

Metrology, electronics, physics, etc.

Jérôme Juillard GEEPS – DEPARTEMENT SYSTEMES ELECTRONIQUES

# Table of Contents

Preamble5			
1 Fundamentals		7	
1.1 Concepts o	of metrology	7	
1.2 Constituen	nts of a sensor	10	
1.3 Error budg	get	12	
1.4 Transduce	ers	13	
1.4.1 Passive	transducers	13	
1.4.2 Active tr	ransducers		
1.5 References	S	17	
2 Conditioning ele	ectronics		
2.1 Bridges an	nd potentiometric readouts		
2.1.1 Potentic	ometric readouts	20	
2.1.2 Wheatst	tone bridge	21	
2.2 Analog Fro	ont-Ends	22	
2.2.1 Current	-output transducers	23	
2.2.2 Voltage-	-output transducers	23	
2.3 Non-ideali	ities of electronics	24	
2.3.1 Passive	components	24	
2.3.2 Active c	components	26	
2.3.3 Intercor	nnects	32	
2.4 References	S		
3 MEMS acceleror	meters and gyroscopes	35	
3.1 MEMS acce	elerometers	35	
3.1.1 Sensitiv	rity		
3.1.2 Noise			
3.1.3 Dynamie	c effects in accelerometers		
3.2 MEMS gyro	oscopes	41	
3.2.1 Sensitivi	rity	42	
3.2.2 Noise		44	
3.2.3 Mode-m	natched gyroscopes	45	
3.2.4 Quadrat	ture error	45	
3.3 References	S	46	
4 Physics of MEM	IS		
4.1 Introduction	on	48	

	4.2	Mechanics of MEMS	
	4.2.1	Fundamental principles	
	4.2.2	Beams and suspensions	50
	4.2.3	Plates and membranes	60
	4.3	Electromechanical transduction	61
	4.3.1	General overview	61
	4.3.2	Detection	62
	4.3.3	Actuation	68
	4.4	Dissipation	
	4.4.1	General overview	
	4.4.2	Fluidic losses	79
	4.4.3	Thermoelastic losses	
	4.4.4	Anchor losses	85
	4.4.5	Electrical losses	
	4.4.6	Other losses	
	4.5	References	
5	MEM	S design and modeling	90
	5.1	Introduction	90
	5.2	MEMS design flow	91
	5.2.1	Geometry selection	91
	5.2.2	High-level modeling	91
	5.2.3	Optimization	92
	5.2.4	Low-level modeling	92
	5.3	Co-simulation	93
	5.3.1	Equivalent electrical models	93
	5.3.2	Hardware description languages	93
	5.4	References	93
6	Arou	nd sensors	94
	6.1	Components	94
	6.2	Processing	94
	6.3	Environment	94
7	FM r	esonant sensors and energy harvesters	94
8	Appe	ndices	96
	A - PSD	, variance and Allan variance	96
	B - Froi	n Hooke's law to Euler-Bernoulli	
	B.1 C	erivation of Euler-Bernoulli's equation	

B.2 Boundary conditions	99	
C - From simple beams to complex suspensions		
C.1Torsion beam – Fixed at $x = 0$ , couple at $x = L$		
C.2 Bending Beam – Fixed at $x = 0$ , force at $x = L$		
C.3 Bending Beam – Fixed at $x = 0$ , couple at $x = L$		
C.4 Bending beam –Fixed at $x = 0$ , sliding + force at $x = L$		
C.5 Bending beam – Spring at $x = 0$ , sliding + force at $x = L$		
C.6 Bending beam – Torsional spring at $x = 0$ , sliding + force at $x = L$		
C.7 General model of a beam with end loads		
D - Eigenmodes and model order reduction		
D.1 Galerkin's method		
D.2 Modal analysis of a simple resonator		
D.3 Numerical models and modal analysis		
E - Beyond plane capacitors		
E.1 Analytical approach		
E.2 Electrostatic boundary element method		
F - From Navier-Stokes to squeeze-film damping		
G - From Fourier to thermoelastic damping	117	
H - Manufacturing processes		
I - Summary of poly-Si / mono-Si / Air properties		
J - Labs		
K - Projects		

# Preamble

This document is intended to serve as a reference for students taking the following courses:

- Electronic Systems (**SE**, CentraleSupélec 1<sup>st</sup> Yr)
- Capteurs Intégrés MEMS / Integrated MEMS Sensors (CIMEMS, CentraleSupélec 2<sup>nd</sup> Yr)
- High-Performance Sensors and TRansducers (**HIPSTR**, SCMA mention, CentraleSupélec 3<sup>rd</sup> Yr, and M2 Integration-Circuits-Systems of UPSaclay)
- Physics and Modeling of MEMS (PHYMEMS, M2 of Nanosciences of UPSaclay)

These courses are all related to my own field of research: the modeling and design of MEMS (Microelectromechanical Systems) devices, in particular MEMS-based resonant sensors and oscillators.

Although I dedicate a large part of this paper to MEMS, I try to do so within the framework of a "system" approach, so that the notions and approaches developed here can be transposed to devices based on other types of transducers.

Part 1 of the document is thus devoted to generalities. It is fundamental for all the courses mentioned above:

- Concepts of metrology
- Components of a sensor
- Error budget
- Transducers

Part 2 is devoted to the electronics associated with the sensors. This part, still very general, is of marginal interest for PHYMEMS students:

- Bridge and potentiometric readouts
- Analog Front-Ends
- Non-idealities of electronics

In part 3, I describe the operation of common inertial sensors, with simplified physics:

- MEMS Accelerometers
- MEMS Gyroscopes

Some metrological characteristics of these sensors are determined, to illustrate the previous parts.

Part 4 deals in broad strokes with the physics of MEMS devices:

- Mechanics
- Transduction
- Dissipation

Part 5 is dedicated to their design and modeling:

- Top-down design approach
- Modal analysis and model order reduction
- Numerical simulation techniques

Part 6 is devoted to more advanced considerations on some sensor-specific processing, and on the components found in the sensor environment:

- Lock-in detection technique
- Closed-loop control
- Testing, calibration compensation
- Data fusion
- Sensor environment (packaging, clocks, voltage references, etc.)

Part 7 deals with resonant sensors and MEMS energy harvesters.

## 1 Fundamentals

### 1.1 Concepts of metrology

A sensor is a device that delivers information about the physical environment in which it operates. The "output" of the sensor takes the form of an electrical signal, analog or digital, whose evolution over time must be similar to that of a given physical quantity at the "input" of the sensor (Fig. 1.1).



Fig. 1.1 - A sensor converts the variations of a physical quantity A(t) into the variations of a voltage V(t), according to a certain relationship, valid over a given operating range, with the exception of errors.

A linear (or affine) relationship is expected to exist between the input A(t) and the output V(t) of the sensor, and that this relationship be valid in the sensor **operating range**, i.e. :

- over a certain measuring interval.
- within a certain bandwidth.
- independently of influence quantities that affect the sensor environment.

with the exception of **systematic** or **stochastic errors**, whose typical values are specified in the operating range.

The **measuring interval** is the set of values of A(t) measurable by the device, i.e.  $[A_{min}, A_{max}]$ . A related notion is that of measurement **range** (or dynamic range) corresponding to  $A_{max} - A_{min}$ .

### Examples:

1 - The typical measuring interval of a clinical thermometer is  $[34^{\circ}C, 43^{\circ}C]$ . Some industrial thermometers have measuring ranges greater than  $1000^{\circ}C$ .

2 - A measuring interval of [-100 G, 100 G] (1  $G = 9.81 m/s^2$ ) is typical of a MEMS accelerometer used as a crash sensor, an interval of [-10 G, 10 G] of an accelerometer used in a game controller, an interval of [-1 G, 1 G] an inclinometer.

The **bandwidth** corresponds to the frequency range  $[f_{min}, f_{max}]$  for which the output of the sensor is a "faithful" image of the input. In control-engineering vocabulary, the transfer function between the input and the output of the sensor is "constant" over this frequency band. The bandwidth thus determines the minimum and maximum rates of variation of A(t) that the sensor is able to transcribe. The narrower the bandwidth of a sensor, the longer its **response time**.

#### Examples:

1 - The response time of a mercury fever thermometer is of the order of a few tens of seconds, which corresponds to a bandwidth of a few tens of mHz. Infrared thermometers allow much higher bandwidths (of the order of kHz), and therefore much shorter response times.

2 - A crash sensor must measure much more sudden changes in acceleration than a joystick sensor, so it requires a higher bandwidth.

The output value of the sensor may be affected by physical quantities other than the one being measured. These quantities are called **influence quantities**. For a given sensor application, it is essential to know the main influence quantities, and how these affect the metrological performance of the device.

#### Examples:

1 - The distance between the sensor and the target is an influence quantity for infrared thermometers. For a mercury medical thermometer, it is the contact surface that will be the main influence quantity (on the response time, in particular).

2 - Temperature is the main influence variable of most sensors that are not thermometers. This is particularly true for MEMS inertial sensors (the mechanical properties of the sensing element vary with temperature).

3 - Many sensors deliver a signal proportional to a reference supply voltage, which is then an influence quantity to be taken into account.

These three aspects determine the **operating range** of use of the sensor, within which its behavior is often (but not always) modeled by

$$V(t) = V_{ref} + S \times (A(t) - A_{ref}) + v_{svs} + v_{sto}(t)$$
 (1.1)

where  $A_{ref}$  is the reference value of A(t) (typically the center of the measurement interval) and  $V_{ref}$  the corresponding sensor output voltage, S is the sensor **sensitivity** (in V by "units of A"),  $v_{sys}$  a **systematic error** term (in V) and  $v_{sto}$  a **stochastic error** term ("noise", also V-shaped). In the rest of this section, we will assume without loss of generality that  $V_{ref}$  and  $A_{ref}$  are zero, so the sensor behavior is modeled by

$$V(t) = S \times A(t) + v_{svs} + v_{sto}(t) \qquad (1.2)$$

or by

$$V(t) = S \times \left( A(t) + a_{sys} + a_{sto}(t) \right) \quad (1.3)$$

where  $a_{sys} = v_{sys}/S$  and  $a_{sto} = v_{sto}/S$ .

Once sensitivity has been factored out, the systematic and stochastic errors are expressed in "units of A":  $a_{sys}$  and  $a_{sto}$  are called **input-referred** errors, while  $v_{sys}$  and  $v_{sto}$  are called **output-referred** errors. In order for the sensor to be "accurate" (this term needs to be specified), the input-referred error terms must be small with respect to the sensor's measuring range.

The **stochastic error** is a random quantity: the output of the sensor does not have exactly the same value when a measurement is repeated under a given set of operating conditions (i.e. same input quantity and influence quantities). The stochastic error is commonly characterized by its frequency spectrum, or by its Allan variance (see Appendix A). The **resolution** of the sensor, which is the smallest variation of A(t) that can be measured, is all the better as the stochastic error is small. A sensor with a low stochastic error is said to be "**precise**".

Remark:

It is possible, within certain limits, to improve the resolution of a sensor:

- by multiplying the number of sensors and averaging their outputs at a given time. This "expensive" solution (in terms of cost, size, consumption...) obviously assumes that the sensors are strictly identical and subject to the same stimuli.
- by using a single sensor and averaging measurements taken at successive times. If the error  $v_{sto}(t)$  is a "white" (flat-spectrum) noise, averaging over N successive samples improves the resolution by a factor  $\sqrt{N}$  but reduces the bandwidth (increases response time) by a factor of N. This is the **resolution-bandwidth trade-off**, limited by the "corner frequency" of the noise (below which the stochastic error is no longer white).

For a given sensor, the **systematic error** is a deterministic quantity, in the sense that it is repeatable. Thus, its value does not change if the operating conditions of the sensor do not change. It includes the **bias error**, the **sensitivity error** and **the non-linearity error**. These three quantities characterize the deviation between the ideal linear model

$$V(t) = S \times A(t) \tag{1.4}$$

and the actual behavior of the sensor. The **bias error** (referred to the output) corresponds to the value of *V* read when the input quantity *A* is zero. When referred to the input, it corresponds to the value of *A* which cancels the sensor output. The **sensitivity error** (or scale factor error) results from a deviation between the nominal sensitivity of the sensor and its actual sensitivity. Finally, the **non-linearity error** concerns all deviations of the sensor behavior from an affine characteristic. These three types of errors are illustrated in Fig. 1.2. A sensor with a small systematic error is said to be "**true**".

The notions of "trueness" (related to systematic error) and "precision" (related to stochastic error) are classically illustrated in Fig. 1.3.

Remark:

The ideal sensor model can of course be made more complex, for example by taking into account dynamic effects (the sensitivity is then modeled as a transfer function), or possible non-linearities (we could replace (1.4) by a polynomial model, for example).



Fig. 1.2 - Systematic errors. From left to right: effects of a bias error, a sensitivity error and a nonlinearity error.



Fig. 1.3 - Trueness and precision.

Exercise:

Make the link between the abstract notions presented above and the datasheet of a commercial sensor. From the datasheet of the <u>ADXL 1001</u> accelerometer, for example, identify the measurement range and the bandwidth of the sensor, quantify the impact of influence quantities such as temperature and supply voltage on the sensor characteristics, and determine the resolution of the sensor on a given bandwidth.

### 1.2 Constituents of a sensor

To choose and use a commercial sensor in a given application, it is sufficient to master the concepts presented in the previous section, and to include other elements such as cost constraints, consumption ... To develop a new sensor (more "optimal" than those commercially available), it is necessary to go down to a lower level and focus on the more fundamental components of sensors. The engineer's objective will then be to select and assemble these subsystems in order to meet specifications, metrological in particular. To this end, it will be essential to understand how the non-idealities of each of the subsystems affect the (systematic or stochastic) error of the overall sensor.

First of all, a sensor consists of at least two distinct parts: a **transducer**, and a **conditioning circuit**, which delivers V(t) –these are the two essential blocks of an **analog** sensor.

A **transducer** is a physical device that converts one type of energy (for example: elastic, kinetic, thermal) into another, in our case: electrical, with the information being coded as voltage or current. From the electrical point of view, we can model a transducer as a source of current or voltage, whose intensity varies according to the physical quantity to be measured. The last section of this part is entirely dedicated to examples of common transducers.

### Remarks:

1 - Transduction phenomena are often reversible. They can be implemented within sensors, but also within actuators (piezoelectric positioners, LED displays, Peltier modules).

2 - Transducers are also essential in energy harvesting applications, where the aim is to convert "ambient" energy (vibrations, temperature gradient) into electrical energy.

The main function of the **conditioning circuit** associated with the transducer is to transform the output of the transducer into an electrical voltage. V(t) whose full scale (from  $V_{ss}$  à  $V_{dd}$ ) corresponds to the measuring range of the sensor (from  $A_{min}$  à  $A_{max}$ ). This generally requires **amplification**: the conditioning circuit of a transducer is therefore **active**.

#### Example:

One wishes to use a thermocouple to make a temperature measurement over a wide range (1000 °C). The sensitivity of the thermocouple is  $10 \,\mu$ V/°C. Thus, the output voltage of the transducer only varies from 10 mV when the temperature spans the measuring interval. In order to expand this voltage excursion to 5 V (the supply voltage of many active circuits), the conditioning electronics must be able to amplify the 10 mV voltage deviation by a factor of 500.

Commercially-available sensors include at least these two blocks. **Digital** sensors also include **analog-to-digital converters** whose function is to sample V(t) at regular times and quantize each sample over a certain number of bits.

Finally, a sensor may also include

- blocks, most often digital, allowing integrated information processing (filtering, correction of non-idealities, self-testing, self-calibration, feature extraction).
- digital blocks for information transmission according to predefined protocols (I2C, SPI...).
- power management units, voltage references, clocks, other sensors (to minimize the impact of influence quantities, for example, or merge measurements in one way or another)

### 1.3 Error budget



Fig. 1.4 - Information processing chain in a sensor.

Fig. 1.4 shows the information processing chain inside a sensor in a symbolic way. When designing a sensor, it is generally important to quantify the impact of the imperfections of each part on the overall accuracy of the measurement. For this, we can consider that only the stages of the analog part have an impact on the accuracy of the sensor, whether in terms of trueness or precision. If each of these stages is characterized by an input-referred error and if all dynamic effects are neglected, we have, starting from the A/D converter :

$$V_q = V + v_{when} \tag{1.5}$$

where  $V_q$  is the quantized and sampled voltage, and  $v_{when}$  represents the quantization error (resulting from rounding errors and miscellaneous non-idealities of the converter), then, for example

$$V = G_{cond} \times (v + v_{cond}) \qquad (1.6)$$

where  $G_{cond}$  is the gain of the conditioning electronics, v the output voltage of the transducer and  $v_{cond}$  the error related to the electronics, and finally

$$v(t) = S_{tran} \times (A + a_{tran}) \quad (1.7)$$

where  $S_{tran}$  is the sensitivity of the transducer and  $a_{tran}$  its intrinsic input-referred error. It is useful to write these relations in a more synthetic form :

$$V_q = G_{cond} \times S_{tran} \times \left(A + a_{tran} + \frac{v_{cond}}{S_{tran}} + \frac{v_{when}}{S_{tran}G_{cond}}\right) (1.8)$$

in order to refer all errors to the sensor input. This makes it possible to compare the contributions of each of the elements to the total error (expressed in the dimension of the physical quantity of interest), and thus to choose all the elements of the chain in a reasoned manner.

Remark:

The same reasoning obviously applies when the transducer output is a current or a charge, as in the case of piezoelectric or capacitive transducers. The error reported at the input of the conditioning stage is then expressed in amperes or coulombs.

This approach is also useful to understand an important principle concerning the sensor conditioning electronics, when it requires several amplification stages to achieve the required gain  $G_{cond}$ . For example, suppose we try to obtain a gain  $G_{cond}$  with 2 amplification stages, with gains  $G_{cond1}$  and  $G_{cond2}$ , as shown in Fig. 1.5, and with input-referred errors  $v_{cond1}$  and  $v_{cond2}$ . We can then write

$$V = G_{cond1} \times G_{cond2} \times \left(v + v_{cond1} + \frac{v_{cond2}}{G_{cond1}}\right) (1.9)$$

Thus, the contribution to the total error of the second amplification stage is all the smaller as the gain of the first stage is high. This first amplification stage (known as an **analog front end (AFE**)) therefore plays a crucial role in the sensor's error budget: it is the stage whose design requires the most attention. It is classically dimensioned to have a high gain and a "low" error (i.e. an error which, relative to the transducer input, is small compared to  $a_{tran}$ ).

