - \sinh(\gamma_n)}.$$
With the inclusion of the effect of dissipation, the dynamical equation for the bending modes takes the form:

\[-\frac{12\rho}{d^2} \frac{\partial^2 D}{\partial t^2} = (E_R + \chi) \frac{\partial^4 D}{\partial z^4}.\]  

(17)

Here, \(E_R\) is the relaxed, low frequency Young's modulus and \(\chi\) is the complex-valued susceptibility function, effectively making the Young's modulus complex.

We assume a solution of the form

\(D(x, t) = Ae^{i(kx - \omega t)}\),

where \(k\) is defined by the complex dispersion relation

\(\omega = \sqrt{E_R + \chi} / (12\rho) k^2\).

Assuming small shifts caused by dissipation (low dissipation regime \(|\chi| \ll E_R\)), \(\omega = k^2/\sqrt{(E_R^2)/(12\rho)}(1 + \chi/2E_R)\) the solution to Eq. (17) is of the form

\(D(z, t) = Ae^{i(kx - \omega_0 t)} e^{-\gamma t}\),

(18)

where \(\omega_0 = k^2/(E_R^2)/(12\rho)\).

The following expressions for frequency shift and dissipation are obtained:

\[\frac{\delta f}{f_0} = \frac{\Re(\chi)}{2E_R}\]

(19)

\[Q^{-1} = \frac{2\Im(\omega_0)}{\omega_0} = \frac{\Im(\chi)}{E_R}\]

(20)

where the standard definition of damping is \(e^{-\omega_0 t^2/2\pi}\) and \(Q^{-1} = \frac{\chi}{\omega_0}\) (for under-damped systems with low dissipation).

The susceptibility function is given by:

\[\chi(\omega) = E_R \left( \frac{2\delta f}{f_0} + iQ^{-1} \right)\].

(23)

The preceding derivation is taken from work done by Seoanez, Guinea, and Castro Neto [40].

A one-dimensional model serves well to illustrate conceptually many aspects of the resonator system. For many considerations, the beam structure can be reduced to a point particle with an effective mass and spring constant, reducing the problem to the familiar damped driven harmonic oscillator. While the previous section describes the dynamic extended shape of a resonator, the damped driven harmonic oscillator equation describes the time-dependent motion, which is the same for each point along the beam, defined by the differential equation:

\[\ddot{x} + \gamma \dot{x} + \omega_0^2 x = \frac{F_d}{m} e^{i\omega_d t}\],

(24)

where \(\gamma = \omega_0^2 Q\) is the damping term proportional to the velocity, \(\omega_0\) is the undamped, undriven resonance frequency, and \(F_d\) and \(\omega_d\) are the drive force and drive frequency, respectively. The displacement \(x(t)\) can be derived in frequency space by Fourier transforming equation (24) to obtain

\[\tilde{x}(\omega_d) = \frac{\tilde{F}_d}{m\omega_0 - \omega_d^2 + i\omega_0 Q}\].

(25)

As measurements are commonly done using a network analyzer or a spectrum analyzer in frequency space, this form is very useful. The derived response can be split into real and imaginary components or amplitude and phase responses as shown in Fig. 4. With the fit to the Lorentzian equation, the amplitude trace directly measured by a network analyzer or lock-in amplifier is used to determine the resonance frequency (\(\omega_R\)). The full width at half maximum (\(\Delta \omega\)) is related to the damping coefficient \(\gamma\). The term \(\tilde{F}_d\) will depend on the drive method, some of which are described in the next section. Note that the natural resonance frequency is shifted due to the drive and damping to become

\[\omega_R = \omega_0 \sqrt{1 - \frac{1}{2Q^2}}\]

\[\phi = \arctan \left( \frac{\omega_0 \omega_d}{Q (\omega_0^2 - \omega_d^2)} \right)\].

(26)

where the phase response is not shifted in frequency space.
Fig. 4. Amplitude and phase response of a damped driven harmonic oscillator $\omega = \omega_0 \sqrt{1 - \frac{1}{2Q^2}}$.

Fig. 5. Standard Linear Solid Model representation. $\epsilon_i$ is the strain acting on each spring, with spring constant $k_i$, $\sigma$ is the stress on the system and $\eta$ the viscosity.
Source: Reprinted with permission from Ref. [41].

2.3. The standard linear solid model

One framework to introduce dissipation is through the Standard Linear Solid Model (SLSM), also known as the Zener Model for anelastic solids. This model describes the intrinsic dissipation mechanism for viscoelastic solids that fully recover to their initial state after a load is removed. (A necessary condition for the assumption that the dissipation is linear and amplitude-independent.) Hooke's law is modified to allow for a finite relaxation and creep time for the stress and strain, respectively. The model, applied to viscoelastic materials, combines the Maxwell and Kelvin–Voigt models and can be represented by two springs and a viscous component illustrated in Fig. 5. This setup models instantaneous elastic strain when a stress is applied, where under constant stress the strain follows asymptotically, and under a constant strain, the stress relaxes asymptotically. The recovery and relaxation of the strain and stress result in characteristic timescales.

Following the derivations by Casula et al. [41] and Cleland and Roukes [42], the modified Hooke's law takes the form

$$\sigma = E_Y \epsilon \rightarrow \sigma + \tau_\sigma \frac{d\sigma}{dt} = E_R \left( \epsilon + \tau_\epsilon \frac{d\epsilon}{dt} \right),$$

(Eq. (27) can also be obtained by Fourier transforming equation (27)). For very slow resonators modes, where $\omega \tau \ll 1$, the Young’s modulus is $E_R$, as expected. For extremely high frequencies, where $\omega \tau \gg 1$, the Young’s modulus becomes
the unrelaxed Young’s modulus \( E_U = E_R \frac{\tau_e}{\tau} \). For the intermediate frequencies, the Young’s modulus is complex and can be written in terms of real and imaginary components

\[
E_Y = E_{\text{eff}} \left( 1 + \frac{i \omega \Delta}{1 + \omega^2 \tau^2} \right),
\]

where \( \tau = \sqrt{\frac{\tau_e}{\tau}} \) is the arithmetic mean of the relaxation times and \( \Delta \tau = \tau_e - \tau \) is a measure of the difference of the relaxation times. \( E_{\text{eff}} \) is the effective Young’s modulus that characterizes the static stiffness of the material. Defining \( \Delta = \frac{\Delta \tau}{\tau} \) as the fractional difference of the relaxation time further simplifies the expression. Note that \( \Delta = \frac{E_U - E_R}{\sqrt{E_R E_U}} \) can also be expressed as the fractional difference of the relaxed and unrelaxed Young’s moduli. Now the complex term is expressed in terms of \( \tau \) alone, a sufficient approximation for the low dissipation regime. This term will be used from now on, as working with a dimensionless parameter simplifies matters. The effective Young’s modulus is given by the real component of Eq. (30)

\[
E_{\text{eff}} = \frac{1 + \omega^2 \tau^2}{1 + \omega^2 \tau^2} E_R.
\]

To understand how this relates to dissipation, one can consider the strain energy as the resonator cycles through one period. The mechanical energy density integrated over one period is given by

\[
\Delta w = \int \mathcal{H}(\sigma) d\sigma \Delta \epsilon = \int_0^{2\pi} \mathcal{H}(\sigma) \left( \frac{d\epsilon}{dt} \right) dt.
\]

Using

\[
\mathcal{H}(\sigma) = \epsilon_0 (E_{\text{eff}} \cos(\omega t) - E_{\text{im}} \sin(\omega t))
\]

\[
\frac{d\epsilon}{dt} = -\epsilon_0 \omega \sin(\omega t),
\]

where over one period the \( \cos(\omega t) \sin(\omega t) \) term will vanish, leaving

\[
\Delta w = \pi \epsilon_0^2 E_{\text{im}}.
\]

Assuming the system is released from maximum strain at \( t = 0 \), the maximum strain energy is given by

\[
w_{\text{tot}} = \int \mathcal{H}(\sigma) \Delta \epsilon |_{t=0} \Delta \epsilon = \frac{1}{2} \epsilon_0^2 E_{\text{eff}}.
\]

Defining the dissipation as the amount of energy lost per cycle as a fraction of total energy, one obtains

\[
Q^{-1} = \frac{\Delta w}{2\pi w_{\text{tot}}} = \frac{\pi \epsilon_0^2 E_{\text{im}}}{2\pi \frac{1}{2} \epsilon_0^2 E_{\text{eff}}}
\]

\[
= \frac{E_{\text{im}}}{E_{\text{eff}}} = \frac{\omega \tau}{1 + \omega^2 \tau^2} \Delta.
\]

For the SLSM, the susceptibility function, defined in Section 2.2, takes the form

\[
\chi = E_R \frac{\omega^2 (\tau^2 - \tau_e^2)}{1 + \omega^2 \tau^2} + i E_R \frac{\omega (\tau_e - \tau)}{1 + \omega^2 \tau^2}.
\]

An expression for the relative frequency shift can be written in terms of the relaxation times \( \tau \):

\[
\frac{\delta f}{f_0} = \frac{1}{2} \frac{\omega^2 (\tau^2 - \tau_e^2)}{1 + \omega^2 \tau^2} = \frac{1}{2E_R} \frac{\delta \mathcal{H}(\chi)}{\chi}.
\]

In expression (35) for the dissipation, there are two terms; \( \Delta \), which gives the magnitude of the dissipation, and \( \frac{\omega \tau}{1 + \omega^2 \tau^2} \), which corresponds to the dynamics of the dissipative mechanisms. One can extract from this expression a number of properties of the dissipation that are independent of the dissipative mechanism at play (illustrated in Fig. 6)

- for \( \omega \tau \ll 1 \) \( Q^{-1} \propto \omega \) isothermal limit
- for \( \omega \tau \gg 1 \) \( Q^{-1} \propto \omega^{-1} \) “frozen” limit
- \( Q_{\text{max}}^{-1} = \frac{1}{2} \Delta \) for \( \omega = \tau^{-1} \).

For intrinsic losses material properties define dissipation. The geometry defines the resonance frequency and can be used to shift the dynamics to a parameter space where dissipation mechanisms are suppressed. When low dissipation is desired, one must make sure that the resonator’s mechanical timescale \( (f_0^{-1}) \) does not match intrinsic dissipative timescales that
define \( \tau \). Some processes may be explained by a distribution of timescales; this can be incorporated in the model by including a number of SLSM elements in series and/or parallel as discussed by Casula et al. [41], each resulting in dissipation peaks as the frequency is swept. Correspondingly, timescale independent \( Q \) models can be forged by a series of parallel SLSM elements with varying time constants. Generally speaking, the SLSM applies whenever there is a process that introduces a timescale that establishes a phase relation between strain and stress. For example, when considering thermoelastic effects in the isothermal limit, there is no transfer of energy. The disturbance is slow enough that no phase difference between the strain and stress can be established. (Physically, this would manifest itself as a temperature gradient.) In the high-frequency adiabatic limit, the disturbance is fast and the response time is correspondingly slow. Hence, no energy transfer can occur. In this limit, the dissipative mechanisms become ‘frozen’, as the reaction time is too slow to thermalize the structure. A temperature gradient is established, but it varies so quickly that no heat flow occurs. Alternatively, energy conversion processes or transfer processes exist that remove energy from the resonator independent of such timescales. In these cases, the SLSM model is not a useful description of dissipation.

2.4. Noise and dissipation

Finite dissipation is integral to the laws of thermodynamics. Through dissipation, a resonator will “feel” the thermodynamic bath and can thereby find its thermal equilibrium. In the resonator model here, where the dissipation is proportional to the velocity, this process is described by the Langevin equation. Following Cleland [39], the Langevin equation is written as

\[
\ddot{x} + \gamma \dot{x} = -m\gamma x + F_N(t),
\]

where \( \gamma \) is the damping factor introduced in Eq. (24), and \( F_N(t) \) is a force introduced by random noise. The ensemble average results in the thermal equilibrium:

\[
2mk_BT\gamma = \int_{-\infty}^{\infty} \langle F_N(0)F_N(t') \rangle dt'.
\]

As shown in Ref. [39], introducing \( K(s) = \langle F_N(t)F_N(t + s) \rangle \) as the correlation function, results in a spectral energy density given by

\[
S(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} K(s)e^{i\omega s} ds.
\]

For uncorrelated force noise the correlation function is a delta function and the spectral noise force results in

\[
S_f(\omega) = \frac{m\gamma k_BT}{2\pi}.
\]

This frequency-independent white noise is a result of any dissipative system in thermal equilibrium, where dissipation is introduced by the Langevin equation and is proportional to the velocity of the oscillator. The equipartition function states that, for a harmonic oscillator, thermal energy is shared between all degrees of freedom equally: in this case, kinetic and potential terms in Eq. (38). The effect of the Langevin equation is to dampen all motion in a given mode through the term proportional to the velocity, but adding and removing energy stochastically through the random force term, resulting in a thermal energy of

\[
\langle E \rangle = \frac{1}{2}m\langle \dot{x}^2 \rangle + \frac{1}{2}k\langle x^2 \rangle = k_BT.
\]

A more detailed description of the preceding discussion is given by Cleland [39] and the equilibrium energy is calculated in Ref. [43].
Finite dissipation inevitably leads to noise in the system, as required by the fluctuation–dissipation theorem. The dissipation provides the coupling to the environment that allows the mode to thermalize with its surroundings. Thermodynamics states that each degree of freedom has $\frac{1}{2} k_B T$ energy. Each mode contains two degrees of freedom, amplitude and velocity, resulting in a total energy of $k_B T$. As described by Ref. [39], thermalization occurs due to a noise force $f_0$ that forces the beam in a manner that is uncorrelated in space and time. This force noise is characterized the spectral density defined by

$$S_f(\omega) = \int_{-\infty}^{\infty} f(t_0)f(t_0 + t) \exp^{i\omega t} dt.$$  \hfill (43)

From this, and the force amplitude relation of a damped, undriven harmonic oscillator, one can derive the spectral amplitude noise for a given mode $n$

$$S_{\text{a}_n} = \frac{1}{(\omega_n^2 - \omega^2)^2 + (\frac{\omega_n}{Q})^2} \frac{S_f(\omega)}{m^2}.$$  \hfill (44)

Using the requirement that the kinetic energy of a given mode is $\frac{1}{2} k_B T$, Ref. [39] shows that

$$S_{\text{a}_n} = \frac{4k_B T\omega_n}{Q}.$$  \hfill (45)

Hence, any resonator mode, independent of dissipation mechanisms or geometry, exhibits spectral amplitude noise of the form

$$S_{\text{a}_n} = \frac{1}{(\omega_n^2 - \omega^2)^2 + (\frac{\omega_n}{Q})^2} \frac{4k_B T\omega_n}{m^2Q}.$$  \hfill (46)

This expression for mechanical thermal amplitude noise (or thermal Johnson noise) is generic for any white-noise source, which is defined by a frequency-independent noise force, where $k_B T$ must be replaced by the noise energy that couples to the resonator. This expression holds true for high temperatures and low frequencies ($\hbar \omega \ll k_B T$). The resulting displacement on resonance is given by

$$\frac{\delta x_{\text{thermal}}}{\sqrt{Hz}} = \sqrt{\frac{4k_B TQ}{m\omega^3}}.$$  \hfill (47)

Complete discussions of noise processes in NEMS devices along with spectral density functions and predicted effects on the frequency and phase noise response are given in Refs. [44,42,45].

In summary, dissipation is the mechanism through which a resonator loses the mechanical energy stored in an oscillation mode. This energy can be transferred into other modes in nonlinear systems; transferred as elastic energy out of the resonator; converted to non-mechanical energy; or lost to the thermal bath as heat. The measured quality factor may be determined by many different sources of dissipation. In the linear regime, $Q$ is defined as the sum of the inverses of each contributing dissipation mechanism (which can easily be seen from $\Delta W_{\text{tot}} = \sum \Delta W_i$):

$$Q_{\text{tot}}^{-1} = Q_{\text{CL}}^{-1} + Q_{\text{CD}}^{-1} + Q_{\text{MM}}^{-1} + Q_{\text{TED}}^{-1} + Q_{\text{TLS}}^{-1} + Q_{\text{MDP}}^{-1} + \cdots.$$  \hfill (48)

Each mechanism describes an uncorrelated linear force that resists motion and irreversibly converts energy at a rate proportional to the resonant velocity. The following chapters consider various dissipation mechanisms and evaluate their relevance to NEMS resonators. These processes are split into three categories: Extraneous Dissipation (Section 4) includes dissipative mechanisms that are not related to the resonator itself but are typically attributed to engineering constraints. Discussed are circuit damping $Q_{\text{CD}}^{-1}$, clamping losses $Q_{\text{CL}}^{-1}$, and dissipation due to multiple materials $Q_{\text{MM}}^{-1}$. Extraneous sources also include the effect of the medium. This is described by various forms of fluid damping $Q_{\text{TED}}^{-1}$. Intrinsically Sources of Dissipation (Section 5) considers losses within the material itself, namely thermoelastic dissipation $Q_{\text{TED}}^{-1}$, phonon–phonon $Q_{\text{p-p}}^{-1}$, surface losses $Q_{\text{MDP}}^{-1}$, mechanical defects $Q_{\text{MDP}}^{-1}$, and quantum dissipation occurring through two level systems $Q_{\text{TLS}}^{-1}$. A section is also devoted to dissipation in carbon nanotubes as well as graphene due to the high interest in such devices and their unique properties (see Section 6). The following chapters look into each dissipation mechanism and evaluate the relevance to the resonator performance and how this may be determined experimentally.

Finally, methods and ideas on how to minimize and avoid dissipation are presented. The last section introduces the concept of nonlinear damping and discusses its origins as well as current experimental studies on the topic.

3. Nems devices

NEMS devices act as transducers of a mechanical force into an electrical signal. As the structures are extremely small, the mechanical response is very sensitive to external forces. This is what gives the devices the ability to act as detectors and sensors for intriguingly small signals, such as electron spin events [3], molecules [4] or even the quantum background [9,22,19,46]. This sensitivity is a result of downsizing to the nanometer scale. The ability to monitor their mechanical motion with
great precision makes them suitable as oscillators [47–49] for frequency standards, mass sensors [50,51,28,52], in nuclear magnetic resonance detection [53], gyroscopes, accelerometers, electro-mechanical filters [54], high-speed switches [55,56], logic elements [32,57], biotechnology applications [58,15,14,8], and signal processing [1,59]. In carbon nanotube resonators, mechanical modes as high as 39 GHz have been observed [60]. To enable these functionalities not only must the mechanical degree couple to the desired force, but the electrical transduction must be strong in order for changes in the mechanical response to be observed. The following section describes common drive and detection methods, as well as a brief fabrication overview. An up-to-date review of current NEMS resonators, including fabrication, drive and detection, applications, carbon nanotubes and discussion of quantum behavior is given by Greenberg et al. [61]. Poot and van der Zant [46] give a extensive overview. An up-to-date review of current NEMS resonators, including fabrication, drive and detection, applications, carbon nanotubes and discussion of quantum behavior is given by Greenberg et al. [61]. Poot and van der Zant [46] give a extensive overview.

3.1. Resonator actuation and detection

3.1.1. The magnetomotive technique

In the presence of a magnetic field, a nanomechanical device can be both actuated and detected by the coupling of the magnetic field with a conductor attached to the NEMS device. This method is known as the magnetomotive drive and detection technique, and it is often the method of choice for very small structures that oscillate at extremely high frequencies [80]. A significant limitation is the need for a strong magnetic field of 1–16 T, usually produced by a superconducting magnet. The high field also adds to circuit loading (discussed below) and makes the setup large and expensive.

As a significant limitation is the need for a strong magnetic field of 1–16 T, usually produced by a superconducting magnet. The high field also adds to circuit loading (discussed below) and makes the setup large and expensive.

The drive force of a current-carrying conductor in a magnetic field is given by the Lorentz force $\vec{F}_d = I\vec{L} \times \vec{B}$. This force is plugged into Eq. (25), which determines the resonance response. For the detection of the resonance, the magnetomotive technique makes use of Faraday’s law of induction, which describes the voltage produced by a conductor moving through a magnetic field. In Fourier space, the induced voltage is given by

$$\tilde{V}_{\text{emf}}(\omega) = i\xi B L \omega \tilde{x}(\omega).$$  \hspace{1cm} (49)

$\xi = 0.83$ is an integration factor. Substituting expression (25) for $\tilde{x}$ one obtains:

$$\tilde{V}_{\text{emf}}(\omega) = \frac{i\xi \omega L^2 B^2/m}{\omega^2 - \omega_r^2} \tilde{I}(\omega),$$  \hspace{1cm} (50)

where it is assumed that the magnetic field causing the Lorentz force is perpendicular to the long axis of the beam.

3.1.2. The capacitive technique

The force generated on the electrodes of a capacitor is derived from the change in capacitor energy $E_c = \frac{1}{2}CV_b^2$ with respect to distance $F_c = \frac{1}{2}\frac{d}{dx}CV_b^2$. By charging the beam and applying a time-varying drive voltage $V_b \rightarrow V_b + V_{dr}$, the damped driven harmonic oscillator equation becomes:

$$\ddot{x} + \gamma \dot{x} + \left(\omega_0^2 - \frac{C''V_b^2}{2m}\right)x \approx \frac{C'V_bV_{dr}}{m}.$$  \hspace{1cm} (51)

where $\frac{d}{dx} \approx C'_{x0} + C''_{x0}(x - x_0)$, and the static terms are dropped. In addition, it is assumed that $V_{dr} \ll V_b$, hence the $V_{dr}^2$ term can be neglected. The resonance frequency is shifted by the applied d.c. voltage and becomes $\omega_{res} = \sqrt{\frac{k}{m} - \frac{C''V_b^2}{2m}}$. Now the displacement as a function of drive voltage and beam voltage becomes:

$$\tilde{x}(\omega) = \frac{C'V_b\tilde{V}_{dr(\omega)}}{m} \frac{1}{\omega_{res}^2 - \omega^2 - i\omega\omega_{res}/Q_{res}}.$$  \hspace{1cm} (52)

In order to balance the charges on the capacitor, a detectable current is induced at the gates. From this current, the motion of the resonator can be inferred. Using $Q = V_bC$, one derives:

$$\frac{dQ}{dt} = V_b \frac{dC}{dt} = V_b \frac{dC}{dx} \frac{dx}{dt} = V_bC' \omega \tilde{x}.$$  \hspace{1cm} (53)

The final signal detected is typically a voltage resulting from feeding the induced current through a transimpedance amplifier. The signal is linear in both drive voltage and beam displacement (assuming small displacement) and proportional to $V_b^2$. The capacitive (also known as the electrostatic) method is extremely successful with larger MEMS structures, where the larger capacitances result in larger signals. Also, higher frequencies are typically more difficult to detect due to parasitic capacitance and losses in the transimpedance amplifier. It must be recognized, however, that small structures [31] and larger high-frequency structures [81] have been demonstrated.
3.1.3. The piezoelectric technique

In piezoelectric materials, a mechanical strain induces a voltage, and, alternatively, an applied voltage will induce a mechanical strain that is proportional to the electric field. This property can be used to detect and drive mechanical motion up to extremely high frequencies well over 1 GHz [54], these modes are typically bulk modes (as opposed to flexural modes, which are the focus of this paper). The transduction efficiency is very high and this method is extensively used in MEMS. The method is less ideal for small structures and flexural modes. The crystal integrity must be maintained for a material to be piezoelectric, ultimately limiting how thin the device layer can become. Typically, piezoelectric materials are thicker than 1 µm, making entry into the NEMS world quite limited. Nevertheless, piezoelectric NEMS devices have been successfully demonstrated with piezoelectric layers as thin as 100 nm for both cantilevers and doubly-clamped beams only 4 µm long [82–84]. Piezoelectric materials transduce a strain into a voltage (or vice versa) given by

\[ \epsilon_i = \frac{\Delta L_i}{L_i} = \frac{V}{t}, \tag{54} \]

where \( L_i \) is the physical dimension, \( E_i \) is the electric field along the \( z \)-axis, \( t \) is the device thickness, \( V \) is the applied voltage, \( d_{3i} \) are the contour piezoelectric coefficients, and \( i \) is the index representing the \( x, y \) and \( z \) components of the crystal. Here, it is assumed that a field is applied only along the \( z \)-axis, which is the case when a piezoelectric material is sandwiched between two conducting layers. Obviously, thickness modes, where the strain is only in the \( z \) direction, are the simplest structures to calculate. For doubly-clamped beams, electrodes are attached close to the base where the strain field is the strongest. An electrode at one end of the beam is used to actuate the structure, and a second electrode at the opposite end detects. The efficiency of the transduction denoted by \( \kappa \) will depend on the placement of the electrodes. For the extension, one writes:

\[ \Delta L_x = \kappa^2 d_{31} L_x \frac{V_{dr}}{L}. \tag{55} \]

Additional static voltages can be applied to add static strain that tune the frequency [83]. On the other end of the beam, the induced strain will set up an electric field that causes charge to flow. The resulting current is the time derivative of the charge

\[ I_{out} = \frac{d}{dt} \int_A (d_{31} E \epsilon_x) dA, \tag{56} \]

where \( A \) is the area of the electrodes. \( \epsilon_x \) is composed of a flexing component and an extensional strain component that can be neglected for this mode and is given by

\[ \epsilon_x^f = -\frac{t}{2} \frac{d^2 D}{dx^2} \tag{57} \]

where \( D(x) = D_0 \phi(x) \) is the displacement amplitude along \( x \) as defined in Section 2.2. Plugging in the solutions for the doubly-clamped beam, one obtains

\[ I_{out} = i \omega_0 d_{31} E w \frac{t}{2} D_0 \frac{d \phi(0.224 L)}{dx}, \tag{58} \]

where \( x = 0.224 L \) is how far along the beam the electrodes range, corresponding to where \( \frac{d \phi}{dx} \) is maximized. \( D_0 \) is the beam amplitude, a linear function of the input voltage derived in Ref. [85]

\[ D_0 = 0.09 \alpha Q d_{31 E} \frac{L^2 V_{dr}}{\rho f_0^2 L^2}, \tag{59} \]

where \( \alpha \) is an empirical factor describing the electrodes and \( f_0 \) the fundamental resonance frequency. This expression holds if the piezoelectric material dominates and the electrode mass can be neglected. A detailed analysis on piezoelectric actuation and detection specifically for flexural modes is given in Ref. [86], where the mechanical response is given in terms of the mechanical impedance. For the doubly-lamped beam mode, one finds

\[ I_{out}(\omega) = \frac{6.05 i \omega (d_{31 E} t w)^2}{mL^2 \left( \omega_0^2 - \omega^2 + \frac{\omega_0^2 Q_{res}}{\rho f_0^2 L^2} \right)} V_{dr}(\omega). \tag{60} \]

Ref. [86] also demonstrated the beneficial scaling properties of the piezoelectric actuation and detection scheme compared to the capacitive method.

3.1.4. Piezoresistive detection

In a piezoresistive material the resistivity changes with strain due to shifts in the conduction band Fermi levels. Typical piezoresistive materials are doped semiconductors, where an applied strain increases or decreases the gap size. The
piezoresistive gauge factor, $\gamma_P$, characterizes the strength of piezoresistivity and hence, the amount of change in resistance that occurs due to an applied strain is expressed as:

$$\gamma_P = \frac{\Delta R}{R_0} \frac{1}{\varepsilon_s} = (1 + \nu) + \frac{\partial \rho_e}{\rho_e} \frac{1}{\varepsilon_s},$$  \hspace{1cm} (61)

where $\Delta R$ and $R_0$ are the change in resistance and total resistance, respectively, $\varepsilon_s$ is the applied strain, $\nu$ the Poisson’s ratio, and $\rho_e$ the resistivity. Unlike piezoelectric materials, this process is not symmetric or reversible, as “resistivity” cannot be applied, hence this method can only be used for detection. This has been implemented in NEMS devices [87,88]. The resonator itself may be manufactured from a piezoresistive material as is the case for boron doped diamond [89] or, alternatively, a piezoresistive material can be integrated into the resonator. Typically this is done at the clamping points of the resonator or wherever the strain is the greatest. An advantage of the piezoresistive detection method is that it can be implemented into a heterodyne down-mixing circuit where the resonant response frequency is multiplied with a bias frequency to obtain a low frequency detection signal that is not sensitive to parasitic capacitances [87].

This method has been successfully used for mass sensors [51] and demonstrated atomic resolution in an atomic-force microscope system [90,91].

3.1.5. The dielectric technique

Dielectric actuation and detection has been proposed as a universal method for NEMS devices [92,93]. It has been shown that using side gates to polarize resonators made of dielectrics is an efficient way to both drive and detect a NEMS device. Using this method, high quality factor devices have been demonstrated.

The force density acting on the resonator is the product of the polarization and the electric field gradient:

$$\vec{f} = (\vec{P} \cdot \vec{\nabla}) \vec{E}. \hspace{1cm} (62)$$

The detection occurs due to changes in the circuit impedance. Rieger et al. [94] use the dielectric drive and detection schematic coupled inductively to a microwave cavity to tune both the resonance frequency and quality factor.

3.1.6. Other actuation and detection techniques

Other techniques include hybrid setups where, for example, a piezoelectric crystal is used for actuation and the readout is accomplished optically where a laser is reflected off a resonator. Optical methods are especially common for larger MEMS structures but have also been demonstrated for NEMS [95–97]. Optical methods can be very successful in NEMS when coupling mechanically to the optical near field [98]. Such opto-electromechanical methods allowed for low dissipative circuits with strong coupling [99].

Effective actuation techniques that are not discussed here include the piezoresistive detection method [87,88] and electrothermal actuation [100]. More exotic methods used only at low temperatures are variations on the capacitor method, which include single-electron transistors (SET) [12] and superconducting microwave cavity [101]. Recently, SET detection of carbon nanotubes have demonstrates high quality factors [102]. D.C. SQUID detectors have also been successfully implemented to detect the resonant motion through changes in flux [103]. Using field emission, where the current is exponentially dependent on the electrode–resonator gap, the motion of carbon nanotubes has been detected as well [104]. These low-noise methods reduce back action noise as well as circuit damping but are typically more difficult to implement than the methods described above. Dissipation resulting from the readout circuit will not be discussed for these methods.

3.2. Fabrication

Most devices discussed here are manufactured using standard top-down optical or e-beam lithography methods. These methods give a high level of control to design and geometry of the resonator and allow for exact device engineering. Typical steps include a lithography step of the device layer that defines the two-dimensional structure. This is followed by metalization (usually thermal or e-beam evaporation) that serves as an etch mask and, if needed, also as the electrical contact. Anisotropic reactive ion etching (RIE) defines the device after which an isotropic dry or wet-etch removes a sacrificial layer after which the resonator is suspended. These steps are illustrated in Fig. 7. Multiple lithography steps can be used for more complex devices and, if needed, non-metallic layers can be included through sputtering or atomic layer deposition. Post processing measures such as annealing, may significantly affect dissipation (change defects stresses and strain, i.e., phonon propagation). Crystal purity and quality will also affect not only the resonance frequency but to a greater degree, dissipation. Also, growth processes can introduce intrinsic strain that will alter the quality factor. These effects are discussed below.

Now that the devices we are considering have been introduced and described, we can start considering sources of dissipation and methods to calculate and measure them. Dissipation is sorted into two broader categories: Extrinsic, which includes effects that can typically be manipulated through engineering and design, but also includes dissipation mechanisms that are related to the environmental conditions such as ambient pressure, and intrinsic, which includes dissipation mechanisms that are intrinsic to the resonant material. These classifications present an attempt to organize the field but are not a result of exact definitions.
4. Extrinsic sources of dissipation

4.1. Circuit damping $Q_{CD}^{-1}$

For any practical application, a purely mechanical resonator will always be coupled electrically to a drive-and-detection circuit; the most common methods are demonstrated in Section 3.1 (even the optical readout method often relies on a form of electrical actuation). In this section, the effects of the circuit on dissipation, known as circuit loading, is illustrated for the magnetomotive, capacitive, and piezoelectric activation and detection methods.

A damped driven mechanical resonator system can be modeled by an RLC circuit. This results in an equivalent mechanical resistance $R_m$, capacitance $C_m$ and inductance $L_m$. From the frequency response in voltage to drive current, the mechanical equivalents can be deduced. The resonance frequency is given by $f_{RLC} = \frac{1}{2\pi\sqrt{LC}}$. Dissipation in this model depends on whether the RLC circuit is in parallel or in series and takes the form $Q_{CD}^{-1} = \frac{1}{R_mC_m}$ or $Q_s^{-1} = R_m\sqrt{C_m}$, respectively. The full-frequency damped driven parallel RLC circuit is given by

$$V_p(\omega) = \frac{\omega I}{\omega^2 + \omega/(R_mC_m)} I(\omega),$$

and the series circuit takes the form

$$V_s(\omega) = \frac{\omega I}{\omega^2 + \omega R_m/L_m} I(\omega).$$

On resonance, this reduces to $V_p(\omega_0) = R_m I(\omega_0)$ and $V_s(\omega_0) = \frac{Z_m^2}{R_m} I(\omega_0)$, the expected voltage drop over an impedance for a given current, with $Z_m = \sqrt{L_m/C_m}$ being the characteristic impedance of the mechanical system.

4.1.1. Magnetomotive damping

Magnetomotive damping ($Q_{MMD}$) occurs through the interaction of the readout circuit and an external magnetic field. It applies whenever the Lorentz force is used to actuate the resonator. The damping effect of the magnetomotive drive-and-detection scheme was analyzed for the magnetomotive method in Ref. [105]. Using the results from Eq. (50) (doubly-clamped beam) together with Eq. (63), the mechanical equivalents to a parallel RLC circuit can be determined (see Fig. 8 for circuit diagram);

$$R_m = \frac{\xi B^2}{m\omega_0} Q_{res}^{-1}$$

$$L_m = \frac{\xi B^2}{m\omega_0^2}$$

$$C_m = \frac{m}{\xi B^2.$$
Here, $Q_{\text{res}}$ is the quality factor of the resonator without the external circuit. These expressions are correct in the infinite source impedance limit ($Z \to \infty$). $Z$ is the impedance that is seen by the $V_{\text{emf}}$ generated by the moving conductor in the magnetic field. Actual circuits have a finite electrical impedance $Z(\omega) = R + iX$ over which the electromotive voltage is generated. Resulting eddy currents induce an additional Lorentz force that is not due to the drive current [108]. This force always opposes the velocity and adds to the dissipation. Here we assume only a small change in impedance with respect to frequency over the resonance frequency (applies for high $Q$ resonators). This impedance must be added in parallel to the circuit. Now the voltage over the circuit (as is eventually measured by the network analyzer) is given by

$$V_{\text{emf}}(\omega) = \left(\frac{\omega_0^2 - \omega^2 + i\omega_0 Q_{\text{res}}}{i\omega/C_m} + \frac{1}{R + iX}\right)^{-1} I(\omega),$$

where $\omega_0^2 = \omega_0^2 + \omega_0 Z_m X / |Z|^2$ is the observed resonance frequency shifted by the circuit and $Q_{\text{tot}}^{-1} = Q_{\text{res}}^{-1} + \frac{Z_m X}{|Z|^2}$ is the corrected dissipation, i.e., the total dissipation of the system including both the intrinsic dissipation and the dissipation caused by the circuit $Q_{\text{MMD}}^{-1}$. From expression (68) the shifted resonance frequency and measurable dissipation can be extracted [105].

$$Q_{\text{MMD}}^{-1} = \left(\frac{Z_m R}{|Z|^2}\right)$$

$$= \left(\frac{\xi L^2 B^2}{\omega_0 m}\right) \frac{R}{|Z|^2}.$$ (69)

The resonance frequency is also loaded, and to the first order in $Z_m / Z$ becomes

$$f_{\text{circuit}} = f_{\text{res}} \sqrt{1 + \frac{Z_m X}{|Z|^2}}.$$ (70)

There is a $B^2$ dependence to $Q_{\text{MMD}}^{-1}$, and $Q_{\text{tot}}^{-1} \to Q_{\text{res}}^{-1}$ for $B \to 0$. Hence, whenever the magnetomotive method is being used, the unloaded dissipation of the resonator can be determined by measuring the loaded dissipation as a function of magnetic field and then extrapolating to zero field [18,109–111,33,34,112]. Note that this dissipation is not affected by material properties such as Young’s modulus but will vary with frequency, mass, and length of the resonator. For the thin-beam approximation, where $f \propto L^{-2}$, reducing the resonator length will have a great effect as $Q_{\text{res}}^{-1} \propto L^4$. The reduced effect of magnetomotive damping for higher frequencies has been illustrated by Imboden et al. [34]. The addition of a magnetic field in a self excited carbon nanotube resonator can significantly increase the dissipation though eddy currents, resulting in the expected $Q^{-1} \propto B^2$ dependency [113]. For low frequencies (and long resonators), the circuit loading grows; as this effect is well understood, it can be measured precisely and removed analytically to determine the true dissipation occurring only due to the structure [33]. For short beams, the low circuit loading of the magnetomotive method paved the way for the first flexural mode detection of a mechanical resonance above 1 GHz [114].

4.1.2. Capacitive circuit damping

Following the same framework, one can determine the mechanical equivalent of the RLC circuit for the capacitive resonator system. In addition to determining the dissipation contribution from the circuit, this analysis is also useful when impedance-matching high-frequency resonator circuits. For the capacitive drive-and-detection scheme, the equivalent RLC series circuit and the circuit impedance are both in series (see Fig. 8). The total voltage drop over this circuit is again given by $V = R_{\text{tot}} I = (R_m + R_c) I$, where $R_c$ is the circuit resistance. Total dissipation can be written as $Q_{\text{tot}}^{-1} = Q_{\text{res}}^{-1} R_{\text{tot}} / R_m$, resulting in

Fig. 8. Electrical circuit equivalent. (a) Magnetomotive [105], and (b) for capacitive [106] and piezoelectric [107] drive and detection methods. This modified the Butterworth–Van Dyke equivalent circuit includes the parasitic capacitance and impedance load.
in an additional loaded dissipation by $Q_{\text{CCD}}^{-1} = Q_m^{-1}R_c/R_m$. The next paragraph illustrates a more intuitive method to obtain the same result.

As was shown in Section 3.1.2, the moving capacitor will induce a current to balance the flow of charge. This current must be pushed through the circuit impedance and will contribute to losses. Unlike in the magnetomotive case, where a high-circuit impedance was needed to minimize the losses due to the electromotive voltage, here the induced current must flow freely to reduce losses. Again, high-frequency signals require impedance-matching, resulting in losses even for well-engineered circuits. Consider a resonator with energy $W_{\text{res}} = \frac{1}{2}kx^2$ capacitively coupled to a readout circuit. The current induced must be pushed through an impedance, typically the circuit impedance to the first amplifier and the input impedance of the readout electronics. The power dissipated in the circuit is given by $P_{\text{CCD}} = \frac{1}{2}R_c\tilde{I}^2$. Using the definition of dissipation (1) and assuming the energy lost per cycle is only due to the circuit ($\Delta W_{\text{CCD}} = \frac{2\pi P_{\text{CCD}}}{\omega}$) then the capacitive damping becomes

$$Q_{\text{CCD}}^{-1} = \frac{\Delta W_{\text{CCD}}}{2\pi W_0} = \frac{R_c\tilde{I}^2}{\omega_{\text{res}}k\tilde{x}^2},$$

$$= \frac{R_cC^2\tilde{V}_b^2}{m\omega_{\text{res}}} = \frac{R_c}{Q_{\text{res}}R_m},$$

(71)

where $\tilde{I}$ is taken from the derivation in Section 3.1.2. $R_m = \frac{m_{\text{bare}}}{V_b^2C^2Q_{\text{res}}}$ is the motional resistance on resonance of the capacitively coupled doubly-clamped beam resonator as derived by plugging (52) into (53) and using $\tilde{I} = R_m\tilde{V}_b$. Analogous to the magnetomotive circuit a mechanical inductance and capacitance can be defined; $L_m = \frac{m}{V_b^2C^2}$ and $C_m = \frac{V_b^2c^2}{k}$. A detailed discussion of capacitively actuated resonators and their equivalent circuits is given in Refs. [54,115]. It is clear that, to minimize dissipation due to the circuit, the impedance must be minimized. Analogous to sweeping the magnetic field in the magnetomotive method, here the beam voltage can be swept to determine the dissipation as $V_b \to 0$, which corresponds to the true resonator dissipation.

If the capacitive actuation-and-detection method is used in the presence of an external magnetic field, then there will be additional loading. An applied magnetic field will induce a $V_{\text{emf}}$ that opposes the motion by charging the capacitor. This force will always act against the motion and is proportional to the beam velocity $\dot{x}$ and will hence appear as a dissipative term in the oscillator equation. This scenario has been illustrated by Schwab [116] for a SET operating in a magnetic field. Capacitive circuit damping is equivalent to ohmic dissipation that occurs whenever a DC voltage is applied to a resistive resonator. The motion changing the capacitance results in a current flow. The mechanically induced electric energy will be dissipated by ohmic heating. Barois et al. [117] reproduce the expected quadratic dependency of dissipation on the applied voltage and demonstrate how the RC timescale sets a fundamental limit to the quality factor in an analogous way to thermoelastic damping. Ref. [117] also includes a summary of capacitive NEMS resonators with quality factors ranging from 1700 to as high as 150,000 [92]. The tunnel current dominates the dissipative mechanism in CNT resonators coupled to a single electron transistor [102]. As for the classical ohmic and capacitive loss mechanism, the characteristic dissipation proportional to the gate voltage squared is observed. Circuit loading in a microwave cavity can result in a fivefold dissipation increase with again a bias voltage squared dependence [94].

### 4.1.3. Piezoelectric damping

Adequate to the charges flowing to and from the capacitor, electrons are swept to and from the electric pads of the piezoelectric resonators. Again, these currents must be pushed through a finite resistance of the electric circuit and will lead to losses. The total voltage drop of the circuit is given by $V = (R_m + R_c)I$, where for a double beam using Eq. (60) on resonance, one obtains

$$R_m = \frac{\rho L^3\omega_{\text{res}}}{6.03w(d_{31}E)^2} \frac{1}{Q_{\text{res}}},$$

(72)

Using this, together with the results for the capacitive method, the piezoelectric circuit damping $Q_{\text{PCD}}^{-1}$ is given by

$$Q_{\text{PCD}}^{-1} = \frac{R_c}{Q_{\text{res}}R_m} = \frac{6.03w(d_{31}E)^2}{\rho L^3\omega_{\text{res}}}R_c.$$  

(73)

The numerical factors are dependent on where the electrodes are placed; for the doubly-clamped beam, they are optimized when located at maximum strain, which ranges $\approx 0.224L$ along the beam measured from the clamps on either side [86]. Typically, $d_{31}$ is extremely small and hence circuit damping can often be neglected when determining dominant damping sources; additionally, dissipation scales with the inverse of the resonance frequency, hence, piezoelectric dissipation scales favorably with frequency.