We will see in part 2 how these high-level notions are translated in terms of electronics.



Fig. 1.5 - Making a gain  $G_{cond} = G_{cond1} \times G_{cond2}$  from non-ideal blocks.

### 1.4 Transducers

A distinction is generally made between two types of transducers :

- **active** transducers, which do not require an external power supply to convert the input variable into an electrical signal.
- **passive** transducers, which require an external power supply.

**Simple** transducers are also distinguished from **composite** transducers, the latter requiring several transduction steps to go from the physical quantity of interest to an electrical quantity.

### Remark:

The distinction between active and passive transducers is universal, but "up to a plus or minus sign" ... Some authors consider as active transducers those requiring an external energy input (such as electronic circuits) and while others consider them as passive.

### 1.4.1 Passive transducers

Simple passive transducers are the most easily "accessible": they are devices similar to electric dipoles whose impedance Z varies according to the physical quantity to be measured. Many of them are simply variable resistors or capacitors. By applying a known voltage across such a dipole, one obtains an electric current whose variations mirror those of Z. By imposing the current that flows through it, one obtains a voltage whose variations mirror those of Z.



Fig. 1.6 - Notations for resistive (left) and capacitive (right) transducers.

Let us for example consider a linear element of a conductor characterized by a resistivity  $\rho$ , a cross-section *S* and a length *l* (Fig. 1.6). Its resistance is given by :

$$R = \rho l/S \tag{1.10}$$

It is apparent that any variation with the input quantity *A* of its dimensions (l, S) or of its resistivity  $\rho$  will result in a variation of *R* according to the formula :

$$\frac{dR}{dA} = \frac{\partial R}{\partial \rho} \times \frac{d\rho}{dA} + \frac{\partial R}{\partial S} \times \frac{dS}{dA} + \frac{\partial R}{\partial l} \times \frac{dl}{dA} = \frac{l}{S} \times \frac{d\rho}{dA} - \frac{\rho l}{S^2} \times \frac{dS}{dA} + \frac{\rho}{S} \times \frac{dl}{dA}$$
(1.11)

or to highlight the relative variation of *R* with *A* :

$$\frac{1}{R}\frac{dR}{dA} = \frac{1}{\rho} \times \frac{d\rho}{dA} - \frac{1}{S} \times \frac{dS}{dA} + \frac{1}{l} \times \frac{dl}{dA}$$
(1.12)

### Examples

1 – An **RTD** (Resistive Temperature Detector) is a typical example of a simple resistive transducer. It is a resistor made of a noble metal (typically platinum) whose resistivity variation with temperature  $\theta$  is perfectly known, and quasi-linear over a very large range (about 1000°C). The coefficient of expansion, which governs dS/d $\theta$  and dl/d $\theta$  also varies in the same manner.

2 - The same physical principle is implemented in **thermistors**, which are ceramics whose relative variation in resistivity with temperature is much greater than that of noble metals. On the other hand, the characteristic  $\rho(\theta)$  is strongly non-linear, which limits the (linear) operating range of these devices.

3 - The other typical example of a simple resistive transducer is the **strain gauge**. A strain gauge is a conductive wire firmly bonded to the surface of a solid whose deformation is to be measured at one point. As the solid deforms, the wire also deforms (its length increases and its cross-section decreases, or its length decreases and its cross-section increases), resulting in a change in the resistance of the wire. The most sensitive strain gages are made of **piezoresistive** materials (presented in more detail in Part 4), whose variation in resistivity with deformation is 1 to 2 orders of magnitude greater than that induced by geometric effects.

4 - Strain gauges are often used in composite transducers, where the physical quantity of interest is first transduced into a deformation, before being transduced into a variation of resistance by the gauge. Thus, many pressure sensors rely on a transducer composed of a membrane separating the pressure to be measured from a reference pressure. Strain gauges are arranged around the circumference of the diaphragm, where the deformation induced by the pressure differential is greatest. In this way, pressure variations are transduced into resistance variations.

### Remark:

All resistive transducers are subject to ohmic losses, which result in heating of the transducer. If not taken care of, the variation in resistance caused by this self-heating can become a major source of error.

Capacitive transduction is also widespread, especially in the MEMS field. For example, the capacitance *C* between two parallel conducting plates with surface area *S* separated by a distance *g* by a medium with permittivity  $\epsilon$  (Fig. 1.6) is equal to

$$C = \epsilon S/g \qquad (1.13)$$

and the relative variation with an input variable *A* is written as follows:

$$\frac{1}{c}\frac{dC}{dA} = \frac{1}{\epsilon} \times \frac{d\epsilon}{dA} - \frac{1}{g} \times \frac{dg}{dA} + \frac{1}{S} \times \frac{dS}{dA}$$
(1.14)

This principle is used to make displacement or position sensors, in particular.

Examples:

1 - Many touch screen technologies are based on a matrix of capacitors whose value is affected according to the proximity of a finger or a stylus. In these cases, the capacitors serve as simple transducers.

2 - In a MEMS **accelerometer**, a test mass *M* (suspended from a "spring") moves under the effect of an inertial force resulting from an acceleration  $\vec{A}$  ( $\vec{F}_{Inertia} = -M\vec{A}$ ). This motion results in a variation in the capacitance between the moving mass and the fixed part of the structure. If a voltage is applied to this capacitor, a "motional" current is generated, from which the acceleration can be deduced.

3 - **Gyroscopes** are devices for measuring angular velocity  $\vec{\Omega}$ . MEMS gyroscopes are based on the conversion of angular velocity into Coriolis force, which requires making a proof mass M oscillate with a certain velocity  $\vec{v}$  ( $\vec{F}_{Coriolis} = -2M\vec{\Omega} \wedge \vec{v}$ ). Under the effect of this force, the mass moves, perpendicularly to  $\vec{\Omega}$  and  $\vec{v}$ . This results in a capacitance variation that can be transduced into a current. Note that in this case, energy is supplied to the system at two points, not only to convert the capacitance variation into an electrical quantity, but also to make the proof mass oscillate. In this respect, all the so-called "resonant" sensors, which require setting a test body in oscillation (mechanical or not) are considered as "passive", even when they are based on active transduction phenomena.

4 - Capacitive pressure sensors are a common alternative to the piezoresistive pressure sensors described above. In this case, a capacitance is formed between the membrane and a fixed opposite electrode. When the membrane deforms, the inter-electrode distance changes, resulting in a change in capacitance.

### Remarks:

1 - Unlike resistive transducers, capacitive transducers are not subject to self-heating. Moreover, since they do not dissipate energy, they are intrinsically compatible with "low power" and "low noise" applications. These reasons, among others, have motivated the progressive shift from piezoresistive inertial sensors (e.g. an accelerometer whose spring elongation is measured with a strain gauge) to capacitive inertial sensors.

2 - Capacitive transduction is a reversible phenomenon: when a voltage is applied between two armatures of a capacitor, they tend to attract each other. This phenomenon is notably exploited in the MEMS gyroscopes described above, to give a speed  $\vec{v}$  to the structure. It is also used to generate test stimuli in accelerometers, to control their position, and so on.

### 1.4.2 Active transducers

Active transducers rely on "effects" whose understanding requires a good command of physics: thermoelectric effect, piezoelectric effect, photoelectric effect, etc. It is not our point here of drawing up an exhaustive inventory but rather of giving a few "classic" examples.

### Examples:

1 - A **thermocouple** is made from a junction between two different conductors. Due to the Seebeck effect (thermoelectric effect), an electromotive force appears between the free ends of the thermocouple. The resulting voltage is, to the first order, proportional to the temperature difference between the "hot" junction (the point whose temperature is to be measured) and the free ends of the two conductors (supposedly regulated at a reference temperature). Thermocouples generally have poorer linearity than RTDs, but a slightly faster response time and a larger measuring range.

2 - Some crystalline materials are **piezoelectric**. This means that they become polarized when they are deformed ("direct" effect) and, vice-versa, that they deform under the effect of an electric field ("inverse" effect). A layer of piezoelectric material between two electrodes thus constitutes an active transducer, generating an electromotive force that is a function of the deformation experienced. This type of transducer can be used in the same type of circumstances as the (piezo-)resistive strain gages seen above, with many advantages (no power supply, linearity, temperature stability) but at a higher cost. In the field of inertial sensors and miniature oscillators (time bases), piezoelectric / quartz technologies have long been the only ones able to address high precision measurement applications (especially in aeronautics). They are now competing with MEMS / silicon (and capacitive transduction) technologies that can offer a comparable level of performance at a lower cost.

3 - The photoelectric effect is implemented in **photodiodes**, which emit a current proportional to the light intensity to which they are subjected. Current imagers are composed of arrays of photodiodes and their associated conditioning electronics.

### Exercises:

1 - What is a Hall effect sensor? Is the Hall effect a passive or active transduction phenomenon?

2 - Give examples of the use of the thermoelectric / piezoelectric / inverse photoelectric effect.

### **1.5 References**

Metrology is the science of accuracy, and the vocabulary of metrology must therefore be particularly accurate. To get a feel for the extreme rigor that is required in this respect, it is interesting to consult the following glossary (freely available on the Internet):

[1] "International vocabulary of metrology - Basic and general concepts and associated terms (VIM), 3rd edition ", JCGM, 2012

Nevertheless, one must be aware that there is a huge gap between the formal definition of the characteristics of a measurement and the information available in the datasheets of commercial sensors (go to the datasheets of Bosch accelerometers, STMicroelectronics or Analog Devices)... and also with the metrology notions presented in this section. In this respect, a rather general piece of advice for choosing between two sensors would be to pick the one with the most complete datasheet.

Regarding the different types of transducers, there are plenty of references. In order to deepen one's knowledge of a particular type of transducer, it may be useful to first refer to the "generalist" articles in Techniques de l'Ingénieur (which can be accessed through the CentraleSupélec library).

## 2 Conditioning electronics

### 2.1 Bridges and potentiometric readouts

Passive transducers are based on the variation of an impedance *Z* with the quantity *A* one wishes to measure. When *A* varies by  $\delta A$ , *Z* varies by  $\delta Z$  which is often small with respect to *Z*, but which alone carries information on *A*. This justifies the use of special bridge circuits to use with passive transducers.



*Fig. 2-1 - Two naive ideas. The arrow symbolizes that the impedance is variable (with the magnitude A).* 

For example, consider the two naive ideas shown in Fig. 2.1. In one case, the transducer is supplied with a voltage  $V_b$  and the current flowing through it is measured using a transimpedance amplifier. The output voltage of this circuit is

$$V_{out} = -\frac{Z_f}{Z + \delta Z} \times V_b \approx -\frac{Z_f}{Z} \times \left(1 - \frac{\delta Z}{Z}\right) \times V_b \quad (2.1)$$

In the other case, the transducer is supplied with current  $I_b$  and the voltage at its terminals is measured with a non-inverting amplifier. The output of the circuit is then expressed as follows

$$V_{out} = \left(1 + \frac{R_2}{R_1}\right) \times (Z + \delta Z) \times I_b$$
 (2.2)

In either case, the output of the amplifier is (or can be approximated as) the sum of two terms: a so-called "common mode" term, which does not carry information on *A* and a "useful" term carrying the information, all the smaller in relation to the common mode, as the impedance variations that are to be detected are small. The amplification that can be brought to the useful signal is therefore limited by the saturation of the amplifier due to the common mode signal.

This problem is essentially unique to passive transducers; active transducers generally deliver a useful signal that is not masked by a common mode signal. To solve this problem, impedance bridges, such as **Wheatstone bridges**, or **potentiometric readouts** are used.



Fig. 2.2 - Examples of potentiometric readouts with current (left) or voltage (right) output.



Fig. 2.3 - Differential strain gauge measurement (left) and differential capacitive detection (right). Images from National Instruments and Analog Devices.

### 2.1.1 Potentiometric readouts

Let us first consider the case of potentiometric readouts. These require the use of two voltage sources of the same value  $V_b/2$  and of a reference impedance  $Z_{ref}$  whose value is equal to the nominal impedance of the transducer. The output quantity is taken at the midpoint of the two impedances. It can be a voltage or a current, as shown in Fig. 2.2. In either case, the common mode is eliminated at the input of the amplifier.

A variant (among others) of this assembly, known as "differential", consists in replacing  $Z_{ref}$  by a transducer identical to the first one, which is affected by the quantity A with an opposite sign, so that  $Z_{ref} = Z - \delta Z$ . Fig. 2.3 gives practical illustrations of this in the cases of strain gages and capacitive transduction. This solution, which is not always feasible, has the advantage that it has better linearity than the previous one, is twice as sensitive to the measurand and is much less sensitive to the influence quantities.

In the case of the differential potentiometer circuit with voltage output (Fig. 2.2), the voltage  $V_{in}$  seen by the AFE is therefore

$$V_{in} = \frac{V_b}{2} \times \frac{(Z - \delta Z) - (Z + \delta Z)}{(Z - \delta Z) + (Z + \delta Z)} = -\frac{V_b}{2} \times \frac{\delta Z}{Z} \quad (2.3)$$

In the case of a current output, the current  $I_{in}$  is written

$$I_{in} = \frac{V_b}{2} \times \frac{(Z - \delta Z) - (Z + \delta Z)}{(Z - \delta Z)(Z + \delta Z)} = -\frac{V_b}{2} \times \frac{2\delta Z}{Z^2 - \delta Z^2}$$
(2.4)



Fig. 2.4 – Quarter-, half- and full-bridge configurations of a Wheatstone bridge supplied with voltage (top) or current (bottom). Illustrations from [1].

### 2.1.2 Wheatstone bridge

A Wheatstone bridge requires four impedances of the same kind (at least one of which is a transducer) instead of two, but only requires one source (of voltage  $V_b$  or current  $I_b$ ). The assembly is called "quarter-bridge", "half-bridge" or "full-bridge" depending on whether the number of transducers is one, two or four. The principle consists in measuring (and then amplifying) the voltage difference across the midpoints of the two branches of the bridge. This is zero when the bridge is balanced, i.e. when all impedances are equal ( $\delta Z = 0$ ). Otherwise, it is proportional to  $\delta Z$  to the first order (Fig. 2.4 provides the expression of the bridge output in 8 possible configurations). The "full-bridge" solution, which requires 4 transducers, has the advantage of being perfectly linear, is 4 times more sensitive to the measurand than the "quarterbridge" solution and is much less sensitive to influence quantities.

Remarks:

1 - The differential potentiometric readout with voltage output is linear in  $\delta Z$ , as opposed to the one with current output.

2 - The current-supplied Wheatstone half-bridge is linear in  $\delta Z$ , as opposed to the voltage-supplied one.

3 - Increasing the bias voltage (or bias current) increases the sensitivity of all these circuits. This approach is nevertheless limited by many practical reasons: consumption, self-heating, limitations specific to the integration technology...

4 - Any fluctuation of  $V_b$  (or  $I_b$ ) causes a fluctuation of the sensor output, decorrelated from the measurand value. The bias voltage (or current) is therefore an influence quantity to be taken into account, which may cause a sensitivity error.

5 - The major difficulty in the practical realization of these circuits is "matching" the different elements constituting them in terms of sensitivity to influence quantities. For example, let us consider a "quarter-bridge" circuit used to carry out a strain measurement, consisting of a strain gauge and 3 "passive" resistors. If the strain gauge and the resistors do not have precisely the same sensitivity to temperature, any variation in this quantity will result in a variation in the output voltage, unrelated to the strain that is being measured. In a potentiometric circuit, the voltage sources also need to be matched.

6 - We have not made any particular hypothesis concerning  $V_b$  (or  $I_b$ ). If simplicity dictates that this source be chosen constant (DC), there are many reasons why this choice is not necessarily optimal, or even feasible:

a - capacitive transducers have an infinite DC impedance, which makes it impossible to bias them with a constant source (except in the special case of resonant transducers).

b - using an AC source means that the impedance variation  $\delta Z$  becomes amplitudemodulated. Useful information, carried by  $\delta Z$ , is thus transposed at a higher frequency. This reduces the impact of low-frequency noise (flicker, drift) on the accuracy of the measurement.

c - power consumption can be greatly reduced if the source has no DC component.

On the other hand, the choice of an AC source requires the implementation of a clock/local oscillator, requires that a demodulation be performed after the amplification stage and imposes constraints on the AFE (many characteristics of opamps degrading with frequency).

## 2.2 Analog Front-Ends

As emphasized in Part 1, the AFE should provide the majority of the gain required by the application, and introduce as few errors as possible. These errors are closely related to the non-idealities of passive components and op-amps, which are discussed in section 2.3.



Fig. 2.5 - Transimpedance circuit for current-voltage conversion. Impedance  $Z_p$  models the interconnection between the transducer (the power source) and the electronics.

### 2.2.1 Current-output transducers

When a transducer has a current (or charge) output, the amplifier of choice is the transimpedance amplifier (Fig. 2.5). Its characteristic (considering an ideal op amp) is

$$V_{out} = -Z_f I_{in} \quad (2.6)$$

Impedance  $Z_f$  often consists of a resistor  $R_f$  in parallel with a capacitor  $C_f$  so that

$$Z_f = \frac{R_f}{1 + jR_f C_f \omega} \quad (2.7)$$

The values of  $R_f$  and  $C_f$  are dimensioned differently depending on whether one wishes to integrate the current  $I_{in}$  (one must then choose  $R_f \gg 1/C_f \omega$  so that the current  $I_{in}$  flows preferably through the capacitor rather than through the resistor).

Remark:

Another interest of a transimpedance amplifier is that it sets to zero (via virtual ground) the voltage of the node through which the current flows. Thus, whatever the parasitic impedance  $Z_p$  of the interconnection between the transducer and the AFE (in red in Fig. 2.5), the characteristic (2.6) is unchanged, within the limit of the  $V_+ = V_-$  hypothesis. This property is particularly advantageous in applications where the transducer and the AFE are not physically close.

### 2.2.2 Voltage-output transducers

A simple non-inverting amplifier may be suitable for a transducer whose output is a "useful" voltage, which is not masked by a common mode signal. This solution is simple and has the advantage of having, in theory, an infinite input impedance.

In the case of a bridge circuit, the instrumentation amplifier (Fig. 2.6) is preferred. From a "macroscopic" point of view, we can simply consider an instrumentation amplifier as a differential amplifier, whose input impedance is infinite. When  $R_1 = R_2 = R_3 = R_4 = R$ , the output of the circuit is ideally

$$V_{out} = \left(1 + 2\frac{R}{R_g}\right) \times \delta V \tag{2.8}$$

where  $R_q$  is external to the amplifier, so that the user can set the gain as desired.



Fig. 2.6 - Instrumentation amplifier. The resistance  $R_g$  sets the circuit gain, the resistors  $R_1$ ,  $R_2$ ,  $R_3$  and  $R_4$  are nominally equal.

Remark:

Instrumentation amplifiers are rarely made from individual discrete components. The main reason is that the proper functioning of the circuit depends on the strict matching of the components (op-amps, resistors) of the upper branch with those of the lower branch. For example, if the values of  $R_1$ ,  $R_2$ ,  $R_3$  and  $R_4$  (Fig. 2.6) are not strictly identical, we have

$$V_{out} = \left(1 + 2\frac{R}{R_g}\right) \times \delta V + \left(\frac{1 + R_2/R_1}{1 + R_3/R_4} - \frac{R_2}{R_1}\right) \times \frac{V_b}{2}$$

The useful signal will thus be marred by a common-mode error proportional to the bridge bias (all the more important as the mismatches are important) and will be more likely to drift with temperature. Commercial instrumentation amplifiers are therefore designed from integrated components, "trimmed" by laser to ensure as good a common mode rejection as possible.

### 2.3 Non-idealities of electronics

This non-exhaustive section lists the main causes that may affect the gap between the nominal behaviour of a signal conditioning chain and its actual behaviour. Therefore, these non-idealities all have an impact on the error, systematic or stochastic, of a given sensor or on the repeatability of measurements from one sensor to another.

### 2.3.1 Passive components

### 2.3.1.1 Variability

The nominal value of passive electronic components, whether discrete or integrated, varies from one component to another. This results from the inevitable dispersion of the manufacturing processes of these components. The other characteristics of these components (intrinsic noise, temperature coefficient, etc.) which depend on this nominal value are therefore also variable.

When required, discrete "precision" resistors with **tolerances** (relative error to nominal value) of 0.005% may be used. The tolerance of discrete precision capacitors is around 1%.

### 2.3.1.2 Drift

The term drift generally applies to any slow fluctuation in the output of a sensor that is unrelated to the measured variable. It covers both the effects of ageing (very slow fluctuation, therefore) and the effects of environmental influence quantities, such as temperature, humidity, etc., which affect the nominal value of passive components.

Typical values of the **temperature coefficients (**TC) of discrete resistors and capacitors are 10-100ppm/°C (compare for example with the relative sensitivity of Pt100 RTDs of 3850ppm/°C).



Fig. 2.7 - Modelling of thermal noise in resistors.



Fig. 2.8 - Examples for noise calculation.

### 2.3.1.3 Noise

All energy dissipating devices are sources of noise. Qualitatively, this is how the fluctuation-dissipation theorem can be summarized. While capacitors (and inductors) are essentially conservative systems, resistors dissipate energy (in the form of heat) when a current flows through them. Thus, any resistor in a conditioning circuit contributes to the stochastic error of the sensor.

This noise, known as **Nyquist-Johnson noise** or thermal noise, can be modeled as the result of a random voltage source in series with the resistor, or as a random current source in parallel with the resistor (Fig. 2.7). It is classically described as "white" (of uniform spectrum), Gaussian, and of density

$$S_{V_{noise}}(f) = 4k_B T \times R \left[ V^2 / Hz \right]$$
(2.9)

or

$$S_{I_{noise}}(f) = 4k_B T/R [A^2/Hz]$$
 (2.10)

where  $k_B = 1.38 \times 10^{-23}$  J/K is Boltzmann's constant, *T* is temperature and *R* is resistance.

To obtain the density resulting from a noise source at another point in the circuit, simply multiply (2.9) or (2.10) by the squared modulus of the corresponding transfer function. To obtain the noise variance at this point on a bandwidth  $[f_0, f_0 + \Delta f]$  the noise density at this point must be integrated between  $f_0$  and  $f_0 + \Delta f$  (which is equivalent to multiplying the density by  $\Delta f$  if the

density is uniform). Finally, the variance (or density) of the noise resulting from several independent noise sources is equal to the sum of the variances (or densities) of the contributions of the sources at that point.

Example:

Consider the transimpedance amplifier in Fig. 2. 8. One seeks the contribution of the resistor  $R_f$  to the input- or output-referred noise, on a bandwidth  $[f_0, f_0 + \Delta f]$ . In this case, the simplest way is to consider  $R_f$  as having a parallel current source, whose density is given by (2.10). This current source can only flow in the complex impedance  $Z_f$  made up of  $R_f$  and  $C_f$  in parallel. Thus, the output-referred noise has a density of

$$S_{V_{out}}(f) = \left| Z_f \right|^2 \times \frac{4k_B T}{R_f}$$

and the input-referred noise density is

$$S_{I_{in}}(f) = \frac{4k_BT}{R_f}$$

The variance of output-referred noise is

$$\overline{V_{out}^2} = \int_{f_0}^{f_0 + \Delta f} S_{V_{out}}(f) df = \frac{4k_B T}{R_f} \times \int_{f_0}^{f_0 + \Delta f} |Z_f|^2 df$$

that of the input-referred noise

$$\overline{I_{in}^2} = \int_{f_0}^{f_0 + \Delta f} S_{I_{in}}(f) df = \frac{4k_B T}{R_f} \Delta f$$

Nyquist-Johnson noise is a **universal** phenomenon that affects **all** dissipative systems, whether "electronic" or not; this is the case for example with resistive transducers, MEMS transducers (whose kinetic energy is dissipated in the ambient fluid), etc.

#### Exercise:

Determine the contribution of the resistors to the input-referred noise of the noninverting amplifier in Fig. 2.8.

### 2.3.2 Active components

We are interested here in the non-idealities of op-amps, seen in a "macroscopic" way as discrete components. The tutorials available on the <u>Analog Devices</u> website will be a real mine of information for anyone who wants to learn more about these issues.

### 2.3.2.1 Variability and drifts

Like passive components, op-amps are subject to drift and dispersion. The nominal values of their characteristics are therefore likely to vary from one amp to another, over time or with operating conditions.

### 2.3.2.2 Gain and bandwidth

An ideal op amp has an infinite open-loop gain  $G_{OL}$  and an infinite bandwidth, it can therefore theoretically amplify a signal with an arbitrarily large gain, regardless of its frequency. It is this infinite gain that imposes that  $V_+ = V_-$  for an ideal op amp operating in the linear regime.

In reality, the open-loop gain:

- o is finite.
- o decreases with frequency.

The gain of an op amp can be approximated as having a first-order low-pass characteristic:

$$G_{OL}(f) = \frac{G_{OLmax}}{\left|1 + j\frac{f}{f_c}\right|}$$
 (2.11)

where  $G_{OLmax}$  is the maximum value of the open-loop gain, and  $f_c$  is a cut-off frequency beyond which  $G_{OL}$  decreases. We note that, for  $f \gg f_c$ , the **gain-bandwidth product** (GBP, expressed in Hz)  $G_{OL} \times f$  is constant and equal to  $G_{OLmax} \times f_c$ . This quantity is used to characterize the op-amp: it corresponds to the frequency for which  $G_{OL} = 1$ .

This characteristic limits, for example, the maximum frequency at which to bias a bridge or a potentiometric circuit, used for conditioning the signal from a passive transducer.

Exercise:

We wish to make a non-inverting amplifier with a gain of 100 from an op amp with a gain-band product of 100MHz. In which bandwidth will we actually have a gain of 100 (to the nearest 3dB)?



Fig. 2.9 - Op-Amp model considering a finite open-loop gain and finite input impedance.

### 2.3.2.3 Input impedance

The input impedance of an ideal op amp is infinite and, as a result, the currents are  $I_+$  and  $I_-$  are theoretically zero. The input impedance of a real op amp :

- is finite.
- o is not purely resistive and decreases with frequency.

This impedance is modeled as in Fig. 2.9, with a "differential" impedance between the + and - inputs of the amp and two "common mode" impedances between the amplifier inputs and the ground. Each of these impedances consists of a resistive branch and a capacitive branch. In low-frequency applications (i.e. for a  $RC\omega \ll 1$ ), the capacitive branch plays no role and the input

impedance is therefore purely resistive. At high frequencies, on the contrary, the capacitive component dominates.

The fact that the input impedance is finite has an effect on the sensitivity of the sensor, especially when the open-loop gain of the op amp is low. Inn fact, if  $G_{OL}$  is infinite, then  $V_+ = V_-$  and no current flows in the differential part of the input impedance (the input impedance could therefore theoretically take an arbitrarily small value).

The resistive component of the input impedance of commercial op-amplifiers has typical values ranging from  $10^6\Omega$  to  $10^{12}\Omega$ . The capacitive component is of the order of a few pF.

#### Exercises:

1 - Calculate the output voltage of a transimpedance amplifier, in the case of finite gain / finite input impedance. Show that the maximum value of the circuit gain is limited by these two quantities.

2 - Show that the output of a non-inverting amplifier is not affected by the value of the input impedance if the gain of the op-amp is infinite. What else changes?

Remark:

In addition to the input impedances that are intrinsic to the op amp, one must also consider those of the interconnects, which will be discussed at greater length below.

### 2.3.2.4 Common mode rejection ratio

The output voltage of a real op-amp is written :

$$V_{out} = G_{0L} \times (V_+ - V_-) + G_{cmn} \times (V_+ + V_-)/2$$
(2.12)

where  $G_{cmn}$  is the common mode gain. This quantity is ideally zero and the **common mode** rejection ratio (CMRR) of the amp

$$CMRR = \frac{G_{OL}}{G_{cmn}}$$
(2.13)

is theoretically infinite (in practice, this quantity is ranges from  $10^4$  to  $10^7$  in commercial components). This non-ideality can for example result in a bias error, such as when the signals at the input of the amplifier are "useful" voltages superimposed on a constant common mode voltage.

Like  $G_{OL}$  or the input impedance, the CMRR of an op amp decreases with frequency.

Remark:

Other "rejection rates" (PSRR, ripple rejection) characteristics of op-amps represent their ability to provide an output voltage independent of fluctuations in their supply voltages.

For a sensor, the stability of the supply, bias or reference voltages of the various elements must be the subject of particular care. There are many methods to "stabilize" voltage sources to some extent, such as the use of decoupling capacitors or active voltage regulators (LDOs).

All of the above-mentioned op-amp characteristics degrade with frequency. However, several other non-idealities, described below, also limit the use of conventional op-amps at low frequencies.

### 2.3.2.5 Offset voltage and bias currents

Offset voltage and bias currents are purely static non-idealities. These non-idealities can be modeled as shown in Fig. 2.10.



Fig. 2.10 - Model of the offset voltage and bias currents in a real op amp (left), and voltage and current noise (right). Only the nature (deterministic or stochastic) of the sources and their frequency content change.

The offset voltage is the DC voltage that should be applied between the + and - inputs to make the output of the op amp strictly zero. It results from mismatches in the first stage of the internal architecture of the op amp (classically, a "differential pair"). The term "precision" op amp refers to an op amp with a typical offset voltage value below 1mV (and low noise, see below).

Bias currents are DC currents, whose intensity can ranges from a few fA to a few nA depending on the technology (bipolar, FET, etc.) of the op amp. These currents are necessary for the proper operation of the circuit: they must therefore be able to flow to the ground or to an ideal voltage source (including the output of the amp). These currents result in voltage drops in the resistors through which they pass, and thus in bias errors.

Many techniques exist to mitigate these two defects. For example, for a given op amp, the choice of adequate compensation resistors can mitigate the error introduced by the bias currents.

Exercise:

Determine the expression of contribution of the bias currents and offset voltage to the input-referred error of the transimpedance and non-inverting amplifiers in Fig. 2.8.

### 2.3.2.6 Voltage noise and current noise

The differential voltage and input currents of a real op amp not only have a static component, as described above, but also a random component (Fig. 2.10). From the point of view of a sensor, this random component leads to a stochastic error, just as the DC component leads to a systematic error.

Accounting for these noises, independent of each other, in an error budget analysis follows the same pattern as in the case of passive components:

- determine the transfer function between the noise source and the point at which the error is to be calculated,
- calculate the noise density at this point by multiplying the source density by the squared modulus of the transfer function,
- integrate the density over the bandwidth of interest to obtain the variance corresponding to each of the random components,
- sum these variances to obtain the total variance.

The only real peculiarity of the noise sources associated with the op amp is their spectrum, which is generally not white. The transition from density to variance therefore requires a little more effort than in the case of Nyquist-Johnson noise in the resistors.



FREQUENCY (Hz)
Figure 51. Voltage Noise Density vs. Frequency

Fig. 2.11 - Voltage noise characteristics for three Analog Devices op-amps: AD8099 (top left, bipolar input stage amp), AD8065 (top right, FET input stage amp) and ADA4528 (bottom, auto-zero amp). Pay attention to the axes units!

Thus the voltage noise of an op-amp has (except in the special cases mentioned below) a **flicker** component (or **pink** noise or **1**/**f** noise), as shown in Fig. 2.11. The density of the flicker noise is inversely proportional to the frequency. Low-frequency applications are therefore particularly sensitive to it. This noise is generally characterized by its **corner frequency**  $f_c$ , for which the pink and white noise components have the same amplitude. This corner frequency varies greatly from one op-amp model to another. The white component of voltage noise is the

result of internal dissipation within the amp (so it is partly Nyquist-Johnson noise) and other phenomena (e.g. shot noise). Flicker noise can have various physical origins, the most common being the co-existence of several independent random phenomena with different relaxation times.



Figure 1. Current noise of AD8099, a bipolar input amplifier.

Figure 2. Current noise of AD8065, a FET input amplifier.



Figure 52. Current Noise Density vs. Frequency

Fig. 2.12 - Current noise characteristics for three Analog Devices amps: AD8099 (top left, bipolar input stage amp), AD8065 (top right, FET input stage amp) and ADA4528 (bottom, auto-zero amp). **Pay** *attention to the axes units!* 

The current noise spectrum of op-amps may also have a flicker component (characterized by a corner frequency not necessarily identical to that of the voltage noise), but this is not necessarily the case. The current noise characteristics of various commercially available op-amps are shown in Fig. 2.12.

Finally, some op-amp architectures, known as "auto-zero" or "chopper-stabilized", can cancel the effect of flicker noise and offset voltage, at the cost of a bandwidth limited to a few tens of kHz.

### Remark:

All the characteristics mentioned in this section are obviously dependent on the ambient temperature.

### 2.3.3 Interconnects



Fig. 2.13 - Ideal interconnect (left) vs. actual interconnect (right).

A sensor is made up of passive and active elements... connected to each other by wires, or tracks in the case of printed / integrated circuits. These interconnects can contribute significantly to sensor performance.

For relatively low-frequency applications, interconnects can be modeled using parasitic resistors and capacitors, as shown in Fig. 2.13. The most critical interconnects are those located furthest upstream in the conditioning chain, i.e. those between the transducer and the AFE.

Consideration of transducer-AFE interconnect resistance is particularly important for transducers with small internal resistance (e.g. thermocouple, RTD) and/or bridge-mounted transducers. In the latter case, for example, the resistance of the interconnects influences the bias and sensitivity of the sensor. For passive transducers, the interconnect between the (current or voltage) source and the transducer is also critical.

The higher the operating frequency, the greater the influence of interconnect capacitance. Depending on which method is used for integration, the transducer-AFE interconnect capacitance can be much higher than the input capacitance of the op amp, which consequently decreases the input impedance of the circuit (see above). The minimization of parasitic capacitance (due to interconnects, but not only) by technological means, their compensation by passive or active techniques (bootstrap), their modeling and calibration are part of the concerns of all "electronic engineers". Fig. 2.14 shows, for example, the set of parasitic capacitances surrounding a MEMS inertial sensor.


C<sub>P sub</sub>: grounding parasitic from structure to substrate  $C_{P \ bond}$ : grounding parasitic introduced by bonding

Cc bond: paralleling parasitic between bonding Cc\_rout: paralleling parasitic between routing

Fig. 2.14 - Sectional view of a packaged MEMS sensor (taken from https://doi.org/10.1016/j.sna.2020.112159). The transducer is on the left, the associated electronics on the riaht.