A comparison of the three discussed loading mechanisms is depicted in Fig. 9 as a function of length and frequency. The numeric values used are typical for NEMS resonators and show that magnetomotive damping is significantly stronger
Fig. 9. Circuit-loading plots for magnetomotive (blue), capacitive (red), and piezoelectric (green) drive circuits as a function of length and frequency. Width, thickness, and material properties are held constant.

up to very high frequencies. Piezoelectrically loaded circuits increase the dissipation with frequency, but the loading is comparatively low.

4.2. Clamping losses $Q_{CL}^{-1}$

Clamping losses occur when strain is concentrated at the attachment points of the resonator. The time-varying strain radiates elastic energy into the support structure. By definition, flexural modes will always have maximum strain at the nodes located at the clamps and hence are subject to this form of dissipation (see Section 2.2). The vibrating shear force and moment at the clamps act as excitation sources of elastic waves that propagate into the support. (It has been shown that the shear force dominates [118].) The amount of elastic energy lost to the base will depend on the actuated mode and the method of clamping.

Cross and Lifshitz [118] have developed a framework to determine the elastic wave transmission at an abrupt junction. The coupling is weak and transmission can be treated as a perturbation. A two-dimensional model can be used to determine the elastic energy transmitted into the base as well as the heat transport from the constricted beam into the base (the base is assumed to be semi-infinite, and it is assumed no energy returns back into the beam). Defining the clamping loss as the energy transmitted at the nodes into the support structure, one can write

$$Q_{CL}^{-1} \propto \frac{v_g}{L} T_n,$$

where $v_g$ is the group velocity of sound and $T_n$ is the $n$th mode transmission coefficient. For the dissipation in cantilevers, this results in two separate expressions for flexural out-of-plane modes and in-plane modes.

$$Q_{2d-CL \text{ in plane}}^{-1} = \alpha \frac{w^3}{L^3},$$

$$Q_{2d-CL \text{ out of plane}}^{-1} = \beta \frac{w}{L}.$$

The parameters $\alpha$ and $\beta$ are material-dependent only. It can be intuitively understood that the in-plane mode is clamped more rigidly to the base than the out-of-plane mode. This is illustrated in the two-dimensional simulation in Fig. 10 (a) and (b). The in-plane results agree with the predicted dissipation first calculated by Jimbo and Itao [119] where $\alpha = 0.46$. It is often assumed that a doubly-clamped beam suffers twice the loss of a cantilever, as there are two clamping points. Closed form calculations of $\alpha$ [120] show results derived explicitly for cantilevers and doubly-clamped beams. The model uses the standard beam theory presented in Section 2.2 and assumes the thickness of the structure $t$ to be much smaller than the transverse elastic wavelength $\lambda_T = \sqrt{\frac{E}{4\pi(1+\nu)\rho}}$, where $\nu$ is the Poisson’s ratio, justifying the use of two-dimensional elastic wave theory. Hao et al. [120] show for a cantilever ($c-f$)

$$\alpha_{c-f \text{ n}} = \frac{(1 + \nu)\psi}{0.24(1 - \nu)}\zeta_n,$$

where $\psi(\nu)$ is a numerical factor related to the average amplitude and $\zeta_n = \chi_n\beta_n$ is a mode-dependent numerical factor derived from the solution to the beam equations (see Section 2.2). The doubly-clamped beam ($c-c$) results in

$$\alpha_{c-c \text{ n}} = \frac{(3 - \nu)(1 + \nu)\Pi}{1.91(3 - \nu)(1 + \nu) + 2.43\Pi}\zeta_n,$$
**Fig. 10.** Comsol finite element simulation of clamping for a silicon cantilever. $L = 10 \, \mu m$, $w = 0.5 \, \mu m$, $t = 0.3 \, \mu m$, base extends 10 $\mu m$ from the clamping position. (a) 2-D model of out-of-plane mode shows large amount of displacement and stress in the base, dissipation described by Eq. (75). (b) 2-D model for in-plane mode. There is virtually no displacement in the base and the stress is localized very closely to the base. Dissipation described by Eq. (75). (c) and (d) 3-D model with idealized clamping for out-of-plane and in-plane modes, respectively. Minimal stress is simulated; dissipation described by Eq. (82). The frequency in (a) is almost half that of (c), illustrative of the poor clamping in (a). The resonator in (d) shows an even higher frequency than (b), indicative of better clamping. This is predicted as the dissipation for (b) scales as $\frac{w}{L^3}$ and for (d) as $\frac{w}{L}$, it makes intuitive sense that if the undercut in scenario (c) and (d) is $\gg w$, $t$, then scenario (a) and (b) is effectively recovered. This must be avoided during the manufacturing process. (e) and (f) show the strain and frequency of perfectly clamped beams for comparison.

**Fig. 11.** Dissipation plots for room-temperature MHz resonators. Two fits are illustrated $aL^{-n} + b$ with $n = \{3, 5\}$. For both sets, $n = 5$ appears to be the best fit for the scaling law. The constant $b$ portrays the length-independent dissipation attributed mostly to surface losses. The magnetic field dependence of dissipation, including a quadratic fit, is illustrated in the inset.

Source: Reprinted with permission from [33]. © 2007, AIP Publishing LLC.

where again, $\Pi$ is a numerical factor related to the average amplitude where $\Pi(\nu) = \frac{\pi}{2}(3 - \nu)(1 + \nu)$ if $L \gg w$. It can be shown that for the Poisson’s ratio ranging from 0 – 0.5$\alpha$ varies only from 1.3 to 1.7. Experimentally, however, a greater range has been reported; for example, $\alpha = 11.2$ for nanocrystalline diamond (calculated using Ref. [118]) [109], $\alpha = 1.5$ for SiC [114] (the authors do not cite this value; here it is assumed that all dissipation can be explain with clamping losses and $\alpha = \frac{1}{3} \left(\frac{L}{w}\right)^3$ with $Q = 500$, $L = 1.1 \, \mu m$ and $w = 120 \, nm$), $\alpha \approx 1.1$ for 3C-SiC [111], and $\alpha \approx 10$ for polycrystalline diamond [33] (see Fig. 11).

It is illustrated by Hao et al. [120] that, cantilevers suffer from less clamping losses than doubly-clamped beams. Higher-order resonances are also less susceptible than the fundamental mode. This makes intuitive sense, as higher-order resonances have more mechanical energy stored at the nodes far away from the clamps. This effect is well-illustrated in Ref. [121] in a high-order collective mode with strain along the entire beam as well as the clamps. Fig. 12 depicts the Lorentzian response of a high-quality factor collective mode at 1.44 GHz. Typically, first-order GHz flexural modes have very low quality factors of around 500 [114]. To reduce clamping losses, it is possible to increase the aspect ratio and simultaneously maintain the same resonance frequency. This will reduce the overall size of the resonator, making actuation and detection more
difficult, and increase the surface-to-volume ratio, which results in new forms of dissipation (as discussed below). Results published by Gaidarzhy et al. [121] show how an array of coupled oscillators arranged to reduce mechanical impedance can result in high-quality factor transverse mode resonators, where quality factors far exceed limits imposed on standard doubly-clamped beam resonators due to clamping losses.

Depending on the actuation method or the application, in-plane resonances may not always be desired. Out-of-plane actuation results in much higher clamping losses if the two-dimensional model applies. Examples in which this simplified model may be valid would be cases where there is a very large undercut after the release step of the fabrication (see Section 3.2) or possibly graphene resonators, which are inherently two-dimensional. The effects of a large undercut due to imprecise fabrication has resulted in a large reduction of the resonance frequency and unusually high dissipation of silicon beams. Most flexural modes are spared this level of loss due to proper clamping to a thick base. Such cases have been studied explicitly for high-frequency NEMS cantilevers by Photiadis and Judge [122], where the model is expanded to include a base structure. Previously, only monopole radiation from the shear force was considered. More generally, dipole radiation from torque and hybrid radiation may also contribute. If the thickness of the base \( t_b \) is much greater than the thickness of the device, only the monopole radiation contributes. This is most often the case. The limit \( \lambda \gg w, t \) is also taken, which applies safely to all structures considered here. For the cantilever with a finite base \( (t_b \ll \lambda_b) \) the clamping loss is derived in Ref. [122]

\[
Q_{\text{CL base}}^{-1} = \frac{\gamma w t^2 L}{\tau_b^2},
\]

(79)

\[
\gamma = 4 p(\nu, k_b/k) \sqrt{\frac{\rho E(1 - \nu_b^2)}{\rho_b E_b}},
\]

(80)

where \( p(\nu, k_b/k) = \alpha_{22} \left( \frac{k_b}{k} \right)^2 \) is the monopole term and \( \alpha_{22} \) is a function of \( \nu \) of order one [123]. The index \( b \) indicates the base properties. For \( t = t_b \) Eq. (79) reduces to the solution found for the two-dimensional model where the loss is given by (75). The unitless \( \gamma \) factor is weakly dependent on Poisson’s ratio and of order one \((\gamma = 0.95 \text{ for } \nu = 0.3)\). The torque term is proportional to \( \left( \frac{t_b}{\lambda_b} \right)^{5/2} \) and the hybrid term proportional to \( \left( \frac{t_b}{\lambda_b} \right)^3 \) and hence contributes less than the monopole term and can be safely neglected. In the opposite, large base limit \( (t_b \gg \lambda_b) \), the structure is effectively clamped to a semi-infinite medium but only at one side. Again, as the thickness of the base is greater than the thickness of the beam, monopole radiation dominates. Using results from Ref. [124] for the impedance of shear loads at a semi-infinite boundary. Photiadis and Judge [122] determine the dissipation to take the form

\[
Q_{\text{CL hb}}^{-1} \to \infty = \frac{\beta^3_b}{9\pi} \frac{w h^3 E}{L} \frac{k_b^2 k_b (1 + \nu_b) (\tau^3 + 2)}{k_b^2 k_b (1 + \nu_b) (\tau^3 + 2)}.
\]

(81)
As mentioned earlier, imperfect fabrication results in a large undercut, and the quality factor suffers as the system approaches the two-dimensional out-of-plane mode limit. In reality, the base will never be perfect and one may expect a variation on Eq. (82). It will be interesting to measure the same beam with excitation both in-plane and out-of-plane. The surface area will be unchanged, but the clamping losses could differ significantly, indicating which model applies. Huang et al. [126] compared the quality factors of a freestanding beam with a doubly-clamped beam of similar size and of the same frequency (~170 MHz), made of silicon carbide. Compared to the free beam, the quality factor dropped from ~11,000 to ~4500 for the doubly-clamped beam. This is a strong indication that clamping losses are central in determining the performance of such NEMS systems. Considering that typical NEMS resonators have a quality factor of under 10,000, and given Eq. (75), it can be estimated that as long as the aspect ratio \( L/w > 25 \), clamping losses are unlikely to contribute as a major dissipation source, as is the case for long thin beams [66]. Fig. 10 illustrates the cases considered above for clamping loss in a cantilever. The simulation does not include dissipation, but considering the frequency shifts and strain distribution, the above discussion is qualitatively validated.

As mentioned, clamping losses can be minimized by increasing the aspect ratio. For the same operating frequency, if the thickness (or width) is halved, then the length must be reduced by \( \sqrt{2} \), resulting in a factor \( \sqrt{2} \) decrease of the aspect ratio. Alternatively, engineering the resonator anchors to maximize the mechanical impedance mismatch will reduce the clamping losses as well. An implementation of this is to design resonators with multiple beams that resonate out of phase. The antisymmetric motion cancels the forces and moments generated at the base and hence the clamping losses. Examples of this are given by the triple-beam design described by DeVoe [86]. Gaidarzhy et al. [121] solved the problem of high clamping losses due to small cantilevers by positioning an array of cantilevers along a central beam. This removed the strain transfer into the base structure and, through synchrony, amplified the response by forcing the entire structure to resonate at the frequency of the much smaller cantilevers (see Fig. 12). This hierarchical-type structure poses a possible solution to evade clamping losses in high-frequency NEMS structures while maintaining a large size [127], two of the central constraints when scaling down. Typically, harder clamps will reduce losses as discussed above; however, inclusion of flexible anchor points at the nodes has also shown to significantly reduce clamping losses (though this structure behaves like a free–free resonator and is no longer a doubly–clamped beam [126,128]). Increasing the number of nodes will increase the strain energy at positions away from the clamps. Hence, where clamping losses dominate, one would expect higher-order resonant modes to show lower dissipation at the corresponding frequency of a shorter beam. This result is demonstrated by Unterreithmeier et al. [37] in high-strain silicon nitride beams; however, dissipative mechanisms other than clamping losses are believed to dominate. Feng et al. [129] achieved very high quality factors for silicon nanowire resonators at frequencies over 200 MHz; the quality factors scaled with the aspect ratio but not as strongly as \( (w/L)^3 \), which may be explained by imperfect clamping conditions. Finally, Table 2 summarizes the clamping losses.

### 4.3. Dissipation due to multiple materials

Most resonator structures are manufactured from multiple materials. Commonly, there is the device layer and a metallic electrode or mirror, depending on the actuation and detection method, but multiple dielectric layers are not uncommon. Each of these layers is not perfect. They contribute to the total dissipation of the system through clamping losses or any internal dissipation mechanism, and must therefore be considered when making precise measurements. Alternatively, if

<table>
<thead>
<tr>
<th>Model</th>
<th>2D in plane</th>
<th>2D out of plane</th>
<th>Base ( L_b )</th>
<th>( \infty ) base</th>
</tr>
</thead>
<tbody>
<tr>
<td>( Q^{-1} ) ( f )</td>
<td>( \frac{1}{\tau} )</td>
<td>( \frac{1}{\tau} )</td>
<td>( \frac{1}{\tau} )</td>
<td>( \frac{1}{\tau} )</td>
</tr>
<tr>
<td>( c - f )</td>
<td>0.46</td>
<td>\approx 0.5</td>
<td>0.95</td>
<td>0.31</td>
</tr>
<tr>
<td>( c - c )</td>
<td>1.57</td>
<td>\approx 1</td>
<td>3.4</td>
<td>45</td>
</tr>
</tbody>
</table>

where \( \tau^2 = (1 - 2\nu_b)/(2(1 - \nu_b)) \) is the "ratio of dilatational to shear wave numbers in the elastic half-space", \( k_n \) and \( \beta_n \) are mode shape parameters, and \( k_b = \omega \sqrt{2(1 + \nu_b)/E_b} \) is the base wave number. Assuming the base to be made of the same material as the resonator, Eq. (81) reduces to

\[
Q_{cl, hh\rightarrow \infty}^{-1} \approx \frac{\sigma}{L} \left( \frac{w}{L} \right)^4.
\]

where \( \sigma \) is again a function of order one that is weakly dependent on the Poisson’s ratio. For \( \nu = 0.3 \), \( \sigma = 0.31 \). This dependency has been illustrated for noncrystalline diamond beams [33], although the mode was in-plane. These results were generalized in Ref. [125] to include doubly-clamped resonators. It was determined that \( \beta \rightarrow \approx 7\beta \) for the thin support approximation and \( \beta \rightarrow \approx 145\beta \) for the semi-infinite base approximation. These theoretical results are fitted successfully to experimental results for resonators in the 10–100 kHz range [125]. The increase can be, in part, attributed to the higher resonance frequency (factor \( 2\pi \) for the same beam dimensions) and the second attachment clamp.

### Table 2

Summary of clamping losses for various models. The coefficients \( \alpha, \beta, \gamma, \delta \) for cantilevers \((c - f)\) and doubly-clamped beam resonators \((c - c)\) are given for \( \nu = 0.3 \).
the device layer dissipation is known with high precision, dissipation due to attached thin films can be studied in resonator structures, as has been done for ultrathin silicon dioxide films [130] as well as metals [131]. This section does not identify actual dissipation sources but illustrates the effect of dissipation due to multiple materials.

When adding a metal (or any other material) to a resonator, the mass loading will alter the resonance frequency [132]. (For this discussion, it is assumed that the device layer is the thickest and hardest material, but the expressions hold for other scenarios as well.)

\[
f_i = \frac{\lambda_i}{L^2} \sqrt{\frac{E_{ib} t_b + E_{im} t_m}{\rho_{ib} A_b + \rho_{im} A_m}}.
\]  
(83)

(Variables are defined in Section 2.2.) In addition, the second material will also contribute to the total dissipation of the composite structure. Starting with the definition of dissipation given in Eq. (1) and following Ref. [111], we assume that not only each dissipation mechanism but also each individual layer contributes linearly to loss. Hence, for a given resonator mode, the total dissipation is given by

\[
Q^{-1}_{tot} = \sum_{j,k,... \neq i} \left( 1 + \frac{t_j E_j + t_k E_k + \cdots}{t_i E_i} \right)^{-1} Q_i^{-1}.
\]  
(84)

The indices sum over all materials that make up the resonator. Typically, no more than two materials are present; the device layer and the metallic layer that serves as the electrode (additional adhesion layers are extremely thin, ranging from 3–10 nm, and do not contribute in any meaningful way). In this case, Eq. (84) reduces to [133]

\[
Q^{-1}_{tot} = \frac{1}{1 + \beta (Q_b^{-1} + \beta Q_m^{-1})},
\]  
(85)

where the \(b\) and \(m\) indices refer to the beam and metal that compose the structure, respectively, \(\beta = \frac{t_m}{t_b}\) is a measure of the contribution of the metallic layer. Often, \(\beta \ll 1\) and the metallic film contribution to dissipation can be neglected without significant loss in accuracy, as both \(E_m \ll E_b\) and \(t_m \ll t_b\) for most resonators. For a gold–silicon resonator with a device layer thickness of 500 nm and a 50 nm thick electrode \(\beta = 0.075\). For harder materials, such as silicon-carbide, titanium-carbide or diamond \(\beta\) is considerably smaller. The mass loading effect on dissipation contrasts the effect of the metallic layer on the resonance frequency, which is generally very easily measured and must be taken into account when determining the Young’s modulus using Eq. (83). Examples of this can be found in Refs. [132,109,111,34].

Ideally, the contribution of each layer should be measured individually for a given resonator. However, this is often not possible nor necessary, and typical internal dissipation values for common materials described in the literature (Al, Ti, Au, Mo, In, Ag, Cu, a-SiO₂) [131] can be used. Such dissipation measurements of thin films have been conducted on large torsion resonators. The results for internal dissipation due to the metallic layers can be used when calculating the contribution of the electrodes. Assuming a thin film, the frequency shift to the first order is given in Ref. [134]:

\[
\frac{f_{tot} - f_b}{f_{tot}} = 3 G f t_f \left( \frac{\rho_b t_f}{G B_b} - \frac{\rho_b t_f}{2 G B_b} \right),
\]  
(86)

where \(G\) is the shear modulus of the bare structure and the thin film and \(t_i\) and \(\rho_i\) the corresponding thicknesses and mass densities. The corresponding dissipation shift due to the thin film is given by

\[
Q^{-1}_{tot} = Q^{-1}_{bare} + \frac{3 G_f t_f}{G_{bare} t_{bare}} Q^{-1}_{film}.
\]  
(87)

Using the above expressions, larger torsion resonators have been used to measure the internal friction of thin films, both metallic and non-metallic, to high accuracy over a large temperature range [135,130,136,131].

Of fundamental interest are the actual dissipation sources themselves, leading to the \(Q_{beam}^{-1}\) and \(Q_{film}^{-1}\) terms that appear in Eq. (84). They are discussed in detail below.

4.4. Resonators in a medium

It is evident that by removing a surrounding material, be it a highly viscous liquid or gas, or even a low-pressure gas, the quality factor will improve. This can be seen by modifying the elasticity Eq. (8), where the total force acting on the resonator is no longer just the drive force but also the damping force, which is proportional to the velocity. Hence, the damping force will add a term of the form \(F_{medium} \propto -\dot{x}\) due to hydrodynamic loading. Experimentally, it may often be possible to measure the resonator in high vacuum, where damping from the surrounding medium can be safely neglected. This is imperative when measuring intrinsic dissipation effects or trying to maximize \(Q\). In applications, however, a surrounding medium is often unavoidable, either for engineering constraints or more fundamentally when the resonator is used to probe the environment. Examples include gas sensing [137], chemisorption mass detection in real time [88], bimolecular recognition [138,58], and potential computation applications [139]. High quality factor NEMS resonators in air have been demonstrated [140].
(reaching quality factors of over 1000 for 5 MHz structures using a parametric drive). Resonators operating in isopropyl alcohol, acetone, water, and phosphate-buffered saline have been reported with quality factors for 20–100 MHz cantilevers and doubly-clamped resonators of 3–10 MHz [96].

The damping caused by a surrounding medium can be analyzed in three regimes described by Newell [16]. At low pressures, known as the molecular regime, the dissipation is dominated by individual molecule collisions. At high pressures or even liquids, the continuum limit applies as described by Stoke’s law for damping. Lastly, depending on the geometry of the resonator, squeeze-film damping occurs when the medium is trapped between the resonator and base, effectively increasing its viscosity.

### 4.4.1. Low pressure dissipation $Q_{LPD}^{-1}$

The low pressure regime holds when the mean free path of the gas molecules is greater than the resonator dimension $l_{mfp} \ll w$. In this case, the gas molecules do not interact with each other and the dissipation is a result of momentum transfer between individual molecules and the resonating beam [16].

$$Q_{LPD}^{-1} = \left( \frac{2}{\pi} \right)^{\frac{3}{2}} \frac{1}{\rho \nu_0} \frac{P}{v} \approx 0.002 \frac{P}{\sqrt{E \rho}} \left( \frac{L}{t} \right)^{2}. \quad (88)$$

The final expression is true for a doubly-clamped beam resonator, $P$ is the pressure and $v = \sqrt{k_B T / m}$ is the gas molecule velocity, and $Lw = A$ is the area of the oscillator moving through the medium. Liu et al. [139] demonstrated room-temperature time-resolved GHz frequency cantilever structures operating at atmospheric pressure, and have shown that clamping losses and low-pressure dissipation accounts for a quality factor as low as $\approx 18$.