## 2.4 References

If you are interested in these topics, the website of Analog Devices is a real mine!

https://www.analog.com/en/education/education-library/tutorials.html

Two books (a little old, but not too wrinkled) in the "Technical Books" section are particularly useful to extend your knowledge of AFE and op-amps in general:

[1] W. Kester, "Practical design techniques for sensor signal conditioning", Analog Devices, 1999

[2] W. G. Jung, "Op Amp Applications", Analog Devices, 2002

Finally, the "Analog Electronics" section of the tutorials contains many short "factsheets" (typically 5-6 pages) dedicated to specific topics related to op-amps and their applications. I copy some of them here:

### **Operational Amplifiers:**

1. Ideal Voltage Feedback (VFB) Op Amp (MT-032)

I. Inverting Amplifier (MT-213)

II. Inverting Summing Amplifier (MT-214)

- III. Half Wave Rectifier (MT-212)
- IV. Full Wave Rectifier (MT-211)
- 2. Current Feedback (CFB) Op Amps (MT-034)
- 3. Voltage Feedback Op Amp Gain and Bandwidth (MT-033)
- 4. Open Loop Gain and Open Loop Gain Nonlinearity (MT-044)
- 5. Bandwidth and Bandwidth Flatness (MT-045)
- 6. Settling Time (MT-046)
- 7. High Speed Voltage Feedback Op Amps (MT-056)
- 8. Input Offset Voltage (MT-037)
- 9. Total Output Offset Voltage Calculations (MT-039)
- 10. Chopper Stabilized (Auto-Zero) Precision Op Amps (MT-055)
- 11. Input Bias Current (MT-038)
- 12. Input Impedance (MT-040)
- 13. Power Supply Rejection Ratio (PSRR) and Supply Voltages (MT-043)
- 14. Input and Output Common-Mode and Differential Voltage Range (MT-041)
- 15. Common-Mode Rejection Ratio (CMRR) (MT-042)
- 16. Outputs, Single-Supply, and Rail-to-Rail Topics (MT-035)
- 17. Output Phase-Reversal and Input Over-Voltage Protection (MT-036)

#### **Using Op Amps**

- 1. Choosing Between Voltage Feedback and Current Feedback Op Amps (<u>MT-060</u>)
- 2. Compensating for the Effects of Input Capacitance on VFB and CFB Op Amps Used in Currentto-Voltage Converters (<u>MT-059</u>)
- 3. Effects of Feedback Capacitance on VFB and CFB Op Amps (MT-058)
- 4. Op Amp Distortion: HD, THD, THD + N, IMD, SFDR, MTPR (MT-053)
- 5. Op Amp Noise (<u>MT-047</u>)
- 6. Op Amp Noise Figure: Don't Be Mislead (MT-052)
- 7. Op Amp Total Output Noise Calculations for Second-Order System (<u>MT-050</u>)
- 8. Op Amp Total Output Noise Calculations for Single-Pole System (<u>MT-049</u>)
- 9. Op Amp Noise Relationships: 1/f Noise, RMS Noise, and Equivalent Noise Bandwidth (<u>MT-048</u>)

#### **Instrumentation Amplifiers**

- 1. Basic Two Op Amp In-Amp Configuration (<u>MT-062</u>)
- 2. Basic Three Op Amp In-Amp Configuration (<u>MT-063</u>)
- 3. In-Amp DC Error Sources (MT-064)
- 4. Auto-Zero In Amps (<u>MT-067</u>)
- 5. In-Amp Noise (<u>MT-065</u>)
- 6. In-Amp Bridge Circuit Error Budget Analysis (MT-066)
- 7. Difference and Current Sense Amplifiers (MT-068)
- 8. In-Amp Input Overvoltage Protection (<u>MT-069</u>)
- 9. In-Amp Input Radio Frequency Interference Protection (MT-070)
- **10.** <u>A Deeper Look into Difference Amplifiers</u>

# 3 MEMS accelerometers and gyroscopes

Before addressing the physics of MEMS, it seems interesting to draw a "high-level" picture of the operating principle and metrological characteristics of the two most common MEMS inertial sensors: accelerometers, which measure accelerations, and gyrometers (or, by misuse of language, "gyroscopes"), which measure angular velocities. Without going into too much detail, this will illustrate some of the concepts mentioned in the previous sections and justify the main developments in the following sections.

Accelerometers and gyroscopes are used in a multitude of applications, in extremely varied fields: detection of motion, vibration, shock, tilt, for consumer applications (e.g. rotating the image on a smartphone screen according to its orientation), for automotive safety (e.g. triggering airbags in case of impact, stabilizing the attitude of a vehicle), for monitoring civil engineering structures or the human body (fall detection, pacemaker regulation), for example. These applications require a wide range of performances from inertial sensors, from a metrological point of view (measurement range, bandwidth, error), but also in terms of price, consumption and sensor size. Instrument navigation applications ("dead reckoning") are the most demanding in terms of error. Indeed, the position of a "vehicle" (autonomous car, missile, drillhead) can theoretically be computed by integrating acceleration measurements twice over time, and its orientation by integrating angular velocity once over time. But measurement errors are also integrated over time, so that, for example, an acceleration bias error results in a position error that increases as  $t^2$ . Other measurements, such as those from magnetometers, velocimeters, barometric pressure sensors, GPS, etc., are necessary to "resetting" such a system on a regular basis.

In what follows, we will focus on sensitivity and noise in inertial (single-axis) MEMS sensors, and the link between these quantities and the parameters of the mechanical structure and electronic architecture constituting them.

## 3.1 MEMS accelerometers

Two essential functional blocks can be distinguished in a MEMS accelerometer: a transducer that converts the acceleration into an impedance variation, and conditioning electronics that transform this impedance variation into a useful electrical signal. Today, the vast majority of commercial MEMS accelerometers rely on capacitive transduction, so this is the case we are looking at.

As mentioned in part 1, capacitive transduction makes it possible to convert a displacement into an impedance variation. A necessary prerequisite to make a capacitive accelerometer is therefore to convert acceleration into displacement.

Macroscopically, this can be done very simply with a dynamometer, or load cell, (Fig. 3.1) whose elongation can be used to determine the weight (i.e. the product of mass by acceleration) suspended from the hook. Typically, a load cell is used to measure an unknown mass assuming constant and known acceleration (gravity). But one can just as well reason with a constant known mass and use the elongation of the spring as a measure of the acceleration experienced.

"Single-axis" MEMS accelerometers are based on the same principle. They consist of a mass *M* suspended by an elastic structure designed to be flexible in one direction, along which acceleration is sensed, and very stiff in the others.



Fig. 3.1 - Examples of load cells, taken from <u>http://www.lecompendium.com</u>, from the "Traité de Physique Elémentaire" (Drion and Fernet, 1885) and the "Dictionnaire de l'Industrie et des Arts Industriels" (E.O. Lami, 1887).

## 3.1.1 Sensitivity

Under the effect of a *constant* acceleration *A* in the direction of interest, the mass moves by

$$X = -\frac{M}{K} \times A \quad (3.1)$$

where *K* is the mechanical stiffness of the transducer. We can consider an acceleration as "constant" if its spectrum is limited to a bandwidth  $[0, \Delta f]$  where  $\Delta f$  is small with respect to the resonance frequency of the structure :

$$f_0 = \frac{1}{2\pi} \times \sqrt{\frac{K}{M}} \quad (3.2)$$

It is this mechanical resonance frequency that sets the maximum bandwidth of an accelerometer.

When *X* varies (in response to acceleration), the capacitance between the moving mass and the fixed part of the device also varies. If possible, the geometry of the structure is designed so that a capacitance increases with *X* and another decreases with *X* in the same proportions (see Fig. 2.3), in order to implement a differential potentiometric readout. This is, for example, the case of the accelerometer in Fig. 3.2 for which we have

$$C^{+} = N \times \frac{\epsilon_0 S}{g + X} = C_0 \times \frac{1}{1 + X/g} \quad (3.3a)$$
$$C^{-} = N \times \frac{\epsilon_0 S}{g - X} = C_0 \times \frac{1}{1 - X/g} \quad (3.3b)$$

where *N* is the number of "fingers" of the structure (we consider here that the total capacitance results from *N* plane capacitors) and  $C_0$  is the capacitance in the absence of acceleration.



Fig. 3.2 - Simulation of a capacitive MEMS accelerometer using MEMS+. On the left, overview of the structure: moving parts (mass) and deformable parts (suspensions / springs) in green, fixed parts (anchors, electrodes) in yellow. On the right, simulation of an acceleration directed towards the right (resulting in a displacement to the left). The capacitance between the moving fingers and the left electrodes increases, that formed with the right electrodes decreases.

The potentiometric readout is biassed with a voltage  $V_b$  at frequency  $f_b$  as shown in Fig. 2.2, so that the quasi-static capacitance variation is modulated around  $f_b$ . In the case of the differential potentiometric readout with voltage output, the voltage  $V_{in}$  seen by the AFE is

$$V_{in} = -\frac{V_b}{2} \times \frac{X}{g} \tag{3.4}$$

and the output voltage of a non-inverting amp

$$V_{out} = -\left(1 + \frac{R_2}{R_1}\right) \times \frac{V_b}{2} \times \frac{X}{g} = \left(1 + \frac{R_2}{R_1}\right) \times \frac{V_b}{2} \times \frac{M}{Kg} \times A \quad (3.5)$$

In the case of a readout with current output, the current  $I_{in}$  is written

$$I_{in} = -C_0 \times \frac{dV_b}{dt} \times \frac{\frac{X}{g}}{1 - \frac{X^2}{g^2}} \approx -C_0 \times \frac{dV_b}{dt} \times \frac{X}{g}$$
(3.6)

where the approximation is valid if  $|X^2| \ll g^2$ . If this current is integrated in a capacitor  $C_f$  using a transimpedance amplifier, we have

$$V_{out} \approx \frac{C_0}{C_f} \times V_b \times \frac{X}{g} = \frac{C_0}{C_f} \times V_b \times \frac{M}{Kg} \times A (3.7)$$

Remarks:

1 - Relationships (3.1) and (3.2) illustrate a first compromise in accelerometer design, between high sensitivity (large M/K) and wide bandwidth (large K/M).

2 - As highlighted in (3.5) and (3.7), the sensitivity of the sensor, in  $V/(m/s^2)$ , depends on parameters that are specific to the transducer, and others that are specific to the electronics. From the point of view of a sensor designer, these expressions are very useful. For example, in the case of a current output, a compromise must be found between M and  $C_0$  to maximize sensitivity. These two quantities cannot be maximized simultaneously for a given transducer area.

3 - The nominal capacitance does not appear in (3.5), which suggests that the accelerometer would work as well with N = 1 finger as with N = 100 fingers. However, this is only true within the limits of the ideal models (of op-amp and interconnects) underlying (3.5)!



Fig. 3.3 - High-level representation of the transduction and conditioning chain of an accelerometer. The spectra marked by numbers correspond to different points in the chain: (1) acceleration or inertial force, (2) total force (inertia + thermomechanical) acting on the structure, (3) displacement or capacitance variation (limited by the resonance frequency  $f_0$ ), (4) bridge output voltage (capacitance variations are modulated around the frequency  $f_b$ ) (5) voltage at the AFE input (and at its output by assimilating the op amp to a static gain), (6)  $V_{out}$  after demodulation, (7) filtered sensor output.

## 3.1.2 Noise

To facilitate the understanding of this section, refer to Fig. 3.3.

As mentioned in Part 2, Nyquist-Johnson noise affects all dissipative systems, including mechanical systems (it is also called thermomechanical noise in this context). Assuming the system is subject to a linear dissipative force  $F_{diss} = -B\dot{X}$  characterized by dissipation coefficient B, one may show that there also exists a random force  $F_{MEMS,n}$  of uniform density ("white")

$$S_{F_{MEMS\,n}} = 4k_BT \times B \left[ N^2 / \text{Hz} \right] (3.8)$$

acting on the system. In the case of MEMS accelerometers, the main cause of dissipation is the friction of the moving mass in the ambient fluid.

This noise obviously has repercussions on the motion of the mass and on the detection of the acceleration (on the "precision" of the measurement). In a bandwidth  $\Delta f$  (supposedly small with respect to  $f_0$ ), one may write

$$X = -\frac{M}{K} \times A + \frac{F_{MEMS,n}}{K} = -\frac{M}{K} \times \left(A - \frac{F_{MEMS,n}}{M}\right)$$
(3.9)

which highlights that the displacement is the superposition of an acceleration-dependent term and a stochastic error term. The term  $F_{MEMS,n}/M$ , which is homogenous to an acceleration, is the "input-referred" thermomechanical noise of the transducer. Its variance is equal to

$$\overline{A_{MEMS,n}^2} = 4k_BT \times \frac{B}{M^2} \times \Delta f \left[ (m/s^2)^2 \right]$$
(3.10)

The accelerometer "noise floor" set by (3.10) is intrinsic to the transducer; it can only be lowered by increasing the mass of the device, reducing friction (e.g. by vacuum-packaging of the transducer) or reducing the bandwidth.

Noise due to the electronic part of the system must be added to this intrinsic noise. For example, in the case of a potentiometric readout with a voltage output, we have

$$V_{out} = \left(1 + \frac{R_2}{R_1}\right) \times \left(\frac{V_b}{2} \times \frac{M}{Kg} \times \left(A - \frac{F_{MEMS,n}}{M}\right) + V_{AFE,n}\right)$$
(3.11)

where  $V_{AFE,n}$  is the voltage noise due to the electronics (passive and active components), referred to the input of the AFE. This noise can be referred to the input of the system as an acceleration noise with variance

$$\overline{A_{AFE,n}^{2}} = \frac{1}{\left(\frac{V_{b,1}}{2} \times \frac{M}{Kg}\right)^{2}} \times \overline{V_{AFE,n}^{2}} \left[ (m/s^{2})^{2} \right]$$
(3.12)

where  $V_{b,1}$  is the amplitude of the first harmonic of  $V_b$  (which may be harmonic, square, etc.) and  $\overline{V_{AFE,n}^2}$  is the variance of  $V_{AFE,n}$  calculated on the bandwidth  $[f_b - \Delta f, f_b + \Delta f]$ . As discussed in Part 1 of the handout, the contribution of the AFE to the stochastic error of the sensor can be reduced by increasing the "gain" (i.e. sensitivity) of the stage preceding the AFE (i.e. the transducer). Similarly, the contribution of a second amplification stage would be reduced by increasing the gain of the transducer or that of the AFE, i.e.

$$\overline{A_{2^{nd}\,stage,n}^2} = \frac{1}{\left(1 + \frac{R_2}{R_1}\right)^2 \left(\frac{V_{b,1}}{2} \times \frac{M}{Kg}\right)^2} \times \overline{V_{2^{nd}\,stage,n}^2} \left[ (m/s^2)^2 \right]$$
(3.13)

The resolution of most commercially available MEMS accelerometers is limited by the noise of the AFE, not by the intrinsic noise of the transducer, i.e.  $\overline{A_{MEMS,n}^2} \ll \overline{A_{AFE,n}^2}$ .

## 3.1.3 Dynamic effects in accelerometers

To account for the effect of broadband accelerations (e.g. shocks and vibrations), a dynamic model of the transducer is required:

$$KX + B\dot{X} + M\ddot{X} = -MA \Leftrightarrow X = -\frac{M}{K} \times H(p) \times A \quad (3.14-a)$$
$$H(p) = \frac{1}{1 + \frac{B}{K}p + \frac{M}{K}p^2} = \frac{1}{1 + \frac{1}{Q\omega_0} + \left(\frac{p}{\omega_0}\right)^2}. \quad (3-14-b)$$

which then replaces (3.1). In (3.14), p is the Laplace parameter,  $\omega_0 = 2\pi f_0 = \sqrt{K/M}$  is the resonance pulsation of the structure and  $Q = \sqrt{KM/B}$  its quality factor. As shown by (3.14-b), above  $f_0$ , which typically ranges between 100 Hz and 10 kHz, the structure responds less and less to acceleration. The transfer function H(p) is plotted in Fig 3.4 for different values of Q and the

response of the transducer to an acceleration step. Strictly speaking, dynamic effects must be taken into account in all noise calculations.



*Fig.* 3.4 - *Frequency response (left) and step response of* H(p) *for*  $\omega_0 = 1$  *and various quality factors.* 

The value Q = 1/2 is optimal in terms of response time without overshoot. A number of physical parameters may have an influence on *B* and therefore on *Q* (essentially viscosity of the ambient gas and dimensions of the gas films formed between the fixed and mobile electrodes). However, increasing *B* also means increasing the noise floor (3.10). This is not necessarily problematic, as long as electronic noise dominates. One may also use feedback control to force the system to reach its steady-state regime faster than the response time dictated by a large *Q*. This solution, which has the advantage of not raising the noise floor of the accelerometer, nevertheless makes the design of the system more complex. It essentially requires two steps, which can be carried out more simply in the digital domain:

- a "state estimation", which consists in determining, from the digitized signals, the estimates  $\hat{X}$  and  $\hat{X}$  the position and velocity of the moving mass. This operation is more or less complex depending on the degree of prior knowledge on the system to be controlled.
- the generation of a control C according to the estimated state. This can be as simple as

$$C = a\hat{X} + b\ddot{X}$$

in which case one modifies the effective stiffness and damping coefficients of the transducer (via the coefficients *a* and *b* respectively). A common alternative is to use a proportional-integral-derivative control.

$$C = \alpha (\hat{X} - X_{ref}) + \beta \int (\hat{X} - X_{ref}) dt + \gamma \frac{d}{dt} (\hat{X} - X_{ref})$$

Because of the integral term, the control becomes larger the more  $\hat{X}$  and  $X_{ref}$  are different over time. In practice, this control law is used to keep the position of the mass set to  $X_{ref}$ , within a certain bandwidth, for which it can be considered that the inertial force acting on the moving mass is exactly compensated by the force resulting from the control.

The control must obviously be transduced into a force in order to act on the moving mass. This requires not only a DAC when the controller is digital, but also specific electrodes to actuate the moving mass, or a time-division multiplexing scheme to perform actuation and detection with the same electrodes. A high-level schematic of such a system is shown in Fig. 3.5.



*Fig. 3.5 - Simplified diagram of a feedback-controlled accelerometer.* 

Exercises:

1 - In the case of an accelerometer without feedback control, how is value of  $\overline{A_{MEMS,n}^2}$  affected by taking into account transfer function H(p)? And that of  $\overline{A_{AFE,n}^2}$ ?

2 - Establish the expression of the transfer function between the inertial force and the control C in the case of a proportional-integral feedback control ( $\gamma = 0$ ), supposing  $\hat{X}$  is proportional to the output of the analog-to-digital converter and  $X_{ref} = 0$  so that one may write:

$$C = \left(\alpha + \frac{\beta}{p}\right)\hat{X}$$

What is the bandwidth of such a sensor? What is the input-referred noise of the sensor?

## 3.2 MEMS gyroscopes

Literally, a "gyroscope" is a sensor used for measuring an angle, and a "gyrometer" is a sensor used for measuring an angular velocity. To my knowledge, all MEMS sensors marketed under the name "gyroscope" are in fact gyrometers. They can therefore measure an angle only by integrating over time the result of an angular velocity measurement, which is not without risk since this measurement is marred by errors, as already mentioned in the introduction of this section. In what follows (and in the rest of this handout), the same abuse of language will be committed and we will specify "true gyroscope" when it comes to a ... true gyroscope.

MEMS gyroscopes rely on the Coriolis force that is exerted on a moving mass when its frame of reference is rotated

$$\vec{F}_{Coriolis} = -2M\vec{\Omega} \wedge \vec{V}, \qquad (3.15)$$

where *M* is the mass of the device,  $\vec{V}$  its speed and  $\vec{\Omega}$  the angular speed at which the rotation occurs. The detection of this force follows more or less the same principle as the detection of the inertial force due to an acceleration, in the case of the accelerometer: the force is converted into displacement, the displacement into a capacitance variation, the capacitance variation into current or voltage.

However, there are many differences between gyroscopes and accelerometers. To mention only the most obvious:

- the transduction of angular velocity into Coriolis force (3.15) is a passive transduction scheme, in the sense that it requires an external (kinetic) energy input. In the absence of this "velocity bias", there is no Coriolis force. In practice, the structure is actuated by electrostatic forces (see part 4) to vibrate harmonically in a given direction, called the drive direction, with frequency  $f_d$ .
- Coriolis force is orthogonal to  $\vec{V}$ . The mechanical structure of a gyroscope must therefore allow motion in this "sense" direction as well as in the "drive" direction. From a mechanical point of view, this makes the design of gyroscopes more complex than that of accelerometers.
- in the case of the accelerometer, the displacements and capacitance variations have the same frequency domain as the acceleration being measured. In the case of the gyroscope, the displacements resulting from the Coriolis force, and the consequent variations in capacitance, are centered around the frequency  $f_d$  (the angular velocity information is modulated at this frequency).

### 3.2.1 Sensitivity

Fig. 3.6 shows a single-axis MEMS gyroscope geometry that can be referred to as an illustration of this section.

The behavior of a gyroscope can be described in a *simplified* way by means of two coupled equations, one for the drive mode, one for the sense mode

$$K_d X + B_d \dot{X} + M \ddot{X} = F_d \qquad (3.16-a)$$
  
$$K_s Y + B_s \dot{Y} + M \ddot{Y} = -2M\Omega \dot{X} \qquad (3.16-b)$$

where *X* is the displacement of the mass *M* in the drive direction, *Y* its displacement in the sense direction, and  $\Omega$  the angular velocity (around the normal to the *XY* plane). This model is valid when  $\Omega \ll \omega_d$ , where  $\omega_d = \sqrt{K_d/M} = 2\pi f_d$ . In addition, we define  $\omega_s = \sqrt{K_s/M} = 2\pi f_s$ ,  $Q_d = \sqrt{K_dM}/B_d \gg 1$  and  $Q_s = \sqrt{K_sM}/B_s \gg 1$ . The structure is actuated in the drive direction by a force  $F_d = F \times \sin \omega_d t$  generated using an oscillator.

As the drive mode is excited at its natural frequency  $f_d$  we can approximate (3.16-a) by :

$$B_d \dot{X} = F_d \qquad (3.17)$$

so that (3.16-b) becomes

$$K_s Y + B_s \dot{Y} + M \ddot{Y} = -2M \frac{F_d}{B_d} \Omega$$
(3.18)

Finally, if  $f_d < f_s$  (so-called "split-mode" gyroscope) and if the bandwidth  $\Delta f$  of  $\Omega$  is small with respect to  $f_s - f_d$ , the right-hand term of (3.18) can be considered to vary slowly. Then, (3.18) is simplified into

$$Y = -2M \frac{F_d}{K_s B_d} \Omega \tag{3.19}$$

Sense displacement *Y* can therefore be described as a harmonic signal at frequency  $f_d$ , amplitudemodulated by  $\Omega$ , which is in phase with  $F_d$  and in quadrature with *X* so that the center of gravity of the mass describes an ellipse, more or less flattened depending on  $\Omega$ .

By forming ad hoc capacitorss between the fixed parts of the device and the moving part, a capacitive transduction of displacement *Y* can be performed so that

$$C^{+} = C_0 \times \frac{1}{1 + Y/g}$$
(3.20-a)

$$C^{-} = C_0 \times \frac{1}{1 - Y/q}$$
(3.20-b)

where we assume a detection via the variation of the gap g of plane capacitors.



Fig 3.6 - Simulation of a capacitive MEMS gyroscope using MEMS+ software. Top: geometry of the structure consisting of a central mass (perforated part) suspended by four beams allowing the movement according to X (blue axis) and Y (red axis), to detect rotations around Z (green axis). The drive mode (bottom left) is activated by means of the electrode combs on either side of the structure. The Coriolis force causes an oscillation of the sense mode (bottom right), which results in a change in the capacitance formed between the moving mass and the electrode underneath.

To convert these capacitance variations into an electrical signal, the variable capacitors are biased using a potentiometric readout. The main difference with the accelerometer case is that here the circuit can be biased with a constant voltage, because  $C^+$  and  $C^-$  are modulated around frequency  $f_d$ . In the case of a voltage output

$$V_{in} = -\frac{V_b}{2} \times \frac{Y}{g} (3.21)$$

and for a current output

$$I_{in} = -C_0 V_b \frac{d}{dt} \left( \frac{Y/g}{1 - Y^2/g^2} \right) \approx -C_0 V_b \times \frac{d}{dt} \left( \frac{Y}{g} \right) \quad (3.22)$$

Remarks :

1 - The sensitivity of a split-mode gyroscope depends not only on the mass and stiffness of the sense mode (like a conventional accelerometer), but also on the excitation force and the dissipation coefficient of the drive mode.

2 - The drive mode can be set to oscillate at its resonance frequency with a voltagecontrolled oscillator (VCO), whose frequency is set so that the phase between driving force  $F_d$  (at the output of the oscillator) and measured displacement X is 90° (which corresponds to resonance, see Fig. 3.4). This obviously requires to transduce X into a voltage.

3 - A differential capacitive detection scheme is obviously not possible with the (simplistic) structure shown in Fig. 3.6. To achieve this, one could fabricate two similar structures side by side, and drive them in phase opposition. In this way, the two structures would see the same Coriolis force, but with opposite signs. Consequently, the displacements according to Y and the corresponding capacitance variations would also be equal of equal magnitudes, but opposite signs.

### 3.2.2 Noise

The same analysis scheme as for accelerometers can be applied to gyroscopes. It is the thermal noise of the sense mode that determines the noise floor, but the noise of the electronics is often a few orders of magnitude above this limit, at least for split-mode gyroscopes. It is therefore electronic noise which determines the sensor resolution.

Taking into account thermomechanical noise, displacement Y is written

$$Y = -2M \frac{F_d}{K_s B_d} \Omega + \frac{F_{MEMS,n}}{K_s} \qquad (3.23)$$

which means that, on the bandwidth of interest  $\Delta f$ , we have

$$\overline{\Omega^2_{MEMS,n}} = 2k_BT \times \frac{B_s}{M^2} \times \frac{B_d^2}{F^2} \times \Delta f \ [(rad/s)^2] \quad (3.24)$$

The noise contribution of the AFE in the case of a voltage output is

$$\overline{\Omega_{AFE,n}^2} = \frac{1}{\left(\frac{V_b}{2g} \times \frac{2M}{K_s} \times \frac{F}{B_d}\right)^2} \times \overline{V_{AFE,n}^2} \left[ (rad/s)^2 \right]$$
(3.25)

where  $\overline{V_{AFE,n}^2}$  is the voltage noise of the AFE on the  $[f_d - \Delta f, f_d + \Delta f]$  frequency band, referred to the input. For a current output, we find

$$\overline{\Omega_{AFE,n}^2} = \frac{1}{\left(C_0 \omega_d \times \frac{V_b}{g} \times \frac{2M}{K_S} \times \frac{F}{B_d}\right)^2} \times \overline{I_{AFE,n}^2} \left[ (rad/s)^2 \right]$$
(3.26)

where  $\overline{I_{AFE,n}^2}$  is the current noise of the AFE on the  $[f_d - \Delta f, f_d + \Delta f]$  frequency band, referred to the input.

Remarks:

1 - The drive mode is also subject to noise: thermomechanical noise, but also electronic noise (generated in the oscillator loop), which obviously affects the gyroscope's performance.

2 - The sense mode of a gyroscope can be controlled to improve its response time and/or increase its operating range, as in the case of accelerometers.

### 3.2.3 Mode-matched gyroscopes

Some high-end gyroscopes rely on the use of two modes with the same resonance frequency. These are called "degenerate mode" or "mode-matched" gyroscopes. In this case, (3.19) can be replaced by

$$\dot{Y} = -2M \frac{F_d}{B_s B_d} \Omega \qquad (3.27)$$

Consequently, the amplitude of the sense displacement *Y* is  $Q_s \gg 1$  times greater than in the case of a "separate mode" gyroscope.

Two modes can be considered degenerate if they cannot be spectrally resolved, i.e.  $|f_s - f_d| \ll f_s/Q_s$ . This condition is very difficult to achieve in practice: the frequencies of the vibration modes must not only have the same value, but also the same temperature dependence; otherwise active drift compensation strategies must be implemented (again, a form of feedback-control). Furthermore, the use of degenerate modes restricts the bandwidth of  $\Omega$  to  $f_s/Q_s$  as opposed to  $f_s - f_d$  in the case of separate modes. This is again a form of compromise between sensitivity and bandwidth.

If the gyroscope has degenerate modes, one may verify that the variance of the inputreferred thermomechanical noise is still given by (3.24). Thus, as far as the noise floor is concerned, there is therefore no difference between separate or degenerate mode gyroscopes. On the other hand, the variance of the noise due to the electronics is reduced by a factor of  $Q_s^2$ compared to (3.25) or (3.26).

## 3.2.4 Quadrature error

A major difficulty encountered in the design of gyroscopes (with separate modes or not) is the existence of a so-called "quadrature" error. This error is caused by a parasitic oscillation at frequency  $f_d$  of the sense mode in the absence of rotation ( $\Omega = 0$ ). This oscillation is itself the result of a coupling between the drive and sense directions (anisoelasticity) related to manufacturing or design defects.

In the presence of such non-idealities, (3.16b) becomes

$$K_{s}(Y+\eta X) + B_{s}\dot{Y} + M\ddot{Y} = -2M\Omega\dot{X} \quad (3.28)$$

If there is no rotation, the error signal (in the "separate mode" case) can be written

$$Y_{err} = -\eta X \qquad (3.29)$$

which should be compared to the displacement due to the Coriolis force, given by (3.19), which is rewritten here as

$$Y_{Coriolis} = -2\frac{M}{K_s}\dot{X}\Omega \quad (3.30)$$

On the one hand, we can see that the two terms  $Y_{Coriolis}$  and  $Y_{err}$  are in quadrature, which, in theory, is sufficient to discriminate by demodulation the useful signal from the error signal. This is easy as long as the amplitude of the error signal  $Y_{err}$  is less than that of the largest signal  $Y_{Coriolis}$  that must be measured (both signals must pass through the same amplification chain). We have, from (3.29) and (3.30)

$$\frac{|Y_{err}|}{|Y_{Coriolis}|} = \frac{\eta |X|}{2\frac{M}{K_s}\omega_d |X|\Omega} = \frac{\eta}{2}\frac{\omega_s^2}{\omega_d\Omega} \quad (3.31)$$

which means that the coefficient of anisoelasticity must at least verify

$$\eta \le 2\frac{\Omega}{\omega_d} \times \frac{\omega_d^2}{\omega_{s^2}} \quad (3.32)$$

In practice, coefficient  $\eta$  can be adjusted post-manufacturing by "trimming" techniques or through the application of electrostatic forces that will "straighten" the structure.

Exercises

1 - Resume the study of quadrature error in the case of "degenerate modes". Is the problem the same? Is it worse? Is it improved?

2 - What is the Analog Devices accelerometer with the lowest noise? On which range (of acceleration, frequency and temperature) does it operate? How does temperature affect its metrological performance? How much does this component cost?

3 - Repeat the same study for STMicroelectronics and Bosch accelerometers.

4 - Repeat the same study for the gyroscopes of these three companies.

## 3.3 References

A similar treatment of MEMS accelerometers and gyroscopes can be found in:

[1] V. Kaajakari, "Practical MEMS", Small Gear Publishing, 2009.

Gyroscopes in particular are discussed in detail in

[2] C. Akar, A. Shkel, " MEMS Vibratory Gyroscopes. Structural Approaches to Improve Robustness ", Springer, 2009

# 4 Physics of MEMS

## 4.1 Introduction

The approach developed in part 3 makes it possible to translate specifications expressed at the "application" level in terms of performance (sensitivity, range, response time, resolution) into specifications at the "system" level, expressed in terms of gains, voltage or current noise and various non-idealities (for the electronic part), in terms of stiffness, damping coefficient, mass, capacitance, forces (for the "electromechanical" part).

A designer of integrated sensors must be able to translate these "system" specifications into "physical" terms: for example, what geometry to give to a MEMS structure to achieve a certain stiffness, a certain damping coefficient, a certain resonance frequency? The same type of work is obviously necessary for the design of the electronic part, but this is largely outside the scope of this course, dedicated more specifically to the MEMS transducer than to its associated electronics.

The design of a MEMS transducer is complex for many reasons. First of all, because MEMS are multi-physical devices, which require mechanical, electrostatic, fluidic and thermal manipulation. The physical phenomena are coupled, more or less strongly, which forces design compromises: for example, reducing the gaps between the moving part and the fixed part of a MEMS structure increases the efficiency of capacitive transduction, but accentuates the influence of the inter-electrode air film on the damping coefficient of the system. Similarly, the geometry of a gyroscope suspensions must allow the displacement of a mass in two orthogonal directions, while minimizing thermoelastic or radiation losses at the anchoring points.

The designer must also have a good knowledge of the technologies at his disposal to fabricate the transducer: obviously of the constraints they impose on the minimum dimensions of the structures or gaps, but also dispersions and other non-idealities (residual stresses, stress gradients, etc.) likely to affect the characteristics of the post-manufactured structure. The packaging that houses the transducer also has a considerable influence on the behavior, both static and dynamic, of MEMS structures and is particularly crucial for their long-term behavior. These technological aspects are outside my area of expertise and are therefore not covered in detail in this handout.

In what follows, we are interested in MEMS physics seen in an idealized light, i.e. forgetting at first most of the imperfections due to the technology. This design predicate effectively makes it possible to translate "system" specifications into "physical" or "geometrical" specifications. Once this step has been completed, it will be possible to validate the transducer design using numerical simulation tools at a lower level of abstraction (e.g. finite elements). This will make it possible to confirm (or invalidate) the simplifying assumptions made to arrive at the "physical" specifications and to verify, by simulation, the influence of the technological manufacturing process, and of the encapsulation, on the performance of the device. These numerical tools are presented in Part 5.

# 4.2 Mechanics of MEMS

## 4.2.1 Fundamental principles

Most MEMS structures are made of silicon, doped enough to be considered conductive. Silicon is a material that is almost perfectly elastic up to its yield strength, i.e. there is a linear relationship between the geometrical deformations of the medium, described by the tensor  $\epsilon$ , which is the symmetrical part of the displacement gradient, and the internal forces, or stresses, described by the tensor.  $\sigma$ .

This relationship, known as Hooke's Law, is written for an isotropic medium

$$\boldsymbol{\sigma} = \frac{E}{1+\nu} \left( \boldsymbol{\epsilon} + \frac{\nu}{1-2\nu} Tr(\boldsymbol{\epsilon}) \boldsymbol{I} \right) \qquad (4.1)$$

where *E* is Young's modulus ( $E \approx 160$ GPa for Si) and  $\nu$  is Poisson's ratio ( $\nu \approx 0.22$  for Si). Hooke's law is simplified into

$$\sigma_{xx} = E\epsilon_{xx} \quad (4.2-a)$$
  
$$\epsilon_{yy} = \epsilon_{zz} = -\nu\epsilon_{xx} \quad (4.2-b)$$

in the case where the medium is subject to uni-axial stress in direction x (Fig. 4.1). Equation (4.2-b) describes the lateral expansion of the material when it is compressed according to x or its contraction when it is stretched. Equation (4.2-a) describes a "spring" type behaviour,  $F = -K_{axial}X$  where

$$K_{axial} = \frac{ES}{L} \qquad (4.3)$$

with *L* the dimension along which the stress is applied, and *S* the surface where it is applied (Fig. 4.1).



Fig. 4.1 - Uni-axial deformation of an elastic material (the force  $F_x$  applies on both sides).

Writing Newton's second law in the medium

$$div(\boldsymbol{\sigma}) + \boldsymbol{f} = \rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2}$$
 (4.4)

where **u** is the displacement field (from which the deformations derive), **f** a volume force and  $\rho$  the density of the material ( $\rho = 2330$ kg/m<sup>3</sup> for Si), and combining with (4.1), we obtain the Navier's equation

$$\frac{1}{2}\frac{E}{1+\nu}\left(\Delta \boldsymbol{u} + \frac{1}{1-2\nu}grad(div(\boldsymbol{u}))\right) + \boldsymbol{f} = \rho \frac{\partial^2 \boldsymbol{u}}{\partial t^2}$$
(4.5)

This is a good starting point for a numerical validation of a choice of geometry of a MEMS transducer, by the finite element method for example. However, this formulation is far too general (and the solution tools it requires are far too costly) to be effectively used for the design and dimensioning of a MEMS structure. In a first approach, it is more interesting to design a MEMS device by reducing it to a set of simple components whose behavior is more easily modeled: rigid masses (as in inertial sensors), membranes (as in pressure sensors or microphones), flexures/suspensions (of inertial sensors) and beams (resonant strain gauges, for example).

Remarks:

1 - The influence of temperature T on the behavior of MEMS structures can be taken into account by modifying (4.1) as follows :

$$\boldsymbol{\sigma} = \frac{E}{1+\nu} \left( \boldsymbol{\epsilon} + \left( \frac{\nu}{1-2\nu} Tr(\boldsymbol{\epsilon}) - \frac{1+\nu}{1-2\nu} \alpha (T-T_0) \right) \boldsymbol{I} \right)$$
(4.6)

where  $T_0$  is a reference temperature and  $\alpha$  the coefficient of linear expansion of the material ( $\alpha \approx 2.6 ppm/K$  for silicon).

2 - In this form, Hooke's law makes it possible, for example, to take into account a variation in the global temperature of the structure (such as that due to a change in the ambient temperature): an increase in temperature thus results in an expansion of the structure if it is free to deform, or an increase in its state of stress in the opposite case.

3 - It also makes it possible to account for local temperature variations resulting from deformations and the thermoelastic losses they generate.

4 - The values of E and v given above are valid for polysilicon, which can be considered as an isotropic material. Monocrystalline silicon, on the other hand, is anisotropic. Its Young's modulus typically varies between 130GPa and 170GPa depending on the considered orientation.

## 4.2.2 Beams and suspensions

A beam is a quasi-monodimensional solid, one dimension of which (the length L) is very large with respect to the other two (the width b, the height h). This section is restricted to the case of straight beams with rectangular cross-section under certain types of loads, as shown in Fig. 4.2.



*Fig.* 4.2 - *Geometry of the beam (left) and considered load cases (right). From top to bottom: pure traction, pure bending and pure torsion.* 

#### Remark:

In this section we avoid talking about "thickness". Thickness is a technological parameter designating the thickness of the structural layer in which the MEMS structure is manufactured (in the case of surface micromachining). A MEMS structure can bend according to its thickness (out-of-plane) or perpendicularly to it (in-plane).