### 4.4.2. Viscous dissipation $Q_{visc}^{-1}$

As the pressure increases, the medium is described by its viscosity rather than by the dynamics of the individual molecules. Unlike in the molecular regime, in the continuum limit the movement of the resonator drags some of the medium with it, increasing the effective mass (see Fig. 14). This results in a strong frequency shift not observed in the molecular limit [16]:

$$Q_{visc}^{-1} \approx \frac{3.8 \mu}{\sqrt{E \rho w}} \left( \frac{L}{t} \right)^{2}, \quad (89)$$

and applies again for the double beam resonator and $\mu = 18.27 \times 10^{-6}$ kg/(ms) is the medium viscosity. The viscosity itself is not dependent on pressure; hence as a function of pressure one expects a transition as described in Refs. [18,141].

The crossover pressure can be determined by equating (88) and (89), which results in $P \approx 0.034/w$; at atmospheric pressure, the molecular regime applies if the resonator beam width $w < 350$ nm. NEMS devices are often in this size scale and therefore fall between the two regimes. In a more viscous fluid, the continuum model applies. But, for lighter gases and lower pressures, the molecular regime defines the dissipative characteristics of NEMS structures. This crossover regime can also be characterized by the Knudsen number $K_n = l_{mfp}/w \approx 1$ [142]. Equivalently, the relaxation time $\tau_g$ of the medium will either be greater or smaller than the resonance frequency; in this case, the crossover regime is described when the Weissenberg number $W_l = \omega \tau_g \approx 1$ [143]. Dissipation measurements from low to high pressures have been performed in Refs. [18,88] up to 126 MHz. The crossover from the molecular to the viscous regime is well-documented. At high pressures, the dissipation is predicted not be pressure-independent as described by Eq. (89). High-frequency structures may deviate from the quasi-steady Stoke’s solution and are described in detail by Bhiladvala and Wang [142]. Experimentally, the dissipation scales as $Q_{visc}^{-1} \propto \sqrt{P}$ in the continuum limit as observed in Refs. [18,45,88,144]. Although the viscosity is pressure independent, the dissipation is given by the losses $\gamma$ divided by the resonance frequency $\omega$. At high pressures, the added mass that is dragged along by the resonator decreases the resonance frequency and hence increases the observed dissipation. In addition, the relaxation time $\tau = \mu \gamma / P$ is inversely proportional to the pressure described by the Yakhot number $\mu \gamma$, significant for the crossover regime [143]. The two pressure regimes and the crossover regime are analyzed by Kokubun et al. [141], albeit for larger quartz structures, and explained by the Boltzmann BGK equation given [143]. For wide structures, shear wave effects occur and further increase viscous damping [145]. Svitelskiy et al. [144] measured the dynamical response of $\approx 10$–250 MHz structures over 10 decades of pressure in He, N$_2$, CO$_2$, and liquid CO$_2$.

### 4.4.3. Squeeze-film damping $Q_{SFD}^{-1}$

If the gap size between the resonator and a fixed structure such as the ground becomes small, the resonator will have to squeeze the medium out of the gap in order to oscillate. This damping can only occur in the continuum limit and is known as squeeze-film damping, as described by Rebeiz [146]:

$$Q_{SFD}^{-1} = \frac{\mu}{4 \sqrt{E \rho w}} \left( \frac{L}{t} \right)^{2} \left( \frac{w}{g} \right)^{3}, \quad (90)$$
where $g$ is the gap between the beam and the substrate. (90) is derived from a model based on the drop in pressure of a fluid flowing between parallel walls. The form is strongly geometry dependent and, for wider structures, can be reduced by introducing holes in the resonator as described in Ref. [146]. Squeeze-film dissipation is important for wide structures, as $w > g$. Most commonly, this must be taken into account in capacitive-switch MEMS; similarly, for capacitively driven NEMS structures, the gap is engineered as narrow as possible so that $w \ll g$. These structures will suffer most in high-pressure environments. Applications of this effect for gas sensors is described for lower frequency structures by Andrewset al. [137], and an overview for squeeze-film damping in MEMS, including various geometries, is given in Ref. [147]. The effects of low pressure and viscous and squeeze-film damping are depicted in Fig. 13 for a typical silicon doubly-clamped beam.

An interesting result of air damping is that reducing the resonator width reduces dissipation. Generally, the smaller the structure is, the lower the quality factor becomes (as discussed earlier). In the presence of a medium, however, this may be reversed, as the smaller structures have less drag [142,88] (see Fig. 14). Yang et al. [148] report very high losses and a large reduction in the quality factor with increased pressure for longer and wider beams compared to what has been described here, emphasizing the importance of scale. Ultra-high-quality-factor resonators will be sensitive to molecular dissipation down to much lower pressures ($P_{\text{min}} \approx 10^{-3}$ Torr in air) as demonstrated in Ref. [25]; this is expected, as the pressure dependency holds until other dissipation mechanisms, such as intrinsic or clamping losses, dominate. Verbridge et al. [149] further demonstrate how to optimize the quality factor in NEMS devices at atmospheric pressures by balancing the benefits of reduced size to reduce air damping and squeeze-film damping as well as minimizing the surface-to-volume ratio to reduce losses due to surface defects (discussed in Section 5).

5. Intrinsic sources of dissipation

As illustrated in Section 2.2, losses can be included by making the Young’s modulus complex. The loss in energy per oscillation is given by the complex strain field, analogous to Eq. (2). $\Delta W$ becomes

$$\Delta W = \pi \int_V E \epsilon_{\text{max}}(\vec{x}) dV,$$

where the difference in Eq. (2) is that the integration is over one period and not time averaged. Combining Eqs. (2), (1), and (91), one finds a very general expression for the quality factor

$$Q^{-1} = \frac{E_I}{E}.$$

$E_I$, being the complex component of the Young’s modulus, will depend on material properties, including defects as well as resonant modes, and can even have explicit dependence on geometry, for example, when used to describe the thermoelastic dissipation.

5.1. Thermoelastic dissipation $Q_{\text{TED}}^{-1}$

Thermoelastic damping is a loss mechanism due to phonon–phonon interactions resulting from the scattering of acoustic phonons with thermal phonons. This effect is strongest in strain fields resulting from flexural modes, weak in compression...
modes, and nonexistent in torsional modes (although Houston et al. [150] point out that all MEMS-resonant modes will have a flexural component, and hence torsional modes and even compression modes can experience thermoelastic losses). Most dissipation mechanisms presented here have an implicit temperature dependence, as the material properties are generally temperature dependent and often strongly temperature dependent at low temperatures (the thermal expansion coefficient, for example, can even change sign, as measured in gallium arsenide [151]; see Ref. [152] for further details on this topic). Acoustic phonons are coupled to the diffusive thermal phonons through the macroscopic thermal expansion coefficient $\alpha = \frac{1}{L} \frac{\partial L}{\partial T}$. As will be seen below, thermoelastic dissipation is also explicitly dependent on temperature and vanishes as $T \to 0$.

Here we follow Zener’s derivation of dissipation due to thermoelastic damping in mechanical beam resonators [153, 154]. He showed that under specific conditions, internal friction due to thermodynamical considerations alone can dominate over all other dissipation sources. This occurs when the thermodynamic equilibrium cannot be established. Similar to a gas, a compressed and stretched solid will heat and cool (assuming a positive expansion coefficient). Usually, the time required for thermalization is greater than the period of the resonance. Transverse excitations create regions of dilation and compression in close proximity and hence disrupt the thermal equilibrium.

As a result of the heating and the finite thermalization time for real materials, Hooke’s law must be extended to include the relaxation time of the strain and stress fields, as has been shown in Section 2.3:

$$\sigma + \tau_c \epsilon = M_R (\epsilon + \tau_s \omega),$$

(93)

where $\sigma$ is the stress, $\epsilon$ the strain, $\tau_c$ the respective relaxation times, and $M_R$ the relaxed modulus. This lag results in an out-of-phase strain and stress field resulting in temperature gradients. As a result of irreversible heat flow, mechanical energy is extracted out of the actuated mode and dissipated into the phonon bath. Thermoelastic damping is particularly strong in a flexural mode because the stress and strain fields are both large and spatially close (separated only by the thickness $t$ of the beam). It is assumed that the heat flow is diffusive, hence the theory only holds if the structure size is greater than the mean free path of the thermal phonons ($l_{ph}$). Considering that only the stress perpendicular to the excitation, the increase in energy per volume $E_v$ is given by

$$dE_v = T dS + \sigma d\epsilon.$$  

(94)

Zener [153] derived the thermoelastic dissipation by calculating the time averages of the stress, strain, and temperature; assuming heat flow through thermal diffusion, no heat flow perpendicular to the surfaces, and fixed boundary conditions,
Zener obtained

\[
Q_{th}^{-1} = \frac{\alpha^2 T E}{\rho C_p} \frac{\omega \tau_{th}}{1 + (\omega \tau_{th})^2},
\]

(95)

where \( \tau_{th} = \frac{\tau T}{\rho C_p} \) is the thermal relaxation time. \( C_p \) is the heat capacity per unit volume at constant pressure, \( \alpha \) the thermal expansion coefficient, and \( \kappa \) the thermal conductivity. (Alternatively, the thermal diffusion coefficient \( D = \frac{\alpha}{\rho \kappa} \) can be used.)

Eq. (95) is of the generic form for dissipation found in Section 2.3.

It is apparent that the relationship between the resonator frequency and thermal relaxation time will determine the strength of thermoelastic dissipation. At very low frequencies \( \omega \ll \tau_{th}^{-1} \), the heat flow is effectively instantaneous; as no temperature gradient can be established, this is the isothermal regime and the damping is low. At very high frequencies \( \omega \gg \tau_{th}^{-1} \), the process is adiabatic; the heat gradient changes so quickly that even though the temperature differences may be large, they oscillate too fast for a heat transfer to occur. The thermoelastic dissipation is maximized for \( \omega = \tau_{th}^{-1} \), which for case the phase difference between the strain and the stress is maximized; this is known as the Debye resonance.

Zener’s calculation assumes that all thermoelastic energy dissipation occurs to do a single relaxation time given by the first model. Lifshitz et al. [155] showed that for a rectangular beam, there is an analytical solution for small excitation amplitudes that includes all relaxation-mode timescales. This correction is less than 2%, but their analysis is important, as they consider thermoelastic damping in NEMS structures, dimensions orders of magnitude smaller than that described by Zener. The equation of motion for the beam described in Section 2.2 must be modified to include thermoelastic strain:

\[
\rho A \frac{d^2 Y}{dt^2} + \frac{d^2}{dx^2} \left( EI \frac{d^2 Y}{dx^2} + E \alpha T \right) = 0
\]

(96)

where \( I_T = \int_2^2 \gamma \varphi \sqrt{\alpha} \) is the thermal contribution to the moment of inertia as a result from the modified Hooke’s law (see Eq. (93)) and \( \theta \) is the thermal field. The thermoelastic coupling must be added, which relates \( \theta \) to the strain of the equation of motion as described by Ref. [156]. The approximations \( \theta \ll T \) and the assumption that thermal gradients along the axis of deflection dominate further simplify the problem. The dissipation is included through a complex frequency or Young’s modulus given by

\[
Q^{-1} = \left| \frac{\tilde{\alpha}(\omega)}{\tilde{\alpha}(\omega)} \right| = \left| \frac{\tilde{\alpha}(E)}{\tilde{\alpha}(E)} \right|
\]

(97)

where the imaginary part of the Young’s modulus relates the adiabatic Young’s modulus to the isothermal Young’s modulus [148].

Lifshitz and Roukes [155] found a resulting equation of motion that only deviates from the form illustrated in 2.2 by a complex frequency-dependent Young’s modulus \( E_{\text{iso}} \). At very low frequencies \( E_{\text{iso}} \rightarrow E \), the isothermal case is retrieved, and at very high frequencies \( E_{\text{iso}} \rightarrow E_{\text{adi}} \), the adiabatic Young’s modulus is obtained. For both extremes, the Young’s modulus is real and hence there is no thermoelastic dissipation. In between, the complex Young’s modulus defines the dissipation:

\[
Q_{th}^{-1} = \frac{\alpha^2 T E}{\rho C_p} \left( \frac{6}{\eta^2} - \frac{6}{\eta^3} \sinh \eta + \sin \eta \right)
\]

\[
\eta = \sqrt{\frac{\omega_0}{2D}}
\]

This equation holds for a beam that fulfills the Euler–Bernoulli conditions of large aspect ratio \( L/t \) and small deflections, i.e., in the linear regime. It is also assumed that the phonons are diffusive, which holds if the system size is larger than the mean free path of the thermal phonons, resulting in the \( t^2/D \) dependence. For smaller structures, phonon transport becomes ballistic and thermoelastic dissipation should scale as \( t/v \), where \( v \) is the phonon velocity. A further analysis of deflection amplitude is given in Ref. [157] following the same framework. Thermoelastic damping has been verified in a large range of structures over a very wide range of frequencies well over 1 MHz [158]. Houston et al. [159] consider the scaling properties of thermoelastic damping. In agreement with Ref. [155], it is shown that \( \omega T_{th} \propto l \), where \( l \) is the length scale, and hence thermoelastic damping will reduce with the system size. Nevertheless, depending on specific aspect ratios, it is possible that thermoelastic damping contributes significantly down to the nano-scale. (It is predicted in Ref. [150] that thermoelastic dissipation may be detected for specific structures with dimensions as small as 50 nm.) To date, the smallest structures in which thermoelastic damping has been confirmed as a dominant source of dissipation are in the thickness range on the order of 1–2 \( \mu \)m [160,161,152] and even at thicknesses as small as 500 nm for short beam lengths of 6 \( \mu \)m [148]. The inability to measure thermoelastic dissipation down to even smaller sizes comes from the increased dissipation due to surface effects as the surface-to-volume ratio increases, as is discussed below. Improvements in resonator design and fabrication, including annealing and high tensile strain resonators may reduce the significance of surface effects and allow thermoelastic dissipation to dominate in nanoscale structures (see Fig. 15).
To generalize thermoelastic damping to include other mode shapes, Zener’s equation (95) must be modified by including a prefactor that accounts for the size of the flexural strain field. This is described by [162] for large plate modes; the approach, however, applies to a large range of geometries and mode shapes. Much more complicated structures and modes can be calculated numerically using finite element simulations and the heat transfer equation

\[ D \nabla^2 T = \frac{\partial T}{\partial t} + \frac{E \alpha T_0}{(1 - 2\nu)C_v} \frac{\partial \epsilon}{\partial t}, \]

where \( \nu \) is Poisson’s ratio, which couples heat with the strain field, as has for example been done in Ref. [163] for ring structures; Ref. [164] for MEMS mirrors; Ref. [165] for complex beam resonators; and Ref. [166] over a large thickness range, with good agreement with the analytical solutions. The preceding analysis assumed no surface stress or strain as a boundary condition. Especially as the surface-to-volume ratio increases, this approximation may not be valid. This situation is considered by Ru [167] for nanowires. It is shown that for isothermal surface conditions surface stress becomes significant when calculating thermoelastic dissipation, while under specific conditions, surface stress can reduce thermoelastic dissipation by 50%. Such thermoelastic limited quality factors in high strain resonators have been demonstrated by Verbridge et al. [23,25]. A more general discussion of the relation of strain and dissipation is given in Section 7.1.

Fig. 16 illustrates the main results for thermoelastic dissipation comparing different materials with theory and experiments. Table 3 gives numerical values for thermoelastic dissipation for silicon and diamond for various beam sizes. This framework can be used to determine the maximum sensitivity of a device, as has been done for force detection in Ref. [152]. Duwel et al. [168] describe how, in the case of a MEMS gyroscope, thermoelastic dissipation can set the sensitivity limit. Through careful engineering of the geometry and choice of materials, the resonance frequency can be moved away from
Table 3

Typical numbers for thermoelastic dissipation in silicon and diamond resonators.

<table>
<thead>
<tr>
<th>L, t, w (µm)</th>
<th>$f_0$ (MHz)</th>
<th>$f_T$ (MHz)</th>
<th>$Q_{TED}$ 300 K</th>
</tr>
</thead>
<tbody>
<tr>
<td>Si 100, 1, 1</td>
<td>0.74</td>
<td>273 $\times 10^3$</td>
<td>10$^{15}$</td>
</tr>
<tr>
<td>Si 10, 0.5, 0.5</td>
<td>37.0</td>
<td>1.1 $\times 10^6$</td>
<td>10$^{17}$</td>
</tr>
<tr>
<td>Si 2, 0.2, 0.2</td>
<td>396.6</td>
<td>6.8 $\times 10^6$</td>
<td>10$^{19}$</td>
</tr>
<tr>
<td>NCD 100, 1, 1</td>
<td>1.78</td>
<td>628 $\times 10^3$</td>
<td>10$^{15}$</td>
</tr>
<tr>
<td>NCD 10, 0.5, 0.5</td>
<td>89.2</td>
<td>2.5 $\times 10^6$</td>
<td>10$^{18}$</td>
</tr>
<tr>
<td>NCD 2, 0.2, 0.2</td>
<td>892</td>
<td>16 $\times 10^6$</td>
<td>10$^{19}$</td>
</tr>
</tbody>
</table>

Fig. 16. Zener model of thermoelastic dissipation. (a) $Q_{TED}^{-1}$ as a function of beam thickness for three different length scales, (b) $Q_{TED}^{-1}$ as a function of beam lengths for three different thicknesses, (c) beam thickness that maximizes thermal dissipation at 300 K as a function of length, (d) $Q_{TED}^{-1}$ for silicon at four different temperatures and constant thickness.

the Debye peak to minimize the intrinsic dissipation. A more detailed background, including results for larger structures, various clamping scenarios, and a comparison with air damping, is given in Ref. [169].

So far only rectangular cross-sections have been considered. Tunvir, Ru and Mioduchowski [170] present a theoretical discussion of thermoelastic dissipation. It is predicted that in the nano to micron scale, there is a crossover where rectangular structures suffer from thermoelastic dissipation at high frequencies and should hence be utilized at low frequency modes. Elliptical and triangular cross sections tend to show the reverse where high frequencies are less prone to thermoelastic dissipation. It is also argued that isothermal surfaces result in lower dissipation than adiabatic ones, where isothermal surfaces are of constant temperature and adiabatic surfaces are defined by a normal temperature gradient that vanishes at the surface.

5.2. Surface losses $Q_{SL}^{-1}$

As the resonators become thinner (or narrower), the surface-to-volume ratio grows ($S/V \approx 2(lw/L) \approx 2(\frac{1}{w} + \frac{1}{t})$), where the doubly-clamped beam has no ends and the cantilever’s end can be neglected. Eventually, the surface properties start to play a significant role in the dissipation, and the bulk properties no longer dominate. The surface contains a large amount of defects, due to the abrupt lattice termination as well as surface impurities. Free dangling bonds, absorbates, and crystal termination defects contribute significantly to damping in thin-beam resonators through forming additional energy reservoirs and/or mediating anharmonic mode coupling [111].

Following Refs. [153,161], it is assumed that the surface Young’s modulus $E_S \rightarrow E_S + i\epsilon E_S$ differs from that of the bulk, precisely due to the defects, and can be used to characterize the material up to a skin depth $\delta$. The total energy calculated in Eq. (2) will not be affected as long as $t, w \gg \delta$; however, the dissipation and hence $\Delta W$ may be. Eq. (32) becomes

$$\Delta W_S = \pi \int_V E_S^2 \epsilon_{\max}^2(x) dV_S,$$

(101)

where now the integral runs only over the surface layer $V_S$. Plugging in $\epsilon$ for the cantilever, one finds
\[ \Delta W_S = \pi \delta E_I^2 \left( \frac{w}{2} \right) \int_0^L \epsilon_m^2(x) dx. \] (102)

\( \delta \) is the thickness of the surface layer where \( E_I^2 \) is defined, analogous to a skin depth. The resulting dissipation takes the form

\[ Q_{sd}^{-1} = \frac{2\delta(3w + t)}{\omega t} \frac{E_I^5}{E}. \] (103)

where the results for the doubly-clamped beam are identical, as this structure only differs from the cantilever in \( \epsilon(x) \), which cancels out. As can easily be seen for very thin beams (\( t \ll w \)), the dissipation becomes proportional to \( \frac{\delta}{t} \). The parameters \( \delta \) and \( E_I^2 \) are theoretically difficult to predict, typically making \( \delta E_I^2 \) an experimental parameter. The surface effects can be observed and quantified experimentally by changing the resonator geometry, such as \( t \) or \( w \) [161,171]. Alternatively, surface treatments such as high-temperature annealing, which removes adsorbates and releases strain as described in Refs. [172, 171,161,173], or chemical treatments [171,173,161,174] have proven to have significant effects on the quality factor, and improvements of over one order of magnitude have been reported. Silicon, for example, has two dangling bonds that will react with oxygen when exposed to the atmosphere; treatment with atomic hydrogen can reduce oxidation and significantly lower dissipation [148]. Both chemical and heat treatment for silicon cantilevers of various thicknesses and lengths are discussed by Yang et al. [171], including a discussion on the type of dangling bonds that affect surface losses.