#### 4.2.2.1 Bending of beams

Under the effect of a moment of force around the y axis or under the effect of a force according to z, the beam bends. This bending is governed by Euler-Bernoulli's theory, valid for small deformations of slender beams.



Fig. 4.3 - Finite element simulation of the bending of a girder fixed at both ends (bridge). At the top, initial mesh of the beam without stress. In the middle, deformation of the girder under the effect of a uniform force applied to the upper side. At the bottom, contours of the beam deformation (red = elongation, blue = shortening).

Euler-Bernoulli's theory is based on the kinematic hypothesis that the sections that are straight and perpendicular to the beam axis remain straight and perpendicular during bending. Consequently, bending results in deformations (and constraints) in the *x* direction which vary linearly according to *z* and of opposite signs on either side of the beam axis (Fig. 4.3, see also appendix B). In this framework, the shape of the beam is entirely determined by the shape w(x, t) taken by its axis. This is obtained by writing the balance between the external forces and moments applied to the beam and the internal forces and moments, resulting from stresses or inertia. In a very general way, we have

$$EI\frac{\partial^4 w}{\partial x^4} - N\frac{\partial^2 w}{\partial x^2} + \rho S\frac{\partial^2 w}{\partial t^2} = f + \sum_i F_i \delta(x_{F_i}) + \sum_j M_j \delta^{(1)}\left(x_{M_j}\right)$$
(4.7)

where  $I = bh^3/12$  is the moment of inertia of the beam, S = bh its section, N the "normal" force along the axis of the beam, f(x, t) the linear density of force applied along the beam, the  $F_i$  are forces concentrated at the points  $x_{F_i}$ , the  $M_j$  are couples concentrated at the points  $x_{M_i}$ .



Fig. 4.4 - Examples of MEMS structures using beams and suspensions operating in bending. Anchor points are circled in red, the main direction of motion is indicated in green. Top: cantilever (left) and bridge (right). Middle: accelerometer suspended by 4 serpentine flexures (left) and detail of a flexure (right). Bottom: SEMATECH micro-switch suspended by four serpentine flexures (left) and Draper Labs double mass gyroscope suspended by "folded-beam" flexures (right).

Some boundary conditions that can be associated with (4.7) are

$$w = 0 \quad (4.8-a), \qquad \frac{\partial w}{\partial x} = 0 \qquad (4.8-b),$$
$$\frac{\partial^2 w}{\partial x^2} = 0 \quad (4.8-c), \qquad \frac{\partial^3 w}{\partial x^3} = 0 \qquad (4.8-d)$$

which correspond respectively to zero displacement, zero slope, no couple around y or force according to z at either end of the beam. Three common boundary conditions are

- fixed end: zero displacement (4.8-a) and zero slope (4.8-b). The term "bridge" is used to describe a beam fixed at both ends.
- free end: zero moment (4.8-c) and zero force (4.8-d). A beam fixed at one end and free at the other end is called a "cantilever".
- sliding end: zero slope (4.8-b) and zero force (4.8-d). Many inertial sensor suspensions may be modeled as assemblies of fixed-sliding beams.

Examples of MEMS beams and suspensions are shown in Fig. 4.4. More in-depth considerations about boundary conditions can be found in Appendix B.

#### Suspensions

The suspensions of MEMS structures, such as accelerometers or gyroscopes, are often composed of straight segments, similar to beams working essentially as flexures, connected at their ends by rigid segments (Fig. 4.4). Appendix C provides some beam design elements that make it possible to go a little further than this first approximation.

The use of "serpentine" suspensions instead of straight ones not only makes the design more compact, but also minimizes the influence of normal stress, which is difficult to control in practice, on the stiffness of the structure. For example, a rise in ambient temperature would result in the appearance of a normal compressive stress in the straight suspensions of the structure in Fig. 4.5-a (the beams seeking to expand, but not being free to deform), and thus in a variation of stiffness. On the opposite, a serpentine geometry, as in Fig. 4.5-b, allows the suspension to deform and limits the occurrence of temperature-related normal stress. Moreover, the surface area occupied by the structure in Fig. 4.5-b is less, for an equivalent stiffness.



Fig. 4.5 - Two accelerometer designs with identical sensitivity.

As the inertia of these suspensions is negligible compared to the inertia of the mass suspended from them, they can be represented using a static model. Each of the segments is assimilated to a fixed-sliding beam subject to a force at its end. Thus, the shape taken by each segment is a solution of :

$$EI\frac{d^{*}w}{dx^{4}} = F\delta(L) \qquad (4.9-a)$$

$$w(0) = 0 \qquad (4.9-b), \qquad \frac{dw}{dx}\Big|_{x=0} = 0 \qquad (4.9-c),$$

$$\frac{dw}{dx}\Big|_{x=L} = 0 \ (4.9-d), \qquad \frac{d^{3}w}{dx^{3}}\Big|_{x=L} = 0 \qquad (4.9-e)$$

We find that the displacement w(L) at the free end of the segment is given by  $F = K_{segment} \times w(L)$  where

$$K_{segment} = 12 \frac{EI}{L^3} \qquad (4.10)$$

is the stiffness of the segment. The stiffness  $K_{suspension}$  of each suspension is obtained by considering that the individual segments are springs placed "in series", end to end

$$\frac{1}{K_{suspension}} = \sum_{i} \frac{1}{K_{segment_i}}$$
(4.11)

Finally, the total stiffness of the structure (in the direction of bending)  $K_{total}$  is the sum of the stiffness of each suspension (often 4, sometimes 2)

$$K_{total} = \sum_{j} K_{suspension_{j}}$$
(4.12)

#### 4.2.2.1.1 Cantilevers

For a cantilever whose inertia can be neglected, subject to a force at its extremity, Euler Bernoulli's equation is written as follows

$$EI\frac{d^{*}w}{dx^{4}} = F\delta(L) \qquad (4.13-a)$$

$$w(0) = 0 \qquad (4.13-b), \qquad \frac{dw}{dx}\Big|_{x=0} = 0 \qquad (4.13-c),$$

$$\frac{d^{2}w}{dx^{2}}\Big|_{x=L} = 0 \ (4.13-d), \ \frac{d^{3}w}{dx^{3}}\Big|_{x=L} = 0 \qquad (4.13-e)$$

which results in an equivalent stiffness of

$$K_{cantilever} = 3\frac{EI}{L^3} \qquad (4.14)$$

Note that a fixed-sliding beam of length L can be seen as two cantilevers of length L/2 put end-toend. This makes it possible to simply deduce (4.10) from (4.14), and vice versa.

From a practical point of view, this static model may be useful for the design of certain accelerometers (those made using volume micromachining techniques) or energy harvesters, whose suspensions can be approximated as cantilevers supporting a large mass. The sensing element of an atomic force microscope in "contact" mode can also be modeled in this way.



Fig. 4.6 - "NEMS" cantilever for mass sensing fabricated by CEA-LETI (taken from the thesis of S. Labarthe). Oscillation is achieved by means of electrostatic forces. The two beams on either side of the base of the cantilever work in tension/compression. They are in fact piezoresistive nano-gauges allowing the detection of movement.

It is the resonant behavior of cantilevers which is most often implemented. For example, some mass sensors rely on cantilevers that have been "functionalized" so that the molecules to be detected are adsorbed onto them (Fig. 4.6). This results in a variation of the density of the cantilever, and thus of its resonance frequency. This resonance frequency can then be measured by using the cantilever as the resonant element of an oscillator (see part 3 and the drive mode of the gyroscopes). The beam, subject to a drive force generally distributed over its length and its own inertia, is governed by

$$EI\frac{\partial^4 w}{\partial x^4} + \rho S\frac{\partial^2 w}{\partial t^2} = f \quad (4.15)$$

where f is the linear density of force (both that of the drive and that exerted by the surrounding fluid). The fundamental resonance frequency of such a system corresponds to that of the first eigenmode of the beam (see Appendix D)

$$\omega_{cantilever} \approx 1.015 \times \sqrt{\frac{E}{\rho}} \times \frac{h}{L^2}$$
 (4.16)

The relative change of cantilever frequency is thus related to that of the density and that of the mass by

$$\frac{\Delta\omega_{cantilever}}{\omega_{cantilever}} = -\frac{1}{2} \times \frac{\Delta\rho}{\rho} = -\frac{1}{2} \times \frac{\Delta m}{m} \qquad (4.17)$$

Thus, the sensitivity of this type of sensor is all the higher as the nominal mass of the resonator is low. This possibility of realizing ultra-sensitive gas sensors has largely motivated the enthusiasm for NEMS technologies.

Resonant cantilevers are also used in various microscopy techniques, such as atomic force microscopy (in "non-contact" mode).

#### 4.2.2.1.2 Bridges

Although bridges are rarely used as suspensions, it is interesting to study from such an example the influence of normal stress on the behaviour of structures. The deformation of a bridge subject to static loading at its center and normal stress N is governed by

$$EI \frac{d^4 w}{dx^4} - N \frac{d^2 w}{dx^2} = F\delta\left(\frac{L}{2}\right) \quad (4.18\text{-a})$$

$$w(0) = 0 \quad (4.18\text{-b}), \quad \left.\frac{dw}{dx}\right|_{x=0} = 0 \quad (4.18\text{-c}),$$

$$w(L) = 0 \quad (4.18\text{-d}), \quad \left.\frac{dw}{dx}\right|_{x=0} = 0 \quad (4.18\text{-e}),$$

Assuming the normal force is low, i.e.  $NL^2/EI \ll 1$ , we can look for an approximate solution to (4.18) in the form

$$w \approx w_{bend} + \frac{NL^2}{EI} w_{corr} (4.19)$$

where  $w_{bend}$  and  $w_{corr}$  verify boundary conditions (4.18-b,c,d,e) and

$$EI\frac{d^4w_{bend}}{dx^4} = F\delta\left(\frac{L}{2}\right) \quad (4.20\text{-a})$$
$$\frac{d^4w_{corr}}{dx^4} = \frac{1}{L^2} \times \frac{d^2w_{bend}}{dx^2} \quad (4.20\text{-b})$$

This results in a stiffness equal to

$$K_{bridge} = 192 \frac{EI}{L^3} \times \frac{1}{1 - \frac{1 N L^2}{40 EI}} \qquad (4.21)$$

The case N = 0 corresponds to "pure" bending: it is notable that the same solution can be obtained by considering a bridge of length *L* as two fixed-sliding "segments" of length *L*/2 loaded at their end (4.10). When N > 0 (tensile stress), the effective stiffness of the bridge increases. It decreases if N < 0 (compressive stress).



*Fig.* 4.7 - *Schematic diagram of a 2-axis resonant accelerometer (left) and practical realization (right). What are the differences? Work carried out at Politecnico di Milano in collaboration with ST.* 

This sensitivity of bridge stiffness to normal stress is used for making resonant sensors. Fig. 4.7 shows the operating principle of a resonant accelerometer: under the effect of acceleration, the mass moves. This generates normal forces (proportional to the acceleration) of opposite signs in the resonators, which can be modeled as bridges, and thus a variation in their stiffness and resonance frequency. Each bridge, subject to a drive force generally distributed over its length and to its own inertia, is governed by

$$EI\frac{\partial^4 w}{\partial x^4} - N\frac{\partial^2 w}{\partial x^2} + \rho S\frac{\partial^2 w}{\partial t^2} = f \quad (4.22)$$

where *f* is the linear density of force (both that of the drive and that exerted by the surrounding fluid). The fundamental resonance frequency of the bridge is given by (see Appendix D)

$$\omega_{bridge} \approx 6.459 \times \sqrt{\frac{E}{\rho} \times \left(1 + 0.0246 \frac{NL^2}{EI}\right)} \times \frac{h}{L^2}$$
(4.23)

The relative variation of the bridge frequency is therefore given by

$$\frac{\Delta\omega_{bridge}}{\omega_{bridge}} \approx 0.0123 \times \frac{NL^2}{EI}$$
(4.24)

Note that it is necessary to be able to discriminate frequency variations due to variations in the physical quantity to be measured (e.g. acceleration) from those due to temperature. For example, the heating of a bridge results in the appearance of a normal force equal to

$$N = -ES \times \alpha (T - T_0)$$
 (4.25)

opposing exactly (under the hypothesis of perfectly rigid anchors) its expansion in the direction x. This normal stress due to thermal expansion justifies the choice of a differential measurement for the accelerometer in Fig. 4.7.

#### Remarks:

1 - The manufacturing processes of MEMS devices result in structures in which stresses (called "residual stresses") remain after release. When the stresses are positive, the devices are stiffer than expected (this is the case of the bridge in Fig. 4.4 whose resonance frequency was 33% higher than predicted with the pure bending model). When the stresses are negative, the devices are less stiff, which may possibly lead to buckling. Buckling is the "catastrophic" collapse of a slender structure under normal compressive stress. The buckling point is reached when the fundamental resonance frequency of the bridge is zero ( $\omega_{bridge} = 0$ ) i.e.

$$N \approx -40.72 \times \frac{EI}{L^2} \qquad (4.26)$$

according to (4.23). An exact calculation gives as a critical value of the normal force  $N = -4\pi^2 \times EI/L^2$  which is very close to (4.26).

2 - One of the main causes for the existence of residual stresses is the difference in the coefficients of expansion of the layers of materials deposited during the fabrication of the device. To quantitatively assess the influence of temperature on the mechanical behavior of a MEMS device (including post-fabrication), it is therefore necessary to take into account the thermomechanical behavior of its entire environment (including the packaging). From this point of view, (4.25), which assumes "perfectly rigid" anchors, must be taken with caution.

3 - A bending bridge sees its length increase (the shortest path between the two ends of the bridge being the straight line). Thus, any bending of a bridge is accompanied by an elongation, and thus an increase in the normal stress. Using Hooke's law, we have

$$N \approx \frac{ES}{2L} \int_0^L \left(\frac{\partial w}{\partial x}\right)^2 dx$$
 (4.27)

Consequently, the more a bridge deforms, the stiffer and more resistant to deformation it becomes. This "hardening" phenomenon results in a non-linear spring behavior ( $F = -K \times (1 + \gamma X^2) \times X, \gamma > 0$ ).

4 - These stresses (residual, thermal, elongation) affect, to some extent, all MEMS structures that are geometrically similar to a bridge. Many mechanical design techniques can reduce their impact, such as the use of right-angled beams ("crab-leg", as shown in Fig. 3.6), serpentine or "box" beams, or the fabrication of the structure around a single anchor point (Fig. 4.9).



Fig. 4.9 - 2-axis accelerometer from ST with capacitive transduction, with a single anchor point (center). The similarity with the structure in Fig. 4.7 is striking. This accelerometer equipped the IPhone 4 (image taken from MEMS Journal).

5 - Structures that are geometrically similar to cantilevers (with a single anchor point, therefore) are sensitive to the stress gradient existing in the thickness of the material. If such a gradient exists, the structure after release is bent out of plane (Fig. 4.10).

6 - Regardless of the boundary conditions, the stiffness of a bending beam is proportional to its moment of inertia.  $I = bh^3/12$  and inversely proportional to  $L^3$  (4.10) (4.14) (4.21). To minimize the stiffness of a beam, it is therefore sufficient in theory to make it as long and as thin as the manufacturing process allows. Many practical considerations, some of which, such as compactness, have already been mentioned, limit this simplistic reasoning:

- devices with minimum dimensions are more prone to manufacturing dispersion (the difference between the predicted dimension and the actual manufactured dimension, which varies randomly from one device to another). This is obviously not desirable if uniformity of sensor performance is to be guaranteed.
- a thin and long beam has an increased sensitivity to normal stress (4.21) (4.23), and thus to temperature (in particular).
- some dissipation mechanisms (such as surface losses) are all the more marked as the flexible parts of the structure are thin.



Fig. 4-10 - Consequence of a stress gradient in a MEMS structure.

### 4.2.2.2 Tension and torsion

The stiffness of a beam in tension/compression (subject to normal stress alone, along the beam axis, and deforming in the direction of that stress) is given directly by Hooke's law

$$K_{axial} = \frac{ES}{L} \qquad (4.28)$$

Finally, the following formula gives the torsional stiffness of a beam used as a torsion spring, fixed at one end and subject to a couple (around the beam axis) at the other end

$$C_{torsion} = \frac{GJ}{L}.$$
 (4.29)

where G = E/2(1 + v) is the torsional modulus of the material, and *J* the torsional moment of the beam around the axis *x*. For a beam with a rectangular cross-section, we have

$$J = Dd^{3} \times \left(\frac{1}{3} - 0.21 \frac{d}{D} \left(1 - \frac{1}{12} \frac{d^{4}}{D^{4}}\right)\right) \quad (4.30)$$

where  $D = \max(b, h)$  is the larger lateral dimension and  $d = \min(b, h)$  the smaller.



Fig. 4.11 - 1-axis ST accelerometer with capacitive transduction (left, same source as Fig. 4.9) and operating principle (Politecnico di Milano).

#### Remark

In some rare cases, it may be necessary to take into account the inertia of a beam working in tension or torsion, or the existence of forces distributed along it. The following models can be used

$$ES\frac{\partial^2 u}{\partial x^2} + f = \rho S\frac{\partial^2 u}{\partial t^2} \quad (4.31)$$

for a beam in tension/compression (the displacements u and forces f are according to x) and

$$GJ\frac{\partial^2\theta}{\partial x^2} + m = \rho S\frac{\partial^2\theta}{\partial t^2} \quad (4.32)$$

for a torsion beam (the angles  $\theta$  and the moments m are around x).

#### 4.2.2.3 Beam assemblies and complex suspensions

Please refer to Appendix C of this document.

### 4.2.3 Plates and membranes

To be completed (plates, membranes, Von Karman, adaptive optics, pressure sensor, microphone).

# 4.3 Electromechanical transduction

## 4.3.1 General overview

The following sections focus on electromechanical transduction at the MEMS scale: how to convert a motion into an electrical signal? How to electrically generate forces to set a MEMS structure in motion (or attenuate the motion of a MEMS structure)? In an over-simplified way, sensing the motion or deformation of a structure is useful for "sensor" type applications (accelerometer, pressure sensor), the generation of forces for "actuator" type applications (e.g. micro-switches, micro-mirrors, DMD Fig. 4.12). However, it is not uncommon to combine sensing and actuation within the same application, as in the following examples.



Fig. 4.12 - Digital Mirror Display (DMD) from Texas Instruments, at various magnification levels. The pixels (which are supported by torsion bars) are turned on or off by means of electrostatic forces.

**Test** - Most commercially available MEMS inertial sensors have a built-in self-test function to verify their functionality throughout their lifetime. This test consists in applying a predetermined stimulus to the sensing element using an electromechanically generated force. If certain characteristics of the response deviate from the specifications, then the measurements made by this sensor can be rejected, or corrected according to the test results. A typical test for an inertial sensor is to apply a force step to determine, based on the step response of the structure, its stiffness *K* and its damping coefficient *B* (the mass *M* being unlikely to change over time). This makes it possible to check the airtightness of the packaging over time, or the wear of the structure.

**Feedback control** - To increase the measurement range of a sensor and improve its dynamic behavior or linearity, it is sometimes essential to use feedback control techniques. In the absence of feedback control, the measuring range of a capacitive accelerometer, with mass M, stiffness K and gap g is strictly limited to  $\pm Kg/M$ . This is the value of the acceleration for which the mass would come into contact with the opposite electrodes. In a feedback-controlled accelerometer, a force F is exerted on the mass in order to keep it at a reference position, typically X = 0. In this way, F opposes exactly the force of inertia (and thus provides a measure of it). The measuring range is then limited to  $\pm F_{max}/M$  where  $F_{max}$  is the maximum value of the force that can be applied to the structure. This is interesting because  $F_{max} > Kg$  is a design objective that can be easily achieved at the MEMS scale, for example using electrostatic actuation.

**Oscillators and resonant sensors** - Finally, micromechanical oscillators, such as time references (Fig. 4.13) and all "resonant" sensors in the broad sense (amplitude-modulated gyroscopes, as seen in Part 3, frequency-modulated mass sensors or resonant accelerometers as seen in the previous sections), are devices that intrinsically combine detection and actuation. Indeed, to set a mechanical structure in oscillation at its resonance frequency, it is necessary to apply a "drive" force in quadrature with its displacement. Thus, one must be able to sense the

motion of the structure, on the one hand, and to generate a force according to a certain control law, on the other hand.



Fig. 4.13 - SiTime's silicon resonators are becoming increasingly fierce competitors to quartz resonators, which are larger in size and more costly, for the generation of ultra-stable clock signals. Images and characteristics taken from the SiTime site. Bottom right, block diagram of the MEMS oscillator.

## 4.3.2 Detection

Capacitive and resistive detection have already been discussed in part 2 from an electronic point of view. In the next two sections, their implementation in the specific field of MEMS is discussed.

### 4.3.2.1 Capacitive detection

Capacitive detection is the sensing principle most commonly implemented in MEMS devices. It requires the fabrication of a fixed electrode in close proximity to the structure whose motion is to be measured. The electrode and the structure can be realized in the same step of the manufacturing process, which makes capacitive sensing a "cheap" method, in addition to being efficient. For example, one can compare the relative complexity of MEMSCAP's SOIMUMPS and PiezoMUMPS processes, in Appendix H, adapted respectively to capacitive and piezoelectric sensing and actuation.

The plane capacitor approximation is to the electrostatics of MEMS what beam theory is to the mechanics of MEMS: an extremely useful starting point for a first drafting of a system using an analytical approach, to be refined in a validation phase by more precise models (analytical or numerical, see Appendix E).

For a rectangular plane capacitor, with gap g and side dimensions  $L_1$  and  $L_2$  ( $L_{1,2} \gg g$ ) we have

$$C = \epsilon_0 \frac{L_1 L_2}{g} \qquad (4.33)$$

Due to the fact that  $L_{1,2} \gg g$ , this capacitance is mainly sensitive to changes in g, much less to those of  $L_1$  and  $L_2$ . Mathematically, we can write

$$\delta C \approx \left(\epsilon_0 \frac{L_2}{g} \delta L_1 + \epsilon_0 \frac{L_1}{g} \delta L_2 - \epsilon_0 \frac{L_1 L_2}{g^2} \delta g\right) = C \times \left(\frac{\delta L_1}{L_1} + \frac{\delta L_2}{L_2} - \frac{\delta g}{g}\right)$$
(4.34)

Thus, a displacement X of the mobile electrode will have a much greater impact on the value of C if it is in the direction of the gap ( $\delta g = X$ ) than in the lateral directions ( $\delta L_{1,2} = X$ ), since  $X/g \gg X/L_{1,2}$ . On the other hand, the maximum displacement is limited in one case to g, in the other to  $L_1$  or  $L_2$ . This is still a form of trade-off between sensitivity and measurement range (which can be overcome by feedback-control techniques in some cases). As a consequence, capacitive "gap-closing" detection is preferable to transduce small displacements into electrical signals, while surface variation is more appropriate to transduce large displacements.

In the case of accelerometers, for example, the fact that gap-closing detection is very widely used is justified by the fact that the displacements which must be detected are very small. The same applies to the detection of the sense mode of gyroscopes. In both cases, the displacements are proportional to the mass of the device, therefore to the surface occupied by the device, which is generally desired as small as possible to reduce the unitary cost of the sensor.

Surface variation detection is mainly used in resonant applications. For example, we saw in part 3 that a gyroscope is all the more sensitive as it oscillates with a large amplitude along the drive direction. Since this large amplitude motion must be transduced into an electrical signal (in order to make the structure oscillate at its resonance frequency), surface variation detection is adapted. In addition, resonant applications generally take advantage of a low dissipation coefficient. As a film of air dissipates much less energy per cycle when it is "sheared" (as in surface variation detection) than when it is "compressed" (as in gap-closing detection, see "slide-film" and "squeeze-film" in section 4.4), this also speaks in favor of surface variation detection for resonant applications limited by fluidic damping.

#### Remarks

1 - To increase the sensitivity of capacitance to displacement, "comb" structures are commonly used (see Fig. 4.4 for an example - bottom right - of surface variation detection, Fig. 4.9 and Fig. 4.14 for examples of gap-closing detection). For a given surface area, the number of "fingers" of a comb is limited primarily by the accuracy of the manufacturing process (which will determine the minimum gap between fixed and movable fingers, as well as the minimum width of the fingers).

2 - Many design trade-offs come into play to determine the parameters of a comb structure. In particular, any surface occupied by combs results in a decrease of the useful mass of the device, which is detrimental to inertial measurements. In addition, a compromise must also be found between the number of fingers per unit length and the stiffness of the fingers (their resonance frequency must be sufficiently high with respect to that of the structure to which they are attached).

3 – On the whole, a manufacturing process is all the more adapted to capacitive detection as it makes it possible to fabricate devices with narrow gaps and large lateral dimensions.  $L_{1,2}$ .



Fig. 4.14 - Mask of an Analog Devices accelerometer. The structure moves laterally under the effect of acceleration, as indicated by the double arrow. The central combs are used to detect the motion of the structure by differential gap-closing detection. Those on the sides are used to generate test stimuli.

#### 4.3.2.2 Piezoresistive detection

At the MEMS scale, resistive sensing - in contrast to capacitive sensing - does not rely on geometric effects (or only to a small extent). Indeed, silicon has quite marked piezoresistive properties: its resistivity varies with its state of stress much more than its geometry. One may plunge into the book by M.H. Bao for a precise description of the phenomenon and the calculation of piezoresistivity coefficients. Here, a more functional description of the phenomenon, borrowed from V. Kaajakari's book, is sufficient.

	n-type			p-type		
	[100]	[110]	[111]	[100]	[110]	[111]
$\pi_l \ [10^{-11} Pa^{-1}]$	-102.2	-31.2	-7.5	6.6	71.8	93.5
$\pi_t \; [10^{-11} \mathrm{Pa}^{-1}]$	53.4	-17.6	6.1	-1.1	-66.3	44.6

Table 4. 1 - Piezoresistivity coefficients of weakly doped silicon (dopant concentration less than $10^{17} cm^{-3}$ ). Taken from Kaajakari.

Piezoresistivity coefficients  $\pi_l$  and  $\pi_t$  (longitudinal and transverse) relate the variation in resistivity  $\rho$  in one direction and the constraints  $\sigma_l$  and  $\sigma_t$ , respectively according to and perpendicular to the considered direction

$$\frac{\delta\rho}{\rho} = \pi_l \sigma_l + \pi_t \sigma_t \tag{4.35}$$

The piezoresistance coefficients (in  $Pa^{-1}$ ) are given in Table 4.1 for common crystallographic orientations in the case of *n*-doped or *p*-doped silicon.

According to (4.35) and Hooke's law (4.1), a piezoresistive gauge of length l and lateral dimensions  $L_1$  and  $L_2$  undergoing a purely tensile stress  $\sigma$  along its length (and therefore a longitudinal deformation  $\epsilon = \sigma/E$ ) will thus see its resistance

$$R = \rho \frac{l}{L_1 L_2} \qquad (4.36)$$

vary from

$$\delta R = R \times \left(\frac{\delta \rho}{\rho} + \frac{\delta l}{l} - \frac{\delta L_1}{L_1} - \frac{\delta L_2}{L_2}\right) = R \times (\pi_l \times E + 1 + 2\nu) \times \epsilon$$
(4.37)

It can be verified that for the values of  $\pi_l$  data given in Table 4.1, the gauge factor

$$GF = \pi_l \times E + 1 + 2\nu \approx \pi_l \times E \tag{4.38}$$

is far greater than 1.

Remarks:

1 - This detection method has a "historical" importance, in the sense that it has now been replaced by capacitive detection in many applications (inertial in particular). The reasons for this are multiple:

- Temperature affects not only the piezoresistivity coefficients of silicon, but also its resistivity. Without a rigorous temperature compensation strategy, piezoresistive MEMS sensor measurements are not usable. This constraint is much less strong for capacitive sensing, which is essentially only affected by the geometrical consequences of temperature variations.
- the fabrication of piezoresistive strain gauges (by local implantation of p-charge carriers in n-doped silicon, for example, or vice versa) typically requires a few specific manufacturing steps, making the manufacturing process longer (and more expensive). On the opposite, the moving and fixed electrodes required for capacitive sensing are realized in a single step of the manufacturing process (at least for structures deforming in the plane).
- As pointed out in part 2, resistors are sources of noise. The same applies to piezoresistors. These generate not only white noise (Nyquist-Johnson noise), but also flicker noise (called Hooge noise, in the case of piezoresistors). The higher the voltage at the terminals of the piezoresistor, the higher the flicker noise. It is therefore not possible to reduce its influence by increasing the bias of the resistors. On the contrary, a capacitance is intrinsically noiseless.

2 - Nevertheless, piezoresistive detection is still used for sensors whose structure deforms out of plane, mainly pressure sensors, as in Fig. 4.15. CEA-LETI has also developed a piezoresistive detection technique for in-plane movements, particularly adapted for gauges at the nanometer scale (cf. mass sensor in Fig. 4.6).



Die Overview

Fig. 4. 15 - Melexis pressure sensor (image from the SystemPlus website). The square membrane in the center bends under the effect of pressure. This deformation is measured using 4 piezoresistive gauges. Those on the left and right are arranged in such a way that they are subject to longitudinal strain (with respect to the direction of current), those at the top and bottom to transverse strain (also with respect to the direction of current). One of the gauges is highlighted in the inset at the bottom left.

### 4.3.2.3 Other detection methods

Other detection principles can be implemented at the MEMS scale to transduce a displacement into an electrical signal:

- Piezoelectric detection: typically using silicon as the structural material and gauges of piezoelectric material to detect its movement (e.g. MEMSCAP's PiezoMUMPS process). It has several advantages (linearity, low intrinsic noise, no need for bias) over piezoresistive and/or capacitive sensing, but is generally more expensive (due to the larger number of manufacturing steps) and is limited to detecting certain types of motion. It is also possible to use a piezoelectric material (instead of silicon) as a structural material. Quartz, for example, makes it possible to fabricate devices, generally resonant (inertial sensors, as at ONERA, clocks...) with a much better temperature stability than silicon, but with larger dimensions and (therefore) a much higher unitary cost.
- FET (Field-Effect Transistor) detection: this is a variant of capacitive detection, where the MEMS structure forms the moving gate of a transistor, so that a displacement of the structure modulates the current flowing through its channel. The very first MEMS device ("resonant gate transistor", Nathanson, Fig. 4.16) is based on this principle, which is only

rarely used commercially - again, mainly due to the complexity and cost of the required manufacturing processes.

- optomechanical detection: in this case, a MEMS resonator is coupled with an optical resonator, whose properties are modulated by the motion of the MEMS structure (Fig. 4.17). This type of device is used in the laboratory for fine material characterization or mass measurement.



Fig. 4.16 - The very first MEMS device actually consists in a "vibrating antenna": when the wavelength of the incident wave is tuned to the length of the beam, the beam enters into mechanical resonance, which modulates the current flowing through the transistor channel at the end. Image from patent US3413573A.



Fig. 4.17 - Principle of an optomechanical AFM: the displacement of the "cantilever" modulates the effective index of the disc (optomechanical coupling), and thus its natural frequencies; this results in a variation of the optical coupling coefficient between the disc and the fiber, and thus of the optical intensity at the fiber output. Work carried out at NIST and U. Maryland. Images taken from DOI: 10.1364/OE.20.018268.

### 4.3.3 Actuation

The physics of MEMS is particular in that surface phenomena play a much more important role than at the macroscopic scale: by reducing all the dimensions of an object by one order of magnitude, one divides the effect of volume forces by three orders of magnitude, the effect of surface forces by two orders of magnitude only. These scaling laws partly explain why MEMS sensors are so singular-looking compared to their macroscopic counterparts.

Capacitive actuation (or electrostatic actuation) is the main method used to apply forces in a controlled manner to MEMS structures, thus setting them in motion, controlling their position, etc. Because of its "surface" nature (it takes advantage of the electrostatic pressure on the surface of charged conductors), it is suitable for the microscopic scale. In addition, it is inexpensive to implement, requiring no special manufacturing step, which explains its great popularity. Most of this section is devoted to it.

#### *4.3.3.1 Capacitive actuation*

### 4.3.3.1.1 Principle



Fig. 4. 18 - Closed system consisting of an ideal voltage source and a capacitor.

The most intuitive way to understand electrostatic forces is to start from the system shown in Fig. 4.18, which describes a capacitor C to which a voltage V is applied. The potential energy of the system is written

$$E_{pot} = E_{source} + E_{capacitor} = -QV + \frac{1}{2}CV^2 \quad (4.39)$$

where Q is the charge provided by the source. As Q = CV, the potential energy is written
$$E_{pot} = -\frac{1}{2}CV^2$$
 (4.40)

Now assume that one of the armatures of the capacitor can move freely in one direction X, so that capacitance C depends on X. In the absence of voltage,  $E_{pot}(X) = 0$  and all positions are equivalent. On the other hand, if  $V \neq 0$  (and regardless of the sign of V), the system evolves towards a minimum of  $E_{pot}(X)$ , that is to say towards a maximum of C(X).

Physically, this reflects the existence of an electrostatic force acting in direction X, deriving from  $E_{pot}(X)$  according to

$$F_{elec} = -\frac{\partial E_{pot}}{\partial X} = \frac{1}{2} \frac{\partial C}{\partial X} V^2 \qquad (4.41)$$

Two configurations, gap-closing or surface variation (also called "comb-drive"), are particularly used.



Fig. 4.19 - Geometry of a "comb-drive". The red electrode is attracted inside the black electrode.

The typical geometry of a "comb-drive" is shown in Fig. 4.19. Using the plane capacitor approximation, one can write

$$C = 2\epsilon_0 \frac{XL_2}{g} \quad (4.42)$$

Using (4.41), we can deduce the following

$$F_{comb-drive} = \epsilon_0 \frac{L_2}{g} V^2 (4.43)$$

In this configuration, a force is generated that attracts the moving fingers between the fixed fingers. Remarkably, this force is independent of *X*. This holds as long as the plane capacitor hypothesis is verified, that is to say neither at the beginning (X = 0), nor at the end of the stroke ( $X = L_1$ ), and, obviously, provided *V* is independent of *X*.

"Comb-drive" actuation is used for the gyroscope in Fig. 4.4., for example



Fig. 4.20 - Gap-closing geometry. The red electrode is attracted to the black electrode.

The gap-closing geometry is illustrated in Fig. 4.20. We have

$$C = \epsilon_0 \frac{L_1 L_2}{g - X} \qquad (4.44)$$

from which we deduce

$$F_{gap-closing} = \frac{\epsilon_0}{2} \frac{L_1 L_2}{(g-X)^2} V^2 \qquad (4.45)$$

Once again an attractive force is generated, but it depends strongly on the position X, to to the point of tending towards infinity at the end of the stroke (X = g). This position dependence may or may not be advantageous, depending on the context.

The test electrodes of the accelerometer in Fig. 4.14 are gap-closing capacitive actuators.

Remarks:

1 – There is a trade-off between force and travel range for capacitive actuation, just like there is a trade-off between sensitivity and travel range for capacitive detection. One has indeed



Fig. 4.21 - Electrode configuration suitable for electrostatic levitation. Image from doi.org/10.1016/j.elstat.2008.03.005

2 - Although electrostatic forces are often attractive, they may also be repulsive. In particular, the geometry shown in Fig. 4.21, found in many comb structures, results in more or less pronounced electrostatic "levitation": the structure rises when a voltage is applied to the electrodes. This effect is all the less pronounced as one gets closer to the plane capacitor hypothesis.

3 - Capacitive <u>detection</u> requires a bias voltage to be applied across at least one pair of electrodes. This generates an electrostatic force and (depending on the configuration of the detection electrodes) a displacement and/or softening of the structure, regardless of whether the bias voltage is DC (case of MEMS resonators) or AC (case of accelerometers). In the case of an AC bias with pulsation  $\omega$ , the force has not only a DC component, but also a component at  $2\omega$ . In particular, it can cause an oscillation of the structure if  $\omega$  is not properly selected.