Typically, the thickness is given by the initial thin film wafer from which the resonator is manufactured. For very thin long beams, surface dissipation is maximized and dominates as the length is reduced until clamping losses increase sharply. This crossover has been observed at lengths of \( \approx 30 \, \mu m \) for 170 nm thick and 4–6 \( \mu m \) wide cantilevers [171]. The significance of the surface over the bulk material is discussed by Wang et al. [174], who note that for a 250 nm thick silicon resonator, the surface monolayer makes up only \( \approx 0.07\% \) of the mass but can effect the quality factor by 60\%, depending on surface treatment. Studying hydrogen and alkyl-monolayers, they conclude that the mechanical properties do not determine the surface dissipation contribution and suggest that the surface defect density determines surface losses. Further studies confirmed the hypothesis that surface loss is dominated by the coupling of the strain field to electrically active surface defects that occur at the interface of the bulk and surface of the resonator [175]. It is estimated that a very low number of these defects, on the order of 10\(^{4}\), may account for a large fraction of the mechanical dissipation. For chemical and biological sensing applications, resonator surfaces are functionalized to become reactive to a given molecule. Removing oxide termination improves the quality factor; it has been known for some time that passivating with thermal oxide through heat treatment also significantly reduces surface loss. This illustrates that the chemical oxidation adds defects, but careful heat treatment will reduce defects [176]. Hence, it is intuitively clear that the chemical etch processes occurring during the fabrication process significantly affect the final dissipation measured; Ref. [51] discusses the effect of Ar\(^+\) damage in a doubly-clamped piezoresistive mass sensor. The functionalization itself (and possibly also the subsequent surface reactions) will significantly alter the resonator losses. The exposure to air and resulting chemical surface reactions have been studied for silicon in detail [177].

The surface area will increase with surface roughness adding to the dissipation. In single crystal silicon or gallium arsenide, surface roughness is generally not an issue. However, NEMS are now readily manufactured from exotic materials, such as nanocrystalline diamond or silicon carbide. The growth processes for these materials have yet to be optimized, and considerable surface roughness (2–10 nm for SiC and even higher for NCD) is not uncommon. The greater surface area leads to increased dissipation, as is indicated in Ref. [178] for SiC large aspect-ratio resonators. It is also estimated that the length-independent dissipation of doubly-clamped NCD resonators can be explained by surface losses [33]. The lack of control, common in CVD growth for crystalline diamond, results in a large variation of surface defects from sample to sample. Single crystal diamond drum resonators have been reported [179]. It is believed that surface losses dominate the dissipation mechanisms at temperatures above \( \approx 60\, K \). Their contribution falls off at lower temperatures and eventually clamping losses are believed to dominate. Surface losses at high temperatures and a mechanical defect dissipation peak at \( \approx 50\, K \) has also been observed in SCD MEMS scale resonators [180], in agreement with the Debye temperature for diamond of nano-crystalline diamond found by Hutchinson et al. [109]. A recent study of high tensile-strain silicon-nitride resonators [37] indicates that a broad spectrum of defects may explain the dominant dissipative source. Although changes in the manufacturing process indicate that surface defects may be the cause, it is difficult to distinguish between surface and bulk defects.

Measurements of nanowires (made from single-crystal silicon [181] and platinum [182]), which have extremely large surface-to-volume ratios, illustrate that surface losses dominate as expected and increase with the surface-to-volume ratio, resulting in quality factors much lower than predicted, if only bulk properties and clamping losses were to be considered. This illustrates how the attempt to evade clamping losses by increasing the resonator’s aspect ratio (typically extremely large for nanowires) inevitably increases the surface-to-volume ratio as well, and thus imposes a new limit on the quality factor. Ultimately, CNT and graphene structures have the largest surface-to-volume ratio; however, their two-dimensional structures justify further consideration. Though surface losses can dominate it is still possible to obtain high quality factors in silicon nanowires at low temperatures [53].

### 5.3. Dissipation due to mechanical defects \( Q_{sd}^{-1} \)

Whereas a perfect crystal loses mechanical energy through mechanisms such as thermoelastic damping, most real structures have measured quality factors much below this ideal limit. As discussed in the previous section, this can be
explained by crystal defects. As described by Greenberg et al. [66], many defects, such as substitutional impurities, interstitial motion, and grain-boundary sliding, just to name a few, contribute to dissipation, as the defects reconfigure between equilibrium and metastable states in the dynamic strain field. The activation energy and relaxation time of these defects characterize the on-resonance dissipation in the same form as found for thermoelastic losses. In this model, an impurity has two metastable states from which it can switch back and forth. The switching barrier is associated with a specific energy, which appears as a peak in the dissipation at a characteristic temperature (thermal-activation peak or Debye peak). The mechanical-defect dissipation can be written as

\[ Q_{MD}^{-1} = \sigma N \frac{\omega \tau_{MD} \exp \left( \frac{E_A}{k_B T} \right)}{1 + \left( \omega \tau_{MD} \exp \left( \frac{E_A}{k_B T} \right) \right)^2} \]

where \( \sigma \) is a unitless constant related to the defect density and \( \tau_{MD} \) is the defect relaxation time. Assuming atomic motion of the defects, the relaxation time will follow the Arrhenius relation; this relationship has been demonstrated in NCD resonators by Hutchinson et al. [109] (see Fig. 18). \( \tau_{MD} \approx 10^{-13} \text{ s} \) is the characteristic atomic vibration period (\( \tau^{-1} \) is the attempt frequency) and \( E_A \) is the activation energy of the defect, typically on the order of 0.01–3 eV, corresponding to the self-diffusion energy of the defect. As for thermoelastic damping, dissipation is maximized when \( \omega \tau_{MD} = 1 \), resulting in a Debye peak. This frequency-dependent dissipation peak in temperature has been observed in a number of NEMS structures and materials; some are summarized in Table 4. The Debye temperature is mode shape independent and is observed in flexural as well as torsion modes. Common defects have been identified, such as the hydrogen-sorption-related defect known to occur in silicon near the surface [183]. More generally, a material may have a wide spectrum of defects, each with a different activation energy and relaxation time. Hence, a sharp Debye peak is not always detected [111]. Assuming a continuum of defects, all contributing equally, dissipation becomes

\[ Q_{MD}^{-1} = \sigma N \sum_{E_A} \frac{\omega \tau_{MD} \exp \left( \frac{E_A}{k_B T} \right)}{1 + \left( \omega \tau_{MD} \exp \left( \frac{E_A}{k_B T} \right) \right)^2} \]

\[ \propto T, \]

where the sum runs over all defects of energy \( E_A \) and \( \sigma N \) is a normalization factor. (Here, it is assumed that the defect density is temperature independent. Strictly, this is not true due to thermal expansion and interactions [184].) For an attempt frequency of 10 GHz and resonator frequency of 10 MHz, energies of \( \sim 0.01–1 \text{ eV} \) contribute in the temperature range considered here.

The temperature dependence of a single defect and a continuum of defects are illustrated in Fig. 17. Apart from the characteristic Debye peak caused by a dominating species of defects, NEMS devices show a power law in temperature-dependent dissipation in the 10–300 K range of the form \( Q^{-1} \propto T^\alpha \) [183,18,109,111]. If the density of defects is independent of defect energy, then \( \alpha = 1 \) is predicted (see Eq. (106)). Mohanty et al. [18] have approximately measured this dependence but as illustrated in Table 4, \( \alpha \approx 0.2–0.3 \) has been observed for a wide range of resonator frequencies and materials. \( \alpha \) can be reduced by increasing the density of low-activation energy defects compared to higher-energy defect densities. Czaplewski et al. [66] have measured the relative defect concentration in amorphous carbon resonators and found only a weak dependence on activation energy, where the density varied by less than 50% for an \( E_A \approx 0.35–0.6 \text{ eV} \) range (decreasing density for increasing activation energy). Where it is possible to introduce a defect density distribution that will result in
Fig. 18. Examples of mechanical damping and the Debye peak of resonators measured at low temperatures. (a) Polycrystalline diamond doubly-clamped resonators over a wide frequency span. Debye peak is indicated by arrows; dotted line shows the power law $Q^{-1} \propto T^{0.2}$. Inset depicts Arrhenius plot following the $\ln(f) \propto -1/T$ dependence. (Reprinted with permission from [109]. © 2004, AIP Publishing LLC.) (b) Temperature sweep of silicon and gallium-arsenide resonators showing stronger temperature dependence; red lines indicate power laws added by authors. Source: Reprinted figure with permission from Ref.[18]. © 2002, by the American Physical Society.

A $Q^{-1} \propto T^{0.2-0.3}$ dependence, a theoretical justification or model is still missing in the literature. For amorphous solids or glassy materials, Ref. [185] describes the dissipation as the result of thermal transition in an asymmetric double-well potential. The high surface-to-volume ratio of NEMS resulting in a large defect density means that even crystalline materials can behave like glassy materials. This is discussed in full in the next section. Here we consider Eq. (106) tuned to this double-well scenario, as has been done by Topp and Cahill [184]

$$Q_{DW}^{-1} = \frac{\gamma^2}{k_B C_{ii}} \int_0^{\infty} \int_0^{\infty} \frac{\omega T}{1 + (\omega T)^2} \text{sech}^2 \left( \frac{\Delta}{2k_B T} \right) f(\Delta) g(V) d\Delta dE_A \approx \frac{\pi \gamma^2 f_0}{C_{ii}} g(E_A) k_B T,$$

(107)

where $E_A$ is now the barrier height, $\gamma$ is the coupling constant of the defect with the strain field, and $C_{ii}$ is the elastic constant. $g(V)$ and $f(\Delta)$ are the distribution functions of the barrier height and asymmetry energy, respectively. The final expression is the result for glassy materials for which $f(\Delta) = f_0$ can be considered constant and $g$ is a smooth, slowly varying function. The variation in dissipation in different materials is explained by variations in $g$ and $f$, which are material dependent.

The activation energy measured through the Debye peak characterizes the type of defect. For example, vacancy self-diffusion and self-interstitial diffusion energies in tetrahedral carbon are 2.3 eV [186] and 1.3 eV [187], respectively. (A full discussion on defects and their activation energy for diamond is given by Ref. [187].) Lower activations, such as the 0.05 eV in amorphous SiO, are attributed to hopping of oxygen atoms between almost degenerate equilibrium positions [188]. Both silicon and gallium arsenide are known to have defects with energy characteristics in the 6–100 K range. A complete discussion of types of defects and associated activation energies is given in Ref. [36]. A range of examples and calculations are given for point defects and Zener, Snoek, dislocation, and boundary relaxation, as well as electronic related dissipation phenomena. In NEMS, which are often far from ideal materials, the type of defect that contributes is difficult to determine;
often, only the general behavior can be determined. As discussed by Czaplewski et al. [66], other amorphous materials will have similar types of defects and associated energies. As the system size is reduced and the surface-to-volume ratio grows, the relative number of defects will grow as discussed above. Eventually, even crystalline structures behave like amorphous solids, often called glassy due to the large surface area. Liu et al. [189] illustrate how increased doping increases dissipation in diamond-like carbon films. In addition, annealing releases stress and further increases the quality factor by reducing the structure disorder. The thermal activation temperature is typically on the order of 10 K and higher, below which no thermally activated transitions are possible and the dissipation becomes temperature independent. As the temperature drops to the sub-kelvin scale, only the ground state and almost degenerate excited states are active, where now transitions occur. This phenomenon, described in the next section, is of fundamental interest and has been studied at depth in thin films as well as NEMS resonators.

5.4. Quantum dissipation $Q_D^{-1}$

Anderson, Halperin and Varma [190] proposed a model, now known as the Standard Tunneling Model of Two-Level Systems, that describes thermal properties of glasses, including specific heat, thermal conductivity, and internal friction. The model assumes that glasses contain a finite number of atoms (or groups of atoms and defects) that occupy and almost degenerate ground state. These form a statistical distribution of localized tunneling states. These almost degenerate states can transition from one to the other by phonon-assisted tunneling, either emitting or absorbing a phonon to conserve energy in the process. This section introduces the Two Level Systems (TLS) model and discusses its role and limitations when applied to NEMS devices.

5.4.1. Standard tunneling model of two-level systems

At low temperatures, the atoms in a glass may relax into an almost degenerate ground state, as illustrated in Fig. 19, where the energy boundary $V \gg kT$ forbids thermal transitions from one local minimum into the other. Quantum mechanically, however, the atoms or groups of atoms can tunnel through the barrier, emitting or absorbing a phonon in the process. The tunneling probability and rate depends not on the absolute energy of the two states but rather the energy difference $\Delta$ and the off diagonal tunneling barrier $\Delta_1$ (also known as the splitting energy). This type of two-state model with weak coupling is described quantum mechanically by a perturbative Hamiltonian

$$H = H_0 + H_1$$
where the zero energy level is set to the center of each well. TLS move in an asymmetric double-well potential with asymmetry $\Delta$ and splitting energy $\Delta_1 \approx 10^{-4}$ eV, which is on the order of the thermal energy at 1 K. A common characteristic of these models is the splitting into two energy regimes. At low temperatures, there is resonant absorption of the TLS, and at high temperatures, the TLS phonons modulate the asymmetry. In glassy materials, there is a continuous distribution of splitting energies ($\tilde{P}(\Delta, \Delta_1) = \delta_0 \delta_{\Delta_1}$), and the number density of two-level systems is assumed to grow linearly with temperature. Where the density of states with small energy difference may be large, the number of systems that are of low-enough mass and sufficiently close in space to allow for quantum tunneling may be relatively low. Still, such systems dominate thermal properties at low temperatures.

As for the thermally activated defects and the thermoelastic dissipation described previously, the TLS are characterized by a relaxation time defined through the tunneling timescale. At very low temperatures, the relaxation rate $\tau^{-1} \ll \omega$ is small compared to the applied strain frequency (in this case, the resonance of the beam); the defects do not reach the thermodynamic equilibrium state. At higher temperatures, the relaxation rate is much faster than the changes in the strain field, and the thermodynamic equilibrium is established. The corresponding dissipation regimes take the form [184]

$$Q_{\text{TLS}}^{-1} = \frac{\pi}{96} C_{l,t} \left( \frac{T}{T_{\text{co}}} \right)^3 T \ll T_{\text{co}}$$

$$Q_{\text{TLS}}^{-1} = \frac{\pi}{2} C_{l,t} T \gg T_{\text{co}},$$

where $C_{l,t} = \frac{\rho v^2}{a^2 l^2}$ is the tunneling strength for longitudinal and transverse vibration modes ($\tilde{P}$ is the spectral density of tunneling states) and $T_{\text{co}}$ is the characteristic temperature defined at which the maximum relaxation rate equals the angular frequency of the resonator ($\tau^{-1} = \omega$).

A corresponding shift in the sound velocity (or resonance frequency) is predicted

$$\frac{\delta v}{v_0_{\text{TLS}}} = \frac{v - v_0}{v_0} = C \ln T \quad T \ll T_{\text{co}}$$

$$\frac{\delta v}{v_0_{\text{TLS}}} = \frac{v - v_0}{v_0} = -\frac{C}{2} \ln T \quad T \gg T_{\text{co}},$$

where $v_0$ is the sound velocity at a fixed temperature, typically the lowest temperature measured. A complete discussion of the standard TLS tunneling model is given by Esquinazi [192].

### 5.4.2. Variations on the standard tunneling model

Glasses are defined here as materials with a broad distribution of defect energies and corresponding splitting energies and barrier heights. Perfect crystals will not fulfill this criteria, but as discusses above, nanocrystal structures with large surface-to-volume ratios have a correspondingly large surface defect density, where the defects act like two-level systems at sufficiently low temperatures, explaining the glassy behavior in crystalline materials. Extensive experiments have been conducted on the dissipation in thin films of both amorphous solids [193,194,133,184,192]; metals [195,192,131,196,197]; and crystalline structures [135,184,192] for temperatures ranging from $10^2$ down to $10^{-3}$ K [198]; a review on the dynamics of two-level systems is given by Leggett et al. [199]. Where at higher temperatures large variation in behavior and dissipation values are recorded, there is an observed universality in amorphous and crystalline structures at low temperatures, where the tunneling model of two-level systems applies. Most of these experiments have been conducted at relatively low frequencies and not in NEMS devices. In this section, we discuss some of the results and interpretations that are then applied to NEMS devices in the following section.

Harrington et al. [200] reported TLS in NEMS gallium nitride torsional resonators and Mohanty et al. [18] describe the effect of two-level systems in a crystal, where only a single defect species is considered. This means only one value of $\Delta$ and $V$ is included in the model. Using Fermi’s golden rule, the single phonon relaxation rate of the excited TLS can be determined. As before, the dynamic response falls into a low-temperature range defined by $k_B T \ll \Delta$ when resonant absorption interactions between an acoustic phonon and the TLS results in an absorption of a phonon of energy $\hbar \omega = \Delta$. If the TLS is in the higher energy level, the phonon will induce a stimulated emission of a phonon of energy $\hbar \omega = \Delta$. At higher temperatures, relaxation absorption dominates as the strain field modulates the asymmetry energy and the thermodynamic equilibrium is established. The dissipation for a single defect in a crystalline material becomes

$$Q_{\text{TMC}}^{-1} = \frac{c}{k_B T} e^{-\frac{\Delta}{k_B T}} T \ll T_{\text{co}}$$

$$Q_{\text{TMC}}^{-1} = \frac{d}{T}, \quad T \gg T_{\text{co}},$$
where \( c = \frac{2ny_{0}^{2}}{\rho v^{2}(\omega_{0}^{2} + \omega_{L}^{2})^{3}} \) and \( d = \frac{\omega_{0}}{\kappa_{B} T_{0}} \). Correspondingly, the frequency shift takes the form

\[
\frac{\delta f}{f_{0}}_{\text{TMC}} = -\frac{ny_{0}^{2}}{\rho v^{2}} \Delta T \ll T_{co}
\]

(113)

\[
= -\frac{ny_{0}^{2}}{2\rho v^{2}k_{B}T} \quad T \gg T_{co}.
\]

(114)

The exponential decrease in dissipation has not been observed; instead, for sub-kelvin temperatures \( T < 100 \text{ mK} \), the dissipation appears to saturate at a finite value \([110]\). A less constrained model than the single crystal described above, but still more rigidly defined than the standard model is proposed by Gilroy and Phillips \([185, 193]\). Specifically for crystals with defects at very low temperatures \([201]\), the density of defect states is a Gaussian (or an exponential) function of the asymmetry energy of an asymmetric double-well potential with a fixed splitting energy \( V \), so that \( P(\Delta) \propto \exp(-\Delta^{2}/2\n^{2}) \) with an experimentally defined width \( \Delta_{1} \). In agreement with the standard glass model, the frequency shift is logarithmic in both temperature ranges above and below \( T_{co} \), with a change in sign of the slope. Dissipation follows a linear dependence on temperature and saturation at higher temperatures.

To describe the large phenomenology observed in experiments, other models have been proposed, including the soft potential model (SPM) \([202, 203]\) that assumes a quartic potential and constant density of states for each mode, based on localized low-frequency elastic waves. The SPM is successful in explaining low-temperature thermal conductivity and predicts a \( Q^{-1} \propto T^{2} \) dissipation law with low-temperature saturation. In the soft potential model, the low energy excitation potential takes the form

\[
V_{\text{SPM}}(x) = W \left[ D_{1} \left( \frac{x}{a} \right)^{4} + D_{2} \left( \frac{x}{a} \right)^{2} + \left( \frac{x}{a} \right)^{4} \right],
\]

(115)

where \( W, D_{1}, \) and \( D_{2} \) define the double-well potential. The density of states is a Gaussian of the form \( P(D_{1}, D_{2}) = P_{0} \exp \left[-0.169D_{1}^{2}/(W/k_{B}T_{g})^{2} \right] \), where \( T_{g} \) is the freezing temperature of the glass. The \( D_{1} \) parameter introduces the asymmetry originating from frozen static thermal strain. \( D_{1} \) values show a Gaussian distribution. \( D_{2} \) values have a cutoff at \( k_{B}T_{g}/2 \) \([202]\). The potential well differs for each localized mode and TLS.

5.4.3. TLS in NEMS

Models specifically applied to tunneling states in TLS have been proposed by Seoanez et al. \([40]\) to include the effect of finite size constraints and the large surface-to-volume ratio. They illustrate how the oscillating strain field pumps the TLS out of their equilibrium states. Two mechanisms of dissipations are identified: the modulated strain fields combined with the finite relaxation time prevents the TLS from establishing a thermodynamic equilibrium and thereby pumps energy out of the strain field, a relaxation absorption mechanism. At lower temperatures, sub-ohmic damping is described through the coupling of the asymmetrical TLS to the ensemble of vibration modes through the strain field through resonant absorption. In addition, most NEMS structures are a composite of multiple materials, such as a dielectric device layer and a conductive surface. Coupling to electronic degrees of freedom sets a lower limit on the dissipation; however, this limit is far below typically measured NEMS devices, even at sub-kelvin temperatures. Mohanty et al. \([204]\) propose that at low temperatures, quantum-mechanical tunneling of two-level systems at high frequency adds dissipation through phonon pumping. This dissipation would only be weakly temperature-dependent and dominate at very low temperatures. Superradiance-enhanced dissipation by collective relaxation may add to dissipation when the structure size is the same order as the phonon wavelength. When including nonlinear behavior, the TLS phonon-emission theory results in a non-vanishing contribution to the dissipation at zero temperature. This is the result of enhanced phonon emission due to superradiance (SR), resulting in the dissipation at zero temperature of

\[
Q_{SR}^{-1} \propto \frac{1}{\omega_{*}^{2}} \frac{ny_{0}^{2}}{\rho v^{2}} \quad T = 0
\]

(116)

where \( \tau_{*}^{*} \) is the renormalized relaxation time per TLS and \( n \) the number of TLS per unit volume. This model agrees phenomenologically with the observed saturation of the dissipation of low temperatures; however, it has not been fitted to NEMS resonators thus far.

At ultra-low temperatures (< 10 mK) interacting TLS must be included in an interaction model, as has been demonstrated by Fefferman et al. \([198]\) for low frequencies. An increase in dissipation has also been demonstrated with the onset of superconductivity. This has attributed to interaction TLS that result in a stain-dependent and hence nonlinear dissipative mechanism described in Ref. \([205]\).

The dissipation mechanisms and dependencies at ultra-low temperatures is of significant importance. As the temperature is lowered higher frequency resonators, the thermal energy eventually drops below the quantum energy of an oscillator \( (k_{B}T < \hbar \omega) \). As discussed above, the direct observation and manipulation of a macroscopic quantum oscillator has been the focus of a large number of theoretical and experimental studies. NEMS experiments on the sub-kelvin temperature dependency of dissipation are, however, relatively rare.
TLS in NEMS. (a) $Q^{-1} - T$ dependence for three resonance frequencies at 3 T. We measure $Q^{-1} \propto T^\alpha$, where $\alpha \approx \frac{1}{3}$. Below 100 mK dissipation saturates. Inset: Dissipation–frequency relation at 6 and 3 T, including fits for magnetomotive damping ($Q_{MMD}^{-1} \propto f_0^{-2}$) and clamping loss ($Q_{CL}^{-1} \propto f_0^2$). $T_{MC}$ is the mixing chamber temperature. (b) $\delta f/f_0 - T$ dependence, $f_0$ is the resonance frequency at 35 mK. Below and above a characteristic temperature $T_c$ ($T_c = T_{co}$ in this manuscript), the shift is logarithmic as predicted by TLS models. The inset depicts the slopes of $\delta f/f_0$ and $f_0^{-1}$ fit for low (up triangles) and high (down triangles) temperature regimes. (c) Normalized dissipation $Q^{-1}/Q_{T_{co}}^{-1}$ for UNCD, Si[110], and GaAs[95] for frequencies ranging from 5 to 45 MHz. The temperature has been scaled to $T_{co}$ (see panel(b)). The black dotted line represents a slope of $\frac{1}{3}$, alluding to the same power law for all three materials. The other slopes represent predictions of various TLS models. (d) $(\delta f/f_0)_{N}$ vs. $T/T_{co}$. $f_0$ is the frequency measured at 0 $T_{co}$. The curves collapse if the ratio $s_2/s_1$ is the same. The dashed lines illustrate four $s_2/s_1$ ratios. $s_2/s_1 = -0.5$ and $s_2/s_1 = -1$ depicts the predictions of the TLS standard tunneling model [192] and the Phillips model [193] respectively. Source: Reprinted figure with permission from Ref. [34]. © 2009, by the American Physical Society.

Zolfagharkhani et al. [110] identified TLS in single crystal silicon in doubly-clamped MHz resonators. The characteristic frequency shift and dissipation power law agree qualitatively with TLS models; however, exact fits to experiments were not made. Using the magnetomotive actuation-and-detection technique, a weaker than expected dependence on temperature was found for the dissipation $Q^{-1} \propto T^{0.36}$. The logarithmic frequency shift at the lowest temperatures indicated resonant absorption of TLS with a broad distribution of energy states corresponding to glassy or disordered crystalline materials. Surprisingly similar results have been measured in gallium-arsenide samples [95] where the dissipation dependence on temperature was determined to be $Q^{-1} \propto T^{0.5}$. Imboden et al. [34] measured the low-temperature dynamics of ultra- nano-crystalline diamond structures over a wide frequency range. The data are compared directly with the experiments in silicon and gallium arsenide by appropriate scaling to the crossover temperature $T_{co}$. The dissipation power law $Q^{-1} \propto T^{3}$ appears to be universal for all materials, a surprising result considering how different single-crystal silicon is to a practically amorphous $sp^3$ carbon material. It is proposed that the response variation in the frequency shift (characterized by the ratio of the slope $s_2/s_1$ below and above the characteristic temperature $T_{co}$) may be explained by universality classes, that differ by the spectrum of density of states. Analogous to the proposals of Phillips [193], where changing the density of states from flat to a Gaussian changed the slope ratio from $- \frac{1}{2}$ to $-1$. Where silicon and UNCD have a slope ratio of $-2$, gallium arsenide shows an even stronger crossover from the resonant to the relaxation regimes with $s_2/s_1 \approx -4$. These results are depicted in Fig. 20. A corresponding density-of-states distribution has not been identified. As pointed out in Ref. [184], the universal behavior at low temperature in amorphous solids and disordered crystals is not reproduced at higher temperatures, indicating that the TLS are not the low-energy tail of the mechanical defects that dominate at higher temperatures that are discussed in the previous section.

A more recent study of gold doubly-clamped nanobeams by Venkatesan et al. [112] demonstrated the dominance of TLS in a purely metallic structure. Again, the temperature dependence $Q^{-1} \propto T^{0.5}$ was lower than predicted by most TLS models.
but higher than observed in Si, GaAs, and UNCD. It is postulated that the effects of size constraints [40] as well as tension may shift the density of states of the TLS and hence modify the low-temperature dynamical response. Hoehne et al. [197] conducted similar measurements in polycrystalline aluminum NEMS structures, where the frequency shift shows again logarithmic behavior and the dissipation scales linearly in the low-temperature regime, followed by a weaker dependence at higher temperatures. The proposed explanation is a one-dimensional phonon mode model with a constant density of states. The results for the dissipation and frequency shift due to TLS are summarized in Table 5. Single crystal diamond (SCD) resonators have been reported with low temperature quality factors as high as $Q = 5.9 \times 10^6$, as well as the strong temperature dependency of $Q^{-1} \propto T^{1.6}$. The results for single crystal diamond are very promising for the next generation of NEMS devices; the structures presented in Ref. [206] are significantly larger and hence orders of magnitude lower in frequency compared to most resonators discussed here. It will have to be seen if the low dissipation in single crystal diamond is maintained as resonant frequency is raised from order $10^4$ to $10^7$ and beyond. While no model seems capable of reproducing the rich behavior observed in experiments, the power law appears to be a universal phenomenon for a whole class of crystal and polycrystalline structures, calling for a search of a theoretical explanation. The interpretation of the saturation at very low temperatures is still open to debate.

Fig. 21 depicts the crossover temperature from the resonant absorption to the relaxation absorption regimes for a very wide range of frequencies in a range of materials. Over six orders of magnitude in frequency $T_{co}$ the power law appears to be a universal phenomenon for a whole class of crystal and polycrystalline structures, calling for a search of a theoretical explanation. The interpretation of the saturation at very low temperatures is still open to debate.

6. Dissipation in carbon materials

Top-down fabrication methods, such as e-beam lithography, allow for a high level of control of device geometry and can reduce device dimensions down to the range of tens of nanometers. The ultimate miniaturization is, however, achieved by manufacturing resonators out of carbon nanotubes (CNT) and graphene sheets directly, where resonators of single atomic thicknesses are now possible. CNT and graphene show extraordinary electrical and mechanical properties (see Table 1). There has been a dramatic amount of research in turning these novel materials into resonant structures. Experiments show promising results in expanding the observed parameter space; however, for flexural mode resonators the quality factors of these novel devices have typically been disappointing and offer a serious limit to performance and sensitivity. A full discussion would merit its own review on this subject; this chapter summarizes some of the current experiments and discusses possible sources of the dominant dissipation mechanisms. In this article, diamond structures have been discussed in full; the following will focus on CNT and graphene structures.

### 6.1. Carbon nanotubes

Tunable CNT resonators have been reported in 2004 by Sazonova et al. [47], who managed to suspend single- and multi-walled carbon nanotube (SWCNT and MWCNT) over an electronic gate and actuate them through capacitive coupling. The motion was detected through the induced strain, which altered the current flowing through the CNT from a source to a ground electrode (analogous to the piezoresistive detection method). By mixing the drive-and-detection signal, a high-frequency mode can be observed at low frequencies. Varying the gate voltage causes significant shifts in the resonance frequency and hence allows the resonator to be tuned, but the gate voltage also contributes to dissipative effects. Later experiments demonstrated the extraordinary properties of CNT resonators, such as mass sensing down to the order of $10^{-24}$ kg [28]; frequencies over a gigahertz at room temperature and pressure [26]; and the smallest radio receiver ever built [213,214]. An example of a freely standing doubly-clamped CNT resonator is depicted in Fig. 22. Typical quality factors are considerably lower for CNT compared to the slightly larger resonators described in previous sections and...
Fig. 21. Characteristic temperature $T_\text{co}$ plotted vs. frequency derived from sound velocity and mechanical measurements for a large range of materials and resonators. Below 100 kHz $T_\text{co}$ lies in the 50–200 mK range, where above the 10 MHz $T_\text{co}$ ranges from 600 mK to 4 K for $\approx$1 GHz frequencies. The dependence $T_\text{co} \propto f^{0.29}$ is a reasonable fit with numerous outliers over a range of six orders of magnitude in frequency. The colors represent the slope ratio of the frequency–temperature relation for below and above $T_\text{co}$. Standard TLS models predict this ratio to be $-\frac{1}{3}$ [185,135] in surprisingly good agreement, considering that the standard model does not explain much of the phenomenology observed. (Black circles, standard glass model) or $-1$ (red circles, Phillips model); these reported ratios are estimates, not exact results. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Source: This data are collected from Refs. [207–209,135,210,195,211,212,196,110,95,34].

Fig. 22. CNT doubly-clamped resonator.
Source: Reprinted with permission from [215]. © 2006, American Chemical Society.

range from 5 to 500 [47,215,26,27,216]; however, lowering the temperature shows a dramatic reduction in dissipation allowing for quality factors up to 2000 [28]. Huttel et al. [78] succeeded in measuring doubly-clamped SWCNT flexural modes with quality factors exceeding $10^5$ at resonant frequencies of 350 MHz when tensile strain is applied. Though these measurements were taken at temperatures as low as 20 mK, the extraordinary high quality factors are not fully explained. It is suggested that the fabrication method leaves the resonators exceptionally clean compared to other methods, indicating that surface contamination may typically contribute to dissipation. Improvement in quality factors through cleaner fabrication procedures is also proposed by Nishio et al. [217]. In addition, tension is known to improve the quality factor, as is discussed in Section 7.1. The novel detection scheme using the CNT as a quantum dot may minimize circuit losses as well.

Seoanez et al. [218] studied the dissipation in CNT and graphene from a theoretical perspective guided by data from Ref. [219]. Where graphene is chosen for the analysis, it is argued that the mechanisms identified would also apply equally to CNT. The dominant dissipation identified at room temperature is predicted to be due to ohmic losses in the graphene or CNT and metallic gates. Doped silicon gates are often used for device actuation and detection [47,220,215,26,27,97,28,78,216] and may contribute in a similar fashion to metallic gates. Electrons (which are known to occur as “electron and hole puddles” even in grounded materials) in the resonating structure couple through a time-dependent potential to the charges in the gates and thereby contribute to energy losses. It is argued that the creation of electron–hole pairs implies the mechanism is ohmic and hence should depend linearly on temperature. Dissipation due to dipole interaction with trapped charges in the silicon base is estimated to be negligible. For low temperatures, Ref. [218] predicts attachment losses should dominate following the discussion in Section 4.2 for radiative losses into the base. Where CNTs are extremely thin are should hence result in low clamping losses, stacked graphene may be limited by such a loss mechanism. Ganzhorn et al. [102] have demonstrated...
CNT with quality factors as high as $2 \times 10^5$ at low temperature, where ohmic losses which scale as $\propto \sqrt{V}$ In the presence of a magnetic field, a CNT resonator will suffer from eddy currents resulting in significant magnetomotive damping [113].

The mechanical impedance mismatch between the carbon materials and the silicon dioxide base will further reduce radiative losses, as the attachment of the graphene and CNT to the base introduces a new loss mechanism known as the Velcro effect. Typically, on a silica base, hydrogen bonds hold the graphene and CNT in place. The binding energy of each hydrogen bond is $10^{-1}$ eV, significantly higher than the total energy contained in a low amplitude mode. However, for strongly driven nonlinear structures, the Velcro effect may very well come into play. This analysis may oversimplify the different fabrication techniques of freely standing CNT structures and, as proposed by Witkamp et al. [215], differences in observed quality factors may be traced back to differences in the attachment mechanisms. Sazonova et al. [47] also predict that “the tube may lose energy by sticking and unsticking from the surface during oscillation” and proposes altering attachment geometry to determine the contribution. It must be noted that the Velcro effect, just like attachment losses, is temperature independent and hence fails to explain the drastic improvement in quality factors observed at low temperatures. Three distinct attachment types studied in Ref. [221] report no correlation with dissipative losses. The role of phonon assisted clamping losses in CNTs at low temperatures has been investigated theoretically by Zhong et al. [222], based on the work of Wilson-Rae [223] which provides a comprehensive theoretical analysis of phonon-assisted clamping losses at low temperatures. This theoretical work is also used as the basis to characterize and predict quantitatively phonon-tunneling attachment losses in a free–free geometry [224].

Jiang et al. [225] investigate the molecular dynamics in singly clamped structures and determine the dissipation in SWCNT to follow a weak temperature dependence, predicting $Q^{-1} \propto T^{0.36}$. The model uses a Tersoff–Brenner potential that allows for covalent bond formation and breaking. It is noted that the low temperature dependence looks similar to that observed in other nano-resonators as discussed in the TLS section (see Section 5.4.3). Where the model is defect free, it is argued that defects may occur in the distorted lattice at the ends of the resonators. Seoanez [218] argues that the TSL can only dissipate energy if they are charged and would most likely occur in the base. Hence they are weakly coupled to the mechnical resonance and do not contribute in any significant manner to dissipation. It is of interest that measurements in the range of $10$ mK $< T < 1$ K by Ref. [78] resulted in the $Q^{-1} \propto T^{0.36}$ dependence, matching the theory for intrinsic losses perfectly and also matching the dependence found for the top–down manufactured nanomechanical resonators discussed above.

Many of the low quality factors observed have been measured at ambient pressure. The diameter of CNT resonators is on the order of 1 nm, much lower than the mean free path of molecules at ambient pressure (68 nm [226]); therefore, the low-pressure model for dissipation discussed in Section 4.4.1 should apply. Garcia et al. [27] measured 3.1 GHz beam with a quality factor of five, a factor of six below that predicted if the dissipation were due solely to air damping. These results are very similar to the measurements of top-down resonators of similar frequency [139]. Sazonova [47] plots pressure vs. quality factor, showing equivalent behavior as seen in Fig. 13. The quality factor improves with falling pressure until it saturates below 1 Torr at $Q \approx 40$ where some sort of clamping or attachment losses most likely begin to dominate. For growing pressure, the quality factor drops below 14; as the pressure exceeds 10 mTorr, there appears to be a $Q^{-1} \propto P^{4}$, but the data is insufficient to exclude a linear dependence.

As will be seen below, intrinsic strain significantly alters the resonance frequency as well as the quality factor of a resonator. It has been shown that strain can be added to CNT to alter the resonance frequency. It is very possible that attachment strain alters the quality factor and may help explain the low values measured. Wu et al. [74] discusses variation in intrinsic strain (both positive and negative) due to the fabrication processes. Witkamp et al. [215] claim to have no compressive strain determined by the SEM image that shows no buckling. Where this may be a good indicator compared to the image of the stacked CNT depicted in Ref. [47], compressive strain well below the buckling instability can result in lower frequencies and higher dissipation; in addition, the quality factors reported in Refs. [215,47] are very similar. It must be noted that the thermal expansion coefficient of CNT is negative at room temperature and below [75]. Hence a CNT will elongate as it is cooled compared to the silicon base on which it is mounted; therefore, one would expect more compressive strain and buckling at lower temperatures. However, the opposite has been reported, as is discussed below. Additionally, as shown by Sazonova et al. [47], the resonance frequency can be tuned by adding strain to the CNT through an applied gate voltage. Typically, additional strain should improve the quality factor as well, as is discussed below; however, the quality factor is shown to decrease rapidly with applied gate voltage indicating that ohmic or electron–hole pair effects dominate.

One way to exclude compressive strain and reducing clamping and attachment losses is through singly clamped cantilevers, as presented by Jensen et al. [213]., Nishio et al. [217], and Weldon et al. [227]. Quality factors over 500 are achievable, an improvement of an order of magnitude compared to many other experiments, but cantilevers typically also have a lower resonance frequency. In addition, the cantilever drive-and-detection methods typically do not require a gate; hence, electron–hole pair coupling is minimized.

It is instructive to discuss the contribution due to thermoelastic dissipation. As noted in Section 5.1, as the device size becomes small the thermal time scale is very short, so unless the resonance frequency grows extremely fast no temperature gradient and hence no heat flow occurs. The system is in its thermodynamic equilibrium and thermoelastic dissipation is minimal. As CNTs are extremely small and in addition the thermal conductivity is relatively high, CNT resonators fall into this category. This can be seen in the lower plot of Fig. 24, where thermoelastic dissipation is shown to increase with increasing resonance frequency. At room temperature the resonance frequency would have to be on the order of a 1 THz before thermoelastic dissipation begins to limit the resonators performance. As noted in Ref. [221] the minimum quality
Fig. 23. CNT resonator dependency on (a) pressure (data extracted from Ref. [47]) and (b) temperature. (Reprinted with permission from [221].) (a) Shows that below 1 Torr the dissipation is pressure-independent. It is predicted that for higher pressure $Q^{-1} \propto P$, hence stronger than indicated here; however, the data range is too incomplete to discard this possibility. (The data is replotted from the published graph.) (b) Shows the dissipation scaling with temperature for a large number of CNT resonators with varying applied voltages. Both $T$ (possibly ohmic damping) and $T^2$ (possibly thermoelastic damping) may fit the data.

factor as defined by the first terms in Eq. (95) can be as low as 20, this is mainly due to the high value of $\alpha^2 E_{\text{CNT}}$. When taking the full expression into consideration, including the thermal time constant the minimum quality factor is increased by $10^3$ for 1 GHz resonances. For lower frequency CNTs the thermoelastic equilibrium is established over a longer time period allowing for the process to occur adiabatically and hence less loss is associated with the process resulting in higher quality factors (see Fig. 24). The top graph of Fig. 24 depicts the temperature dependency for the maximal possible thermoelastic dissipation as well as what is expected for a 1 GHz resonator, it includes the explicit temperature dependence derived in Section 5.1 as well as the temperature dependencies of the material properties. Due to the unusual temperature dependency of the thermal expansion coefficient $\alpha$ the quality factor diverges both at zero temperature and around 1000 K as this is where $\alpha$ changes sign. The data points in Fig. 24 are calculated using results for the thermal dependence of $\alpha$ [77], $\kappa$ [73], $C_p$ [76] as well as using $\rho = 2300 \, \text{kgm}^{-3}$ and $E = 1000 \, \text{GPa}$ [221], where the authors are not aware of any temperature dependence reported for the Young's modulus for SWCNT. Calculations by Jiang et al. [228] indicate that for graphene no significant temperature dependency of the Young's modulus is expected. It must be noted that the values used here are meant to be representative, but considerable variations of the reported values for thermal properties exists in the literature of CNTs as well as variations in different species of CNTs.

So what does explain the high quality factors, exceeding $10^3$ at room temperature and reaching as high as $10^5$ measured in Refs. [220,28,78]? For the highest quality factor the experiments are taken at low temperatures, and the intrinsic dissipation predicted in Ref. [225] is achieved. Hall et al. [220] achieve a quality factor over $10^3$ at room temperature for a torsional mode, where clamping losses are typically lower than for flexural modes, this modes is of considerably lower frequency compared to the others reported here. Clamping and attachment losses are temperature independent, therefore these effects cannot be the sole culprit.

One explanation is that defects and absorbed impurities occurring during the fabrication process contribute as the leading loss mechanism. Typically if such defects manifest themselves as electron–hole pairs that couple to the vibrational mode one would expect ohmic losses and a linear dependency on the dissipation with respect to temperature as discussed in Ref. [218], this dissipative coupling is also considered in Ref. [216] to be the leading cause of dissipation in CNT resonators used as quantum dots, where the quality factor oscillated with increasing gate voltage. Sazonova et al. [221] studied a
large number of CNT resonators over a large parameter space (see Fig. 23). No deterministic behavior can be correlated to device resistance or attachment type. Strong dependencies are observed however with temperature and gate voltage. Thermoelastic damping would explain dissipation dependency proportional to $T^2$ when taking the material properties dependencies into account. The data would also allow for a linear dependence. The complex temperature dependencies of the physical processes and material dependencies make the analysis tricky. For more precise theoretical modeling to be fitted meaningfully more experiments are needed. Vallabhaneni et al. [229] analyze dissipation in SWCNT using molecular dynamics simulations. It is shown that the dissipation in axial vibration modes are length and diameter independent, whereas in transverse modes dissipation drops with increasing length scales and increases with growing diameter. Including quantum effects it is predicted that dissipation scales as $Q^{-1} \propto T^\alpha$ with $0 < \alpha < 1$, slower than other predictions [229].

6.2. Graphene

The ability to suspend and measure the resonant mechanical response of graphene sheets has opened the door to unprecedented NEMS devices. As is the case for CNT resonators, graphene’s unique two-dimensional structure as well as its exceptional mechanical and electrical properties allow for new studies on the nanoscale. As with CNT resonators, the quality factor at room temperature proved to be a limiting factor to sensitivity and device performance. The theoretical study of dissipation in graphene resonators is given in the previous section, following Seoane et al. [218]. It is assumed that dissipative mechanisms are the same for graphene and CNT resonators. Kim et al. [230] modeled the temperature dependence of dissipation for single- and multi-layered graphene and found in agreement with MWCNT studied in Ref. [225] that intrinsic dissipation is increases in multi layers due to translational slipping made possible by the weak interlayer van der Waal bonds. The resulting temperature dependency $Q^{-1} \propto T^\alpha$ is weaker than $\alpha = 1$. Varying the van der Waal bond strength with regards to a silicon base significantly increases the dissipation and reduces $\alpha$. Increasing the attachment strength reduces dissipation by preventing translational slipping.

Measurements by Garcia et al. [29] done in air resulted in low quality factors between 2 and 30. This can be explained by air damping; due to the small gap size squeeze-film damping discussed in Section 4.4.3 can explain the observed values. Following the work done by Bunch et al. [219] on the detection of graphene resonators, there has been an explosion of interest. Room temperature quality factors are typically on the order of 50–250 [219, 231, 232, 30], with lower values reported for resonators operated at ambient pressures [29]. Significant improvements have been observed at lower temperatures [219, 232, 30, 33] with values reaching as high as 14,000 at $T = 5$ K [79]. An example of a graphene resonator and an AFM detection technique is illustrated in Fig. 25. Bunch et al. [219] was not able to discern a size dependency on the dissipation as has been done for top-down fabricated NEMS devices, where it was shown that increasing surface-to-volume ratio leads to an increase in dissipation (see Section 5.2). Whereas increasing the thickness of graphene (by adding monolayers) may lead to a smaller surface-to-volume ratio, it is expected that interlayer slipping will cause a significant amount of dissipation. Geometry and clamping conditions may dominate variation in dissipation over changes in thickness.

As mentioned, a significant improvement in the quality factor has been observed by lowering the temperature. Chen et al. [79] determined the dissipation to scale as $T^{0.36}$ below $T = 90$ K. This low temperature dependence has been observed in a number of NEMS devices including in CNT as discussed above. At higher temperatures, the temperature dependency increases to $Q^{-1} \propto T^2$ (in Ref. [79], it is written $Q^{-1} \propto T^2$; this is probably a typo, as the data fits $Q^{-1} \propto T^2$ very well), faster than predicted by theory. Singh et al. [232] observe a similar crossover at a temperature at 120 K but measured weaker power laws of $T^{0.2 \pm 0.07}$ and $T^{1.73 \pm 0.15}$ for the low and high temperature regimes, respectively. (The authors took the liberty to make their own fits to the data presented in the supplementary material of Ref. [232].) Temperature dependency on dissipation has also been measured by Zande et al. [30]; again, a crossover temperature can clearly be identified, this time around 50 K.
below which $Q^{-1} \propto T^{0.35}$, and for higher temperatures $Q^{-1} \propto T^{2.3}$. The data from these papers are displayed in Fig. 26. Singh et al. demonstrate how added strain, due to differences in the thermal expansion coefficients $\alpha$ of the graphene, metal electrodes and silicon/silicon dioxide base, shifts the resonance frequency [232]. This is true for all resonators in which tensile strain dominates the frequency response, in contrast to rigid structures in which the frequency is determined by the Young’s modulus. As is shown in Section 7.1 below, high tensile strain does not only increase the resonance frequency but can also significantly reduce dissipation. The large increase in resonance frequency has been observed in all experiments discussed above. This change cannot be explained by the temperature effect on the Young’s modulus, which is predicted to have negligible temperature dependency and hence is attributed to strain. The low temperature, $Q^{-1} \propto T^{-0.4}$ behavior agrees with measurements and theory in CNT discussed above, where the mechanism is believed to be intrinsic and caused by distortions of an otherwise perfect lattice structure. The higher temperature dependency where $Q^{-1} \propto T^{-2}$ is also in agreement with measurements from Ref. [221] and can most likely be attributed to changes in tensile strain. Chen et al. [79] and Singh et al. [232] suspended graphene resonators from gold electrodes over silicon dioxide films mounted on a silicon base, whereas in the sample described in Ref. [30], the oxide film is not continuous. This will result in different thermal strain contributions and may explain the lower crossover temperature. The effects of strain on dissipation is discussed in Section 7.1. Qi and Park demonstrate using molecular dynamics how grain boundaries in graphene will reduce the quality factor by 1–2 orders of magnitude [234]. They also suggest however that adding small amounts of tensile strain can restore
quality factors to close to that of defect free graphene. The improvement of the quality factor is not only a result from an increased frequency, but specifically flattens out buckling caused by the defects and thereby suppresses dissipation. Hence strain actively reduces dissipation, opposed to increasing the stored mechanical energy as described in Ref. [37]. It would be of great significance if this result can be observed experimentally.

Shivaraman et al. [231] demonstrate significant increases in frequency and reduction in dissipation after annealing suspended structures. The Ar/H₂ atmosphere at 400 °C annealing procedure intended to remove residue photoresist. This alone would improve the resonance frequency, but in addition, SEM images show a reduction of the side flanges, implying increased tension, which would further improve the resonance frequency and quality factor.

Table 6 summarizes a number of results for low-temperature dissipation measured to obey a \( Q^{-1} \propto T^{\frac{7}{2}} \) power law. This includes four different carbon materials, as well as silicon and gallium arsenide.

### 7. Evading dissipation

To improve device performance, methods have been developed to reduce dissipation. Here, we present three methods that can have a profound influence on the dissipative nature of NEMS structures and hence device sensitivity. The first is mechanical in nature, where an applied tensile strain significantly increases the quality factor. The second is a chemical method where dissipation is reduced through annealing. The third method uses active feedback control circuits that, through parametric amplification, squeeze noise from one parameter space into another.

#### 7.1. High tensile-strain

The elasticity equations used to describe the dynamical response of doubly-clamped beams and cantilevers assume that there is no strain in the material if no external force is applied. For doubly-clamped beams, it is possible to add compressive or tensile strain which will have an effect on both the resonance frequency and quality factor. The resonance frequency becomes [236]:

\[
\frac{f_i}{f_i^0} = \frac{L^2}{3L_S^2} \left( \frac{\varepsilon_i}{\rho_o} \right) \left( 1 + \frac{L^2 S_0}{4\pi^2 Ef} \right),
\]

where the second term includes \( S_0 \), which is the applied tension at 300 K. For very high strain, the doubly-clamped beam behaves more like string, which means the dynamics are no longer dependent on the Young’s modulus and purely on the tension applied, and the length scaling changes from \( f_i \propto 1/L^2 \) to \( f_i \propto 1/L \) [236]. Intrinsic strain can be a result of the crystal growth process, such as lattice mismatch or thermal expansion variations. For example, CVD diamond is often grown at over 800 °C. Upon cooling, the thermal expansion difference of the substrate (silicon \( \alpha = 2.5 \times 10^{-6} \)) and NCD \( \alpha = 1 \times 10^{-6} \) will result in compressive strain, and can cause a nanobeam to buckle [237]. Metallic structures typically build up tensile strain upon cooling [112]. Alternatively, strain can also be directly applied through mechanical engineering, such as bending the wafer with an adjustable screw [24].

Residual strain has resulted in extraordinarily low dissipation at room temperatures. Values obtained for the quality factor of stressed (1200±50 MPa) silicon nitride have exceeded \( 10^6 \) for a 1 MHz structure [25], even though their surface-to-volume ratio is extremely high. Such structures without strain typically have a quality factor of \( \leq 10^4 \), more than two orders of magnitude lower. Similar structures over a wide frequency range have been limited by thermoelastic damping at low frequencies (≈4–40 MHz) and probably by clamping losses at higher frequencies (≈100–200 MHz) [23]. Without high stress, it is not expected that thermoelastic damping can be measured in submicron diameter structures. Although the effect of intrinsic strain is greatest in doubly-clamped beams, it has also been observed in cantilevers. Similar results have been obtained by epitaxially grown multilayers of GaAs to produce NEMS resonators with high intrinsic strain with reduced dissipation by a factor 10 [238]. Such molecular beam epitaxy growth methods allow for “train-tailoring” to controllably vary the intrinsic strain.

External control and tunable strain has been demonstrated using a screw to bend a wafer that served as the base of a NEMS resonator. This allows for the study of tunable strain and the corresponding effects on the resonance frequency and quality factor. Variation of both the resonance frequency and quality factor of over 100% has been demonstrated in
silicon nitride and single-crystal silicon resonators [24]. Where tensile strain increases the resonance frequency and quality factor, compressive strain can be applied to increase dissipative effects. For structures with low intrinsic strain, the quality factor could be improved by a factor of ≈5–10 by externally applying strain, where beams with intrinsic strain showed little effect when external tensile strain was applied. However, the dissipation increased when compressive strain was applied. Electrically induced strain through capacitative coupling has resulted in comparatively minor tuning capabilities of the resonance frequency and quality factor, due to the much smaller forces that can be applied.

Until recently, a compelling theoretical explanation for the impressive increase of the quality factor due to high tensile strain has been lacking. It had been suggested that increased tension increases the acoustic impedance mismatch between the resonant structure and the base, effectively reducing clamping losses. Initial tests appeared to support this proposal; however, the other effects of strain on intrinsic and surface defects should be studied theoretically and experimentally as well, considering the low dissipation measured in nanowire resonators with high surface-to-volume ratios. Ru [167] demonstrates theoretically how surface stress can reduce thermoelastic dissipation by two-thirds in rectangular and circular cross-sectional nanowires. Similar considerations may apply to bulk stresses as well. More recently, Unterreithmeier et al. [37] proposed a solution to the observed increase in quality factor for high-tensile beams based on Zener’s model, for comparison to the strain-free case). Depending on the beam amplitude and the intrinsic strain, either the bending energy or strain energy will dominate. Therefore, the elongation term due to \( \sigma \) is still dropped; however, the elongation term proportional to \( \sigma_0 \) is kept (see Eq. (6) for comparison to the strain-free case). Depending on the beam amplitude and the intrinsic strain, either the bending energy or strain energy will dominate. To further investigate the effect of strain on dissipation, we will assume the resonator is a doubly-clamped beam actuated at the fundamental frequency. Using the definition of the quality factor \( Q = 2\pi W_0/\Delta W \) one quickly can see that \( \Delta W \) is not affected by strain, as \( E_1 \) is strain independent [239]. One obtains

\[
W_{\delta V} = \delta V \left( \sigma_0 \epsilon + \frac{1}{2} E \epsilon^2 \right),
\]

and then plugging in \( \epsilon \) and integrating over the width and thickness, one obtains

\[
W(\sigma_0) = \int_1 \left[ \frac{wt_0 \sigma_0}{2} (\partial X)^2 + \frac{E w t^3}{24} (\partial^2 X)^2 \right] dx,
\]

where now the elongation term due to \( X \) is still dropped; however, the elongation term proportional to \( \sigma_0 \) is kept (see Eq. (6) for comparison to the strain-free case). Depending on the beam amplitude and the intrinsic strain, either the bending energy or strain energy will dominate. To further investigate the effect of strain on dissipation, we will assume the resonator is a doubly-clamped beam actuated at the fundamental frequency. Using the definition of the quality factor \( Q = 2\pi W_0/\Delta W \) one quickly can see that \( \Delta W \) is not affected by strain, as \( E_1 \) is strain independent. One obtains

\[
Q_{\text{strain}} = Q_0 + Q(\sigma_0)
\]

\[
= Q_0 + 2\pi \frac{W(\sigma_0)}{\Delta W}
\]

\[
= E \frac{12 \sigma_0}{E_1 t^2} \int \left( \partial X \right)^2 dx
\]

\[
= E \frac{12 \sigma_0}{E_1 t^2} \int \left( \partial X \right)^2 dx \]

\[
= E \frac{12 \sigma_0}{E_1 t^2} \int \left( \partial X \right)^2 dx
\]

where the last term is the expression for the fundamental mode of a doubly-clamped beam and \( I = \frac{wt^3}{12} \) is the second moment of inertia. As can be seen the relative ratio of the Young’s modulus \( E \) and \( \frac{12 \sigma_0}{E_1 t^2} \) will determine what dominates the quality factor. Using \( \sigma_0 = \frac{S_0}{L} \) we can rewrite Eq. (120) in terms of intrinsic tension for comparison with the expression for the frequency of a beam under tension (Eq. (122))

\[
Q_{\text{strain}} = \frac{E}{E_1} + \frac{1.774 L^2 S_0}{4\pi^2 E_1 I}.
\]
Fig. 27. Example of tensile strain in nanostring resonators. Tensile strain can be tuned resulting in up to a magnitude increase in the quality factor of an initially low strain resonator. As tension is applied, the quality factor increases by almost a factor seven and the resonance frequency more than doubles. Color traces are added by the authors to compare experiment with linear (red) and quadratic (blue) dependence on the quality factor with regard to strained resonance frequency. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.) Source: Reprinted with permission from [24]. © 2007, American Chemical Society.

Fig. 28. Frequency–temperature relation for a graphene resonator. (IOP Publishing. Reproduced by permission of IOP Publishing. All rights reserved.) Source: Graphic from Ref. [232].

graphic of a thermally strained graphene resonator and the corresponding frequency dependency on temperature. Singh et al. [232] use the known thermal expansion coefficients of the silicon and silicon dioxide base along with the gold clamps to extrapolate the thermal expansion coefficients from the frequency dispersion relation. The resulting values for $\alpha_{\text{graphene}}$ are relatively close to theoretical predictions by Mounet and Marzari [240], where the uncertainty is relatively large due to the uncertainty in the clamping widths, graphene length, and Young’s modulus. Before contemplating the implications of this complicated setup, one may ask what the simplest case predicts. For simplicity, let us assume the initial temperature is 300 K and the only intrinsic strain is due to the thermal expansion mismatch of the resonator and the substrate upon which it is clamped. The thermal tension becomes

$$S(T) = wtE \epsilon = wtE \int_{300\text{ K}}^{T} (\alpha_{\text{res}}(t) - \alpha_{\text{base}}(t)) dt = wtE \Delta \alpha \Delta T,$$

(123)

where $t$ and $w$ are the thickness and width, $E$ the Young’s modulus of the resonator, and $\alpha_i$ are the expansion coefficients, which may be temperature dependent. The last expression is true for a temperature-independent expansion coefficient (or small changes in temperature). The dissipation scales with tension and the tension is strongly temperature-dependent due to the thermal expansion mismatches of the numerous materials making up the resonator, clamps, and base of the entire
Fig. 29. Analyzing strain effects in graphene. (a) Thermal expansion coefficients vs. temperature. (b) Thermal tension vs. temperature for various \( \alpha \) parameters, \( S(T) = w t E / \alpha dT \) with \( t = 0.3 \text{ nm} \) \( w = 0.4 \mu \text{m} \) and \( E = 10^{12} \). (c) Dissipation vs. temperature according to Eq. (122); \( L = 1.43 \mu \text{m} \), and \( E_I = 8 \times 10^{12} \) is chosen so that the dissipation approximately matches that measured for graphene at 300 K (see Fig. 26). (d) Resonance frequency vs. temperature according to Eq. (117). Here, \( S_0 = 2.5 \times 10^{-6} \) is a free parameter chosen so that the resonance frequency is approximately that given in Ref. [232].

System. For complex structures, one can calculate the resulting tension (as done in the supplementary material of Ref. [232]) and plug that into Eq. (122). For simpler structures, where all that matters is the difference of the thermal expansions of the resonator and the base, the quality factor becomes

\[
Q(T) = \frac{E}{E_I} \left( 1 + \frac{0.539 L^2}{E w t^2} S_0 + \frac{0.539 L^2}{t^2} \int_{300 \text{ K}}^{T} \Delta \alpha dt \right). \tag{124}
\]

This last expression indicates that extremely thin structures such as graphene would experience the greatest shifts in quality factor (and frequency) due to thermal strain.

Fig. 29 depicts the thermal expansion coefficients, thermal tension, and resulting quality factors and frequency. It is assumed that there is temperature-independent tension and that the thermal tension is zero at 300 K. Where the absolute value of dissipation and frequency is strongly dependent on the initial tension at 300 K, the general shape will remain the same. It can be seen that high initial tension reduces the effect of thermal tension. (This is analogous to the way the effect of tension introduced by electrical means (\( V_G \) in Ref. [232]) is minimized at low temperature where thermal strain dominates.) Whereas the power laws illustrated in Fig. 26 are not recovered, one does observe a strong temperature dependency at temperatures ranging between 100 and 300 K; below 100 K the dependency is strongly reduced. This result is unaffected by small variations in \( \alpha \) and the resulting thermal tension.

The results indicate that thermal strain can cause significant increases in the quality factor, where, as one may expect, the higher the intrinsic strain is the less the thermal strain will contribute to further changes. What is striking is the relatively small impact variations in \( \alpha \) has on the resulting dissipation. As \( \alpha \) typically approaches zero for low temperatures, it can safely be assumed that the dissipation due to TLS discussed above is affected by this mechanism. What one observes is a very strong temperature dependency on dissipation around 300 K, much higher than that observed in graphene (see Fig. 26). At low temperatures, for \( T < 100 \text{ K} \) the dissipation becomes essentially temperature independent, whereas experiments show a \( Q^{-1} \propto T^\frac{5}{2} \) power law. One may conclude that, while thermal tension should affect dissipation, there must be other mechanisms at work to reproduce experimental observations. What has not been considered here, and hence needs to be revised, is the temperature dependency of \( E_I \). While it is argued that \( E \) is roughly temperature independent for graphene, this need not be the case for \( E_I \).

7.2. Surface and heat treatment

Surface contamination can reduce the quality factor of a NEMS resonator. For silicon resonators, this manifests itself through oxidation of the surface layer. This can be reduced by hydrogen termination using an ammonium fluoride
treatment [173] and then storing and operating the resonator in ultrahigh vacuum. It has also been observed that over time the quality factor of a resonator may degrade; however, high-temperature annealing, even in air, can restore the resonator to its initial condition [24]. As expected, thin structures are more sensitive to surface degradation and treatment than thicker structures. Diamond resonators with a room temperature quality factor of over $10^6$ have been demonstrated where a strong dependency on thickness and surface treatment is observed, where oxygen passivation results in the lowest surface dissipation [206]. Yang et al. [172] demonstrated how annealing 170 nm thick silicon cantilevers at 1000 °C for only 30 s in ultrahigh vacuum increases the quality factor by an order of magnitude by de-oxidizing the surface. The effect becomes weaker for shorter beams where clamping losses most likely contribute to dissipation along with surface losses. Wang et al. [174] shows the effect of different alkenes and hydrogen surface termination on silicon resonators. Optimizing surface treatment not only increases device performance but also reduced degradation with time. Also, when using NEMS as molecular detectors, as has been proposed for chemical and biological applications, surface functionalization will affect dissipative mechanism and therefore must be taken into account.

7.3. Dissipation control circuits

In a parametric resonator, a parameter such as the spring stiffness is modulated at twice the natural resonance frequency. This alone will not cause resonant motion; however, when tuned to a specific phase in addition to a resonant drive, a gain in the response amplitude is observed. The gain factor is phase sensitive and can range from orders of magnitude to less than one as the phase sweeps from zero to $\pi/2$. Rugar et al. [241] demonstrated parametric amplification and thermal noise squeezing in a MEMS resonator by capacitively modulating the spring constant at twice the resonance frequency of a piezoelectrically driven cantilever. The expression for the amplitude gain due to parametric drive takes the form

$$G = \sqrt{\left(\frac{\cos \phi}{1 + Q \Delta k/2k_0}\right)^2 + \left(\frac{\sin \phi}{1 - Q \Delta k/2k_0}\right)^2},$$

where $\Delta k$ is the amplitude of the modulated spring constant. Since then, parametric amplification has been observed in higher frequency structures where mechanical strain is used to modulate the effective spring stiffness in silicon resonators [242,48,243] as is illustrated in Fig. 30. For a purely linear system, the parametric amplification has no upper bound, as $\Delta k \to \frac{2k_0}{Q}$; for real samples, nonlinearities limit the maximum possible gain. By tuning the pumping phase $\phi$ to zero deamplification can be achieved for the same pumping power as resulted in the maximum gain, with a minimum gain of 0.5 (−3 dB). Where parametric amplification does not reduce dissipation, it can amplify signal sizes and narrows the resonance

Fig. 30. Example of parametric amplification. (a) Gain for phase dependent parametric amplification and deamplification. (b) Gain as a function of phase. Source: Reprinted figure with permission from Ref. [241]. © 1991, by the American Physical Society.
trace effectively increasing the sensitivity of the device [242]. The parametric response is no longer a Lorentzian and hence the definition of the quality factor from the linear response becomes less rigorous. However, comparing the full width at half maximum of the modulated and unmodulated responses can be used as an effective measure of the Q enhancement, where the signal amplification factor will not match the reduction in dissipation. This technical trick can be extremely useful when amplifying tiny signals above the noise floor or counteracting the damping effects that are present in a viscous medium. Sekaric et al. [140] successfully demonstrate how light-induced parametric amplification can enhance the quality factor of a resonant structure in vacuum from 3800 to 17,000 and obtain a quality factor in air of over 1000. Comparable resonators without parametric amplification would exhibit strong air damping and a quality factor on the order of 100. This order of magnitude improvement significantly enhances the chemical or biological detection capability of NEMS.

Another method to counter the effect of dissipation is by intrinsically driving the resonator. This is in many ways equivalent to that of a closed loop oscillator, however in piezoresistive materials the change in ohmic heating resulting from the change in resistance can drive the resonator. This replaces some of the energy lost though dissipation and acts somewhat like a high Q resonator [244], where the quality factor is pulled from 2000 to over 2 million by tuning the bias current.

### 8. Nonlinear dissipation

So far, all dissipation mechanisms considered are a result of linear dissipation, which appears as a damping term proportional to the velocity of the resonator as described in Eq. (24). Some of the interest in NEMS devices, however, originates in their nonlinear behavior. Due to their small sizes, very little power is needed to obtain a nonlinear response that differs starkly from the Lorentzian form considered so far. The simplest way to add nonlinearity is given by the Duffing equation which expands Eq. (24) to a higher order in \( x \), resulting in

\[
m\ddot{x} + m\gamma \dot{x} + kx + k_3x^3 = 0,
\]

where \( k_3 \) is the nonlinear coefficient and no driving is included. An \( x^2 \) term may be included but the \( x^3 \) term is required to ensure that the energy is bounded. A positive \( k_3 \) will make the structure stiffer and results in maximum amplitudes at higher frequencies. Correspondingly, a negative \( k_3 \) results in beam softening and lowers the frequency of the maximum displacement. After a critical drive force, the beam displacement bifurcates, resulting in a parameter space where the total displacement depends on the history of the resonator. This can be observed by changing the sweep direction of the drive frequency. This behavior is easily observable in NEMS resonators and has lead to the study of controlled switching between stable states [55,31], stochastic resonance [245–247], and control of nonlinearity in resonator devices. In some cases, it is needed to further expand the Duffing equation to include higher order terms such as \( k_5x^5 \) to describe the observed resonator response [248]. Applications of nonlinear resonators are numerous and include signal processing [32], improved sensing sensitivity [249], resonator energy measurement through nonlinear mode coupling [250], and energy harvesting [251,252] (It should be noted that so far, energy harvesting has only been demonstrated in macroscopic systems at much lower frequencies.), just to name a few. A review of mechanical non-linear dynamics is given by Rhoads et al. [253].

The origins of the \( k_3 \) nonlinearities are beautifully explained by Lifshitz and Cross [254]. Intrinsic nonlinearities of the resonator material exist, but typically contribute only at displacements far greater than nonlinear effects due to the geometry and external potentials. An example of nonlinearities due to an external potential can be illustrated by capacitive driving and-sensing methods described in Section 3.1.2. Eq. (51) that describes the drive force results from an expansion voltage acting between the beam and the gate that forms the capacitor. Simply including higher order terms in \( x \) will result in terms not considered above proportional to \( x^2 \) and \( x^3 \), resulting in an explicit term for \( k_3 \) and \( k_5 \) where applicable. As illustrated in Ref. [254] geometric nonlinearities occur when the thin-beam approximation is no longer valid. Typically, the thin-beam approximation holds when \( L \gg t, w \). However if the displacement of a high aspect-ratio structure becomes large compared to its width (or thickness, depending on oscillation mode), the equation of motion becomes nonlinear. As a perturbative expansion, of a doubly-clamped rectangular beam, it can be shown that the nonlinearity coefficient takes the form

\[
k_3 = \omega_n^3 \frac{S}{2I} \beta_n,
\]

where \( \omega_n \) is the \( n \)-th mode eigenfrequency, \( S \) is the cross-sectional area of the beam and \( I \) the second moment of inertia. \( \beta_n \) is a mode-dependent numerical constant, where \( \beta_1 \approx 0.2 \) and approaches \( \frac{1}{3} \) as \( n \to \infty \). Typical displacements are small < 10 nm compared to device widths that are usually greater than 100 nm, so with the exception of graphene and CNT devices, nonlinearities due to external potentials will often dominate over geometric nonlinearities.

Following Lifshitz and Cross [254] we may consider what happens when one includes nonlinear damping terms. One may consider \( x_1^2 \), \( xx_2 \), \( x_2^2 \) which are all on the same order as \( x_3 \). The following is known as the van der Pol–Duffing equation:

\[
m\ddot{x} + m\gamma \dot{x} + m\alpha_0^2x + k_3x^3 + \eta x^2 \dot{x} = F \cos(\omega t),
\]

where \( \eta \) is the nonlinear damping coefficient and \( F \) the drive force at frequency \( \omega \). Just like the standard Duffing equation, an approximate solution can be obtained using secular perturbation theory, valid in the limit of low dissipation and small oscillations. In addition, it is assumed that the response is only large close to the resonance frequency and hence the perturbation is conducted close to the resonance. The details of this method are illustrated clearly in Ref. [254], so only
the results are presented here. The nonlinear oscillator displacement amplitude becomes

\[ x_0 = \frac{\left( F_{\text{drive}} \right)^2}{\left( \frac{\omega_0 - \omega}{\omega_0} - \frac{3}{8} \frac{k_3}{m\omega_0^2} x_0^2 \right)^2 + \left( \frac{1}{2} Q^{-1} + \frac{1}{8} \frac{\eta}{m\omega_0} x_0^2 \right)^2}. \]  

(129)

In contrast to the linear response, the maximum displacement is now shifted from the resonance frequency, where the frequency shift is dependent on \( x^2 \). This results in the backbone expression given by

\[ \omega_{\text{max}} = \omega_0 + \frac{3}{8} \frac{k_3}{m\omega_0} (x_0)^2_{\text{max}}. \]  

(130)

For increasing drive force, Eq. (129) eventually bifurcates resulting in a frequency space that is multivalued in \( x_0 \) given by

\[ \omega_{\text{SN}}^\pm = \omega_{\text{max}} \pm \sqrt{\frac{3}{16} \left( 3 - \left( \frac{\eta \omega_0}{k_3} \right)^2 \right) \left( \frac{k_3}{m\gamma \omega_0} \right)^2 - \frac{\eta \omega_0}{k_3} x_0^2 + \frac{k_3}{m\gamma \omega_0}}. \]  

(131)

where if the value is real, there exist two stable solutions, as depicted below. From this equation, the critical amplitude (and corresponding force and frequency) can be determined, defining the onset of the bifurcation

\[ x_0^2_{\text{SN}} = \frac{8 m\gamma \omega_0}{3} \frac{1}{\sqrt{3 - \eta \omega_0/k_3}}. \]  

(132)

For \( \eta = 0 \), one sees that this expression is inversely proportional to \( k_3 \), meaning that as \( k_3 \) approaches zero (i.e., approaches the linear regime), the amplitude will diverge before bifurcation occurs, which is the expected result. It can also be seen that as the nonlinear damping increases, the critical displacement increases as well and even diverges as \( \eta \rightarrow \sqrt{3}k_3/\omega_0 \), above which no bifurcation occurs. This is a fundamental change compared to the case without nonlinear damping, where bifurcation will eventually always occur.

Experimentally, there are multiple ways the effect of nonlinear damping may be observable. For low drive forces, where bifurcation is not observed and the response is essentially Lorentzian, one may observe a widening of the resonance given by

\[ \Delta \omega = \gamma + \frac{1}{4} \omega_0 \eta x_0^2. \]  

(133)

and corresponding a displacement dependent dissipation of the form

\[ Q^{-1} \approx \frac{\gamma}{\omega_0} + \frac{1}{4} \eta x_0^2. \]  

(134)

can be written.

As both the errors in the dissipation measurement and NEMS linear displacements are very small, this would typically be a small effect that is relatively hard to observe. As proposed in Ref. [254] the effect of \( \eta \) can be observed in the high drive regime where the response bifurcates. It is shown that for \( \eta = 0 \), the normalized response (where the amplitude is divided by the drive force) is independent of the drive force. For \( \eta > 0 \), however, the normalized response drops with increasing drive force, as depicted in Fig. 31. This means that the maximum amplitude response is not affected by \( k_3 \) and Hooke’s law still applies as long as the maximum displacement is considered and not the displacement at the linear resonance frequency. Correspondingly, if nonlinear damping is present, a displacement–force trace will no longer fall on a straight line but fall off at higher drive forces. It is interesting to note that choosing \( x^2 \) or \( \dot{x}^2 \) instead of \( x^2 \dot{x} \) does not change the derived results. In the secular perturbation theory approach, the other terms result in different effective expressions of \( \eta \), but the general response is unchanged, as is demonstrated in Ref. [254], where all possible damping and displacement terms are included in the perturbative expansion up to order 3. This ambiguity is unhelpful, considering that the physical mechanisms that actually cause nonlinear damping are not known, and in the derivation so far, none have been proposed.

Nonlinear effects have been observed in NEMS devices [255], and it is proposed that nonlinearities due to geometry may also result in nonlinear dissipation [256]; however, such mechanisms are still unknown. While smaller resonators often are easier to drive, nonlinear larger structures have also shown frequency broadening with respect to increased strain [257]. These effects are considerable, but again, no cause is proposed. It is found by Suh et al. [258] that parametric amplification can be limited by nonlinear damping, as was observed in a qubit-coupled nanoresonator. In addition, nonlinear dissipation must be considered in active feedback systems, where out-of-phase feedback will result in nonlinear dissipation [259], as well as noise squeezing [260]. Strong nonlinear damping has recently been observed in CNT and graphene resonators under tensile strain and string like structures [261]. In these structures, nonlinear damping is responsible for destroying the hysteresis. The frequency widening is greater than the frequency shift and is analyzed in the zero linear dissipation limit where

\[ \Delta f \propto \frac{\eta}{m} \left( \frac{F_{\text{drive}}}{f_0} \right)^2. \]  

(135)
While again no particular damping mechanism is discovered, it is suggested that linear damping mechanisms coupled to geometric nonlinearities or external mechanisms such as clamping losses may contribute to the effect. The relatively high nonlinear damping may also point to a mechanism specific to graphene and CNT such as sliding of the carbon resonator over its metal electrodes. Croy et al. [262] present a model where tuning the strain using both bias dc and ac voltages can shift a graphene resonator in or out of the non-linear dissipation regime. They observe the non-linear dissipation as a result of coupling of the flexural modes and in-plane phonons. More research is clearly needed to determine what parameters affect the contribution of $\eta$ such as device aspect ratio, clamping strength, temperature, and more.

The Fig. 32 depicts a power-sweep of a magnetomotively driven doubly-clamped diamond resonator (top) and a capacitively driven and detected resonator at room temperature (bottom). The experimental details the same as described in Section 5.4.3. The resonator has been cooled to 40 mK and shows strong nonlinearity and nonlinear dissipation. Ideally, a fit would be needed only once, after which all traces should be reproduced by Eq. (129), providing the relative actuation force is known (which it is). In reality, the fits proved very difficult and even these imperfect results were only obtained by allowing for different values for $k_3$ and $\eta$ for each trace. Also, the error bars are very large, hence they should only be considered as a guide rather than exact values. It is interesting to note that in frequency space $x^2 \dot{x}$ becomes $\omega^2 x^3$, and in this case, $\omega \eta$ is of the same order as $k_3$. This is important, as it justified that the nonlinear and nonlinear damping terms from Eq. (128) are treated as the same order in the perturbative method that was used to obtain the displacement expression given by Eq. (129). These results dramatically show the effects of nonlinear damping very similarly as described by the theory presented in Fig. 31. Why more precise fitting was not possible may indicate that further mechanisms that are not accounted for are at play. For comparison, the lower trace in Fig. 32 shows results for a capacitively driven and detected beam at room temperature. Even though the response is highly nonlinear, there is no measurable nonlinear damping (as the normalized maximum amplitude is constant and no peak broadening is observed within experimental errors). One obvious difference is a quality factor that is much higher at low temperatures, resulting in larger amplitudes for the equivalent drive force. The drive powers cannot be compared directly as the transduction efficiency of the magnetomotive and electrostatic methods are very different. In principle, only the absolute amplitudes should factor in.

![Fig. 31. Nonlinear dissipation. (a) Normalized response of Duffing equation without nonlinear damping. Colors indicate different drive forces. (b) Same as above but with nonlinear damping. Source: Reprinted with permission from [89]. © 2013, AIP Publishing LLC.](image-url)
Recent data illustrates the transition from linear to nonlinear damping in diamond resonators with changing temperature. The data depicted in Fig. 33 illustrates nonlinear damping appearing in an initially linear system as the temperature drops from 300 to 77 K. Here the maximum amplitude is plotted vs. drive force. Although the quality factor changes it is not sufficient to explain the deviation from Hooke’s Law. Fig. 33 depicts Hooke’s law adapted for linear damping only, nonlinear damping only, and both linear and nonlinear damping combined. Starting with Eq. (129) one can express the maximum amplitude as a function of drive force:

$$x_{\text{max}} = \sqrt{\frac{32}{3} \frac{Q^{-1} k_0}{F_{\text{dr}}} \frac{1}{\sqrt{9 F_{\text{dr}} \eta^2 \omega_0^2 + \sqrt{3} \sqrt{27 F_{\text{dr}}^2 \eta^4 \omega_0^4 + 16 Q^{-12} \eta^7 \omega_0^4 k_0^2}}} + \sqrt{\frac{2}{9 \eta \omega_0 F_{\text{dr}}} \frac{1}{\sqrt{9 F_{\text{dr}} \eta^2 \omega_0^2 + \sqrt{3} \sqrt{27 F_{\text{dr}}^2 \eta^4 \omega_0^4 + 16 Q^{-12} \eta^7 \omega_0^4 k_0^2}}} \right)$$

$$\Rightarrow \frac{2}{\eta \omega_0} \left( F_{\text{dr}} \eta^2 \omega_0^2 + \sqrt{F_{\text{dr}}^2 \eta^4 \omega_0^4} \right)$$

(136)

where the last expression is the limit for $Q^{-1} \rightarrow 0$. This expression is analogous to the backbone curve used to characterize the nonlinear component of the spring constant. This dissipation backbone curve can be used to characterize the strength of the nonlinear dissipation which results in a deviation from Hooke’s law. The backbone curve for the nonlinear spring constant allows one to determine the strength of the nonlinearity by measuring the change in frequency with regard to drive amplitude or forcing (see Eq. (130)). Analogously the nonlinear dissipation becomes apparent by the deviation from Hooke’s law, that is measuring the normalized amplitude with respect to drive forcing. This would be a constant for a sys-
tem without nonlinear dissipation, even if the spring constant is strongly nonlinear. The conventional method for measuring nonlinear dissipation though the widening of the FWHM is problematic for systems with strong nonlinear spring constants. Both backbone curves decouple the problem, making it more easily quantifiable.

Although we do not have sufficient data to make confident statements, measurements have indicated that out-of-plane modes typically show more nonlinear damping than in-plane modes of equivalent structures. This is something that would be relatively easy to verify magnetomotively, where a magnetic field can be applied at an angle that would actuate both in-plane and out-of-plane flexural modes. König et al. [205] have observed strain-dependent dissipation and a shift in sound velocity that can be explained by coupling of TLS. This mechanism is a good candidate as a source of nonlinear dissipation, as TLS are known to be significant in both the diamond structure presented above and are likely also to occur in the graphene and CNT resonators that have exhibited nonlinear dissipation [261].

Further systematic studies would give hints to the origins of nonlinear damping and narrow down the parameter space where such effects must be taken into account.

9. Concluding remarks

NEMS typically suffer from considerably higher dissipation than their larger counterparts in the macroscopic world, counteracting some of the sensitivity gained through miniaturization. Linear dissipation of doubly and singly clamped NEMS resonators is discussed in detail. Dominant sources typically encountered are categorized in terms of extrinsic and intrinsic dissipation mechanisms. The relevant parameter space for given resonator sizes and frequencies is investigated, allowing for the identification of dominant dissipation sources. This should provide NEMS users with a framework to understand, account for, and at times evade unwanted losses in resonant structures.

Dissipation is a fundamental characteristic of MEMS and NEMS devices as dissipation (inverse quality factor, 1/Q) is the predominant factor in determining the ultimate device performance in many applications. For instance, MEMS- and NEMS-resonator based timing oscillators use the resonator as the frequency source. Intrinsic dissipation (often quantified by an unloaded quality factor) ultimately limits the most fundamental characteristics of the timing oscillator: phase noise and jitter. As the resonator is shrunk in size for increased normal mode frequency, considerations of size effects, discussed at length in this review, become even more relevant. Timing oscillators used in applications such as GPS receivers have

Fig. 33. Nonlinear dissipation backbone curve. (a) Normalized amplitude vs. drive force at various temperatures. (b) Low temperature data and nonlinear backbone curves for only linear dissipation, only nonlinear dissipation and both dissipation mechanisms combined. Source: Reprinted with permission from [263]. © 2013, American Chemical Society.
stringent requirements on close-in phase noise (1–100 Hz offset from the carrier), primarily determined by defect-induced 1/f noise, which can be improved by reducing or passivating the volume and surface defects.

Front-end bandpass filters are used in many wireless communication applications. Currently, these filters use BAW (Bulk Acoustic Wave) and SAW (Surface Acoustic Wave) resonators for band pass around a precise frequency with a well-defined bandwidth. The most important characteristic of front-end filters is insertion loss, which quantifies the amount of power lost within the pass band. The insertion loss is directly related to the quality factor or dissipation in the resonator. Following the push towards reducing the device footprint, reduction in resonator size from microns to hundreds or nanometers or less is highly desired. As the resonator size is reduced, quality factor is also reduced, which makes the arbitrary reduction of the device size challenging.

Some motion sensors such as accelerometers and gyroscopes also employ resonators for detecting the directional movement of inertial masses. Resonant-mode sensing approach involves motion detection at a normal mode frequency as the inertial mass, a part of the resonator itself, moves. The signal size on resonance depends on the quality factor; hence dissipation in the resonator can be the limiting factor in such devices. Cantilever-based resonators are often used for the detection of chemical or biological entities either in air or fluid. Though these approaches have been in use for a long time in atomic force microscopy, nanoscale counterparts are being studied for potential use as biological and chemical sensors. Other variations include mass sensors and magnetic resonance force microscopy. In all these applications, higher quality factor or lower dissipation is highly desired. Usually, the resonator size is reduced in some of these applications for probing smaller length scale or achieving higher sensitivity, where dissipation becomes the predominant factor determining the ultimate device performance. Other offbeat applications of micro- and nanoscale resonators include thermal noise thermometry where signal size on resonance varies as a parameter free function of temperature [264]. Once again, the on-resonance measurements benefit from higher quality factor and lower dissipation.

Examples discussed above have been explored in depth for resonators or systems on micron scale for a long time. However, miniaturization of the underlying resonator in all these applications has been challenging, even though there are obvious benefits of reduced length scale such as high speed operation, higher density and better integration. The fundamental reason behind these challenges often relates to higher dissipation at smaller length scales. The study of this problem of increased dissipation at reduced sizes needs to be performed in depth, as so many dissipation mechanisms are at play for any given resonator. Enhancement of surface-to-volume ratio makes surface effects more prominent, but the relative contribution of various dissipation mechanisms and their dependence on relevant length scales must be taken into account for a given resonant structure. As resonator size is further reduced, clamping losses become detrimental but can be suppressed with appropriate device design. Also, as the surface-to-volume ratio grows, inevitable surface defects contribute significantly to losses, thereby limiting device sensitivity. Surface treatments such as annealing and introducing tensile strain can significantly improve device performance. Jensen, Peng and Zettl propose an upper limit performance device engineering idealized clamps and a defect free surface to obtain high Q and high frequency resonators [265]. With further control of dissipation, future NEMS devices will become even smaller, faster, and more sensitive enabling new technologies.

Apart from scaling, the aspect of dissipation that is hardly studied involves the dynamics of dissipation itself—i.e. how energy is lost from a resonator into an environment. Anchors in nanomechanical resonators are becoming the focus of both experimental and theoretical studies, as the proper design of anchors can help control the flow of energy out of the resonator mode into the environment or base. Extensive studies have been done on the wave mechanics of acoustic and heat wave propagation through anchors [118]. Another focus is on the design of soft anchors, which decouple the resonator mode, and hence reduce the energy loss. There has been a lot of recent works on the concept of phononic crystal anchors [266], where the anchors are designed to contain phononic crystal bandgaps due to a material or structural periodicity. These phononic crystal structures, for instance the periodic structure of air holes, are being made on nanoscale [267]. Their contribution to dissipation dynamics and the behavior of the resonator performance are now an active field of study [268,269].

An important topic of dissipation, discussed here at length, is the concept of nonlinear dissipation. NEMS devices are explored more and more in the nonlinear regime. Examples include sensing [245,253,250] or for enhanced performance in timing oscillators [270,271], and nanomechanical computation [246]. Nonlinear dissipation sets a limit to the efficiency of dissipation evading techniques such as parametric amplification and active feedback circuits. Emergence of nonlinear dissipation mechanisms will become more important in determining the device performance and discerning the dominant contribution to dissipation in these nonlinear devices.

Acknowledgments

We thank our group members Alexei Gaidarzy, Guiti Zolfagharkhani, Josef-Stefan Wenzler, Tyler Dunn and Diego Guerra, as well as Cesar Seoanez, Eduardo Lopez, Antonio Castro Neto and Francisco Guinea for their help and support. Special thanks to G. Zolfagharkhani et al., S-B Shim et al. and V. Singh et al. for making their data available to us. This work was supported by NSF.

References


G. Bycroft, Forced vibrations of a rigid circular plate on a semi-infinite elastic space and on an elastic stratum (Philosophical Transactions of the Royal Society of London. Series A), Mathematical and Physical Sciences 246 (948) (1956) 327–368.


A.Rahafrooz, S.Pourkamali, Activeself-q