#### 4.3.3.1.2 Electrostatic softening and pull-in

As the electrostatic force is conservative and generally depends on *X*, it is customary to compare it to the restoring force of a spring by defining its "electrostatic stiffness"

$$K_{elec}(X) = -\frac{\partial F_{elec}}{\partial X} = -\frac{1}{2} \frac{\partial^2 C}{\partial X^2} V^2 \qquad (4.47)$$

In the case of gap-closing actuation, when  $X \ll g$ , the electrostatic stiffness is equal to

$$K_{elec-gap-closing}(0) = -2\epsilon_0 \frac{L_1 L_2}{g^3} V^2 \quad (4.48)$$

Since the electrostatic stiffness is negative, this phenomenon is called "electrostatic softening". The effective stiffness of a structure subject to electrostatic softening is  $K + K_{elec}$  where K is mechanical stiffness. The effective stiffness of a MEMS structure - and thus its resonant frequency - can therefore be adjusted by playing on the voltage applied to the structure. This softening phenomenon provides a powerful means for post-manufacturing adjustment of the resonance frequency of a MEMS structure.



Fig. 4.22 - Potential energy of a gap-closing structure subject to an elastic restoring force.

A phenomenon closely linked to electrostatic softening is "pull-in". Pull-in corresponds to a "collapse" of the moving electrode on the fixed electrode, above a certain value of displacement *X*. It usually results in irreparable damage to the electrodes and the structure. Fig. 4.22 provides a qualitative understanding of the phenomenon (neglecting all dynamic effects): it shows the potential energy of a structure subject to an elastic restoring force (F = -KX) and gap-closing actuation, for different values of *V*. For

$$V < V_{pi} = \sqrt{\frac{8}{27} \frac{KG^3}{\epsilon_0 S}} \qquad (4.49)$$

there are two positions of equilibrium, solutions of

$$-KX + \frac{\epsilon_0}{2} \frac{S}{(g-X)^2} V^2 = 0$$
 (4.50)

where *S* is the surface area of the capacitor. One,  $X_s$ , is stable (of positive effective stiffness), the other,  $X_i$ , is unstable (of negative effective stiffness). The amplitude of displacement of the structure is then limited by  $X_i$ , above which the structure collapses. When the voltage reaches  $V_{pi}$  we have

$$X_i = X_s = g/3$$
 (4.51)

and the effective stiffness is zero at this position. For  $V \ge V_{pi}$ , there is simply no longer a stable equilibrium position.

For a "comb-drive", according to (4.43) and (4.47), the electrostatic stiffness is zero. Therefore, there is no electrostatic softening or pull-in in the *X* direction. However, the structure necessarily has a finite stiffness in the *Y* direction (be it the suspensions or the fingers of the combs), along which there can be pull-in.

#### Remark:

MEMS micro switches (Fig. 4.23) are mechanical switches on a microscopic scale. Their interest compared to electronic switches is the outstanding quality of the insulation they provide in the open state. On the other hand, they are much slower, even if they take advantage of the pull-in phenomenon to cause an abrupt collapse of the structure. These devices have only recently become widely available on the market, although they have been under study for more than 40 years. The main difficulty is to limit the wear (mechanical, electrical, thermal...) of the structure that undergoes repeated switching, to guarantee the repeatability of the switching operation over a large number of cycles.



Fig. 4.23 - Analog Devices MEMS micro-switch (image extracted from their site).

#### 4.3.3.1.3 Voltage linearization

There are several ways to linearize the relationship between the electrostatic force and the dynamic control to be applied to the structure. These are very dependent on the type of structure (resonant or not), on the control one wishes to apply (broadband or not, to control the structure to a certain position or to generate an oscillation) and on the amplitude of the displacements one wishes to obtain. Here we only describe a few principles, in the case where electrostatic softening is negligible.

The most immediate solution, at the electronic level, is to work around a voltage operating point, i.e. to choose  $V = V_{bias} + V_{ctrl}$  where  $V_{ctrl}$  is proportional to the control and  $V_{bias} \gg V_{ctrl}$  is a bias voltage, usually DC. We then have  $F_{elec} \propto V_{bias}^2 + 2V_{bias}V_{ctrl}$ . The electrostatic force is therefore the sum of a constant term and a term proportional to the control. If the static displacement due to the constant term is not inconvenient, then the desired linearity is achieved.



Fig. 4.24 - Differential capacitive actuation.

For the static component of the force to be zero, it is necessary to act at the geometry level. Thus, differential actuation can be achieved by using electrodes located on either side of the structure, as shown in Fig. 4.24. Each electrode can then "pull" the structure in its own direction. For example, by applying a voltage  $V_+ = V_{bias} + V_{ctrl}$  on one side and  $V_- = V_{bias} - V_{ctrl}$  on the other, we obtain  $F_{elec} \propto 4V_{bias}V_{ctrl}$ .

Finally, using a control on two "logical" levels makes it possible to overcome the  $V^2$  nonlinearity. It is in fact possible to translate an "analog" command into a "bitstream" by means of a  $\Sigma\Delta$  modulator (Fig. 4.25) or PWM or to use a "bang-bang" control scheme. Similarly, the use of a voltage square wave or pulse train (Fig. 4.25) may be just as effective as a harmonic excitation to bring a structure to resonance.



Fig. 4.25 – On the left,  $\Sigma\Delta$  modulation of an analog signal (sine wave) into a binary bitstream. Locally, the average values of the two signals are equal. On the right, the amplitudes of the first harmonics of the three signals are the same.





Fig. 4.26 - Before and after fabrication (from Pierre Prache's thesis). On the left, layout of the mask of a cantilever with capacitive gap-closing actuation and detection. The cantilever and its electrodes (green with brown border) are defined in the same metal layer of a standard CMOS process. A "window" (gray) is defined in the uppermost layer of the process. An acid etch (HF) is used to etch away the oxide (black) located under the window and to simultaneously release the cantilever and the electrodes. On the right, obtained result. Bottom, equivalent electrical diagram.



Fig. 4.27 - 1-port configuration for simultaneous capacitive actuation and detection.

#### 4.3.3.1.4 Simultaneous actuation and detection

When it is necessary to implement capacitive actuation and detection simultaneously, the simplest solution is to use different sets of electrodes. This is the case for the cantilever in Fig. 4.26 or for the bridge in Fig. 4.14: the beam carries a DC voltage  $V_{bias}$ , the left electrode drives the structure at resonance with an AC voltage  $V_{ctrl} \ll V_{bias}$ , and the right electrode is held at virtual ground, being connected to a transimpedance amplifier. Neglecting the direct capacitive coupling ("feedthrough") between the left and the right electrode, we have

$$i_{meas} = \frac{d}{dt}(C(X) \times V_{bias}) = \frac{dC}{dt}V_{bias} = \frac{\partial C}{\partial X}V_{bias} \times \dot{X} = i_{mot} \quad (4.52)$$

The measured current is thus equal to the motional current, and proportional to the velocity of the structure.

It is also possible (and sometimes desirable for the compactness of the design) to use the same electrodes for detection and actuation. In this case, one difficulty is to separate, in the measured signals, the motional component from the parasitic component resulting from feedthrough. In the so-called "1-port" configuration, shown in Fig. 4.27, for example, when the voltage  $V_{bias}$  is DC, the measured current is written as follows

$$i_{meas} = \frac{d}{dt} (C(V_{bias} + V_{ctrl})) \approx \frac{dC}{dt} V_{bias} + C \frac{dV_{ctrl}}{dt} = i_{mot} + i_{feedthrough}$$
(4.53)

When possible, increasing  $V_{bias}$  is the simplest way to obtain a good ratio between the useful signal and the parasitic signal. Another solution is also to separate in time the actuation phases from the detection phases.

#### Remark:

Only the exact calculation of the capacitances between the different conductors constituting the MEMS makes it possible to quantify precisely these phenomena. This is an essential step in the validation of a new design.

#### 4.3.3.1.5 Illustration

A 3-axis gyroscope (capable of detecting an arbitrarily oriented rotation) from ST Microelectronics is shown in Fig. 4.28. By studying the geometry of the suspensions, and armed with your knowledge of capacitive actuation and detection, try to interpret the operation of this structure from the elements shown.



Fig. 4.28 - View of a 3-axis gyroscope from ST Microelectronics at the scanning microscope (top, taken from the Twitter Nanographs account), at the optical microscope (bottom, taken from the MEMSJournal site).

### 4.3.3.2 Other actuation principles

Piezoelectricity is a reversible phenomenon that can be used both to detect the motion of a MEMS structure and to control it. Piezoelectric actuation has essentially the same advantages and limitations as piezoelectric sensing described above.

Fig. 4.29 shows some principles used in electro-thermomechanical actuators, which take advantage of thermal expansion to induce the deformation of a structure. Heating is obtained by the Joule effect, when a current flows through the structure. The advantage of this actuation technique is that it can generate very large displacements or forces, but it does not necessarily require a particular manufacturing step. Its main defects are linked to its "resistive" nature (consumption, noise...).



Fig. 4.29 - Two geometries of electro-thermomechanical actuators (images from doi:10.3390/act8040069) that can generate in-plane motion. On the left, so-called "Guckel" geometry: the thin arm heats up more than the wide arm, due to the difference in resistance, resulting in a bending movement of the structure, in the direction of the cold arm. On the right, so-called "chevron" geometry: the structure can expand neither to the right and nor to the left and thus deforms upwards.

# 4.4 Dissipation4.4.1 General overview

The *predictive* quantitative determination of dissipation is probably the most difficult part of MEMS device design. On the one hand, because of the very large number of physical phenomena responsible for this dissipation: fluidic, electrical, thermal, acoustic, etc. On the other hand, because the analytical models of these phenomena are limited to simplistic device geometries or to hypotheses that are not always verified in practice. Therefore, it is advisable to be particularly cautious with regard to the estimates provided by the few models presented in this section: more complete references are given at the end of the chapter, to which it will be useful to refer if necessary.

The primary cause of dissipation on a microscopic scale is the viscosity of the fluid surrounding the device. The effect of viscosity is accentuated in thin films of fluid, such as those that are formed by the electrode armatures used for capacitive actuation and detection. For a given geometry, fluid damping becomes progressively less marked as the pressure of the surrounding fluid is reduced, until other dissipative phenomena become predominant: electrical losses, thermoelastic losses, anchor losses, etc. Ultimately, dissipation is limited by losses intrinsic to the material being used.

Within certain limits, one can model the influence of dissipation on the dynamics of a device by a viscous frictional force.

$$F = -B\dot{X} \qquad (4.54)$$

where the coefficient B results from the superposition of the different dissipative phenomena, according to

$$B = B_{fluid} + B_{electric} + B_{thermoelastic} + \dots + B_{material}$$
(4.55)

or

$$\frac{1}{Q} = \frac{1}{Q_{fluid}} + \frac{1}{Q_{electric}} + \frac{1}{Q_{thermoelastic}} + \dots + \frac{1}{Q_{material}}$$
(4.56)

Different design objectives can be pursued, depending on the type of transducer being considered. In the case of resonant transducers, such as gyroscopes, a large quality factor Q i.e. low dissipation B is desirable. In the case of non-resonant transducers such as capacitive accelerometers, a moderate quality factor is preferable (Q = 1/2 being optimal from the point of view of response time). This requires to focus essentially on fluidic phenomena in the non-resonant case, whereas all phenomena are important in the resonant case. From a practical point of view, this also means that there are important differences in the packaging - and therefore in the cost - of these two types of transducers: sealed for non-resonant transducers, sealed *and* vacuum-packaged for resonant transducers.

Finally, one should keep in mind that dissipation also results in a random force  $F_n$  that can be treated as a white noise with density

$$S_{F_n} = 4k_B T \times B \left[ N^2 / \text{Hz} \right]$$
 (4.57)

which fundamentally limits the precision of MEMS transducers.

### 4.4.2 Fluidic losses

### 4.4.2.1 Navier-Stokes equations

The constitutive relationship describing the behavior of a viscous fluid is known as the Stokes relationship. It relates the value of the stress field  $\sigma$  (internal forces) in the fluid to its deformation rate e (symmetrical part of the velocity gradient v) and pressure P

$$\boldsymbol{\sigma}_{visc} = 2\mu \boldsymbol{e} + (\lambda T r(\boldsymbol{e}) - P) \boldsymbol{I} \quad (4.58)$$

via viscosity coefficients  $\mu$  and  $\lambda$ . It should be noted that at the microscopic scale, or at low pressure, the description of air as a continuous medium, and therefore the notion of viscosity, may be inappropriate. This is discussed in the next section.

In a fluid medium, the fundamental law of dynamics is written in a slightly different form from (4.4), to take into account the effects of convection

$$div(\boldsymbol{\sigma}) + \boldsymbol{f} = \frac{D(\rho \boldsymbol{v})}{Dt} = \rho \frac{\partial \boldsymbol{v}}{\partial t} + \boldsymbol{v}. \boldsymbol{grad}(\boldsymbol{v}) \qquad (4.59)$$

where  $\rho$  is the density of the fluid and f a volume force field. Combining (4.58) and (4.59), we the Navier-Stokes equations are obtained

$$(\lambda + \mu)grad(div(v)) + \mu\Delta v + f - grad(P) = \rho \frac{\partial v}{\partial t} + v.grad(v) \quad (4.60)$$

which, coupled with the conservation of the mass

$$div(\rho \boldsymbol{v}) + \frac{\partial \rho}{\partial t} = 0 \qquad (4.61)$$

and the law of perfect gases

$$\frac{P}{\rho} = C^{te} \tag{4.62}$$

are necessary to describe an (isothermal) flow of air in an arbitrary geometry. The peculiar dimensions and geometries of MEMS devices generally allow to make simplifying assumptions to make these equations more meaningful, as explained further in the case of air films. The validity of these assumptions will obviously have to be verified, or even tested by numerical simulation.

### 4.4.2.2 From the continuous regime to the free molecular regime

The assumption that air is a continuous medium is challenged at a small scale and/or at low pressure. This can be accounted for through the Knudsen number

$$Kn = \frac{\Lambda}{D} \tag{4.63}$$

where  $\Lambda$  is the mean free path of air molecules, i.e. the average distance they travel between two interactions, and *D* is a characteristic dimension of the flow (typically the inter-electrode distance). If *D* is large compared to  $\Lambda$  (*Kn*  $\ll$  1), we can consider air as a continuous medium. On the contrary, when *Kn*  $\gg$  1, the air molecules no longer interact with each other at the scale of the flow ("free molecular" regime) and the effective viscosity of the fluid is zero.

The mean free path  $\Lambda$  is equal to  $\Lambda_0 = 64$  nm at atmospheric pressure ( $P_0 = 1$  bar) and is inversely proportional to pressure

$$\Lambda = \Lambda_0 \times \frac{P_0}{P} \qquad (4.64)$$

Provided the Knudsen number is not too large, continuum models can be used in which viscosity  $\mu$  is replaced by an effective viscosity  $\mu_{eff}$ , which depends on *Kn* and flow geometry. For a squeezed film of air ("squeeze-film damping"), we have

$$\mu_{eff}(Kn) \approx \frac{\mu}{1+6Kn} \qquad (4.65)$$

whereas for a sheared film of air ("slide-film damping") we have

$$\mu_{eff}(Kn) \approx \frac{\mu}{1+2Kn} \qquad (4.66)$$

These expressions are valid for  $Kn \le 1$ . We can refer to the articles by Bao and Frangi for cases where Kn > 1.

### 4.4.2.3 "Slide-film damping"

Slide-film damping occurs when a film of air is sheared between two surfaces moving parallel to each other (as in a "comb-drive"). When the lateral dimensions of the surfaces are large in relation to the gap separating them (Fig. 4.30), and provided that the flow can be considered extremely laminar (Couette flow), i.e.  $Re \ll 1$  (where

$$Re = \frac{\rho VD}{\mu} \qquad (4.67)$$

is the Reynolds number, and velocity *V* and distance *D* are characteristic of the flow), the Navier-Stokes equations can be summarized as follows

$$\frac{\partial^2 v_X}{\partial z^2} = 0 \tag{4.68}$$

The velocity profile in the film is therefore a linear function of height z and, according to (4.58), a purely viscous force acts on the moving surface.

$$F_{slide} = -\mu_{eff} \frac{s}{g} \dot{X} \qquad (4.69)$$

where *S* is the surface area and *g* is the gap.



Fig. 4.30 - "Slide-fim damping".

By relaxing the assumption that  $Re \ll 1$  and by taking into account inertial effects (Stokes flow), we end up with a more complex behavior

$$\mu_{eff} \frac{\partial^2 v_X}{\partial z^2} = \rho \frac{\partial v_X}{\partial t} \qquad (4.70)$$

The velocity profile in the film is no longer linear and the force acting on the moving surface verifies

$$F_{slide} + \tau_F \dot{F}_{slide} = -\mu_{eff} \frac{s}{g} (\dot{X} + \tau_{\dot{X}} \ddot{X}) (4.71)$$

where  $\tau_F < \tau_{\dot{X}}$  these two quantities being proportional to  $\rho g^2 / \mu_{eff}$ . At intermediate frequencies, a "dynamic spring" behavior ( $\dot{F}_{slide} \propto \dot{X}$ ) prevails. At high frequencies, we find a purely viscous behavior, as at low frequencies.

### 4.4.2.4 "Squeeze-film damping"

Squeeze-film damping occurs when a thin film of air is compressed between two surfaces moving perpendicular to each other (Fig. 4.31). It is widely associated with gap-closing actuation and detection. Fluidic phenomena, like electrostatic phenomena, are more complex in this configuration.



Fig. 4.31 - "Squeeze-film damping".

Under the same assumptions as above ( $Re \ll 1$ , large lateral dimensions compared to g), one may show that the velocity profile in the air film is parabolic (Poiseuille flow). Taking advantage of (4.61) and (4.62), assuming that the surfaces facing each other remain parallel, that the displacements X are small compared to g, that pressure variations p are small with respect to the ambient pressure  $P_{amb}$  and that p is independent of height in the air film, we end up with the linearized Reynolds equation

$$\Delta p - \frac{12\mu_{eff}}{g^2 P_{amb}} \frac{\partial p}{\partial t} = -\frac{12\mu_{eff}}{g^3} \dot{X} \quad (4.72)$$

whose exact solution depends on the geometry of the surfaces and their boundary conditions (p = 0 on free edges,  $\partial p / \partial \mathbf{n} = 0$  if the air cannot escape). Regardless of the exact geometry, when the squeeze number  $\sigma$  is small,

$$\sigma = 12 \frac{\mu_{eff} \omega D^2}{P_{amb} g^2} \tag{4.73}$$

where *D* is a characteristic dimension of the lateral extension of the surfaces, the behavior is purely viscous. When  $\sigma \gg 1$ , the term in  $\Delta p$  can be neglected in (4.72), and a dynamic spring behavior can be highlighted once more. One can write, when  $\sigma \ll 1$ 

$$F_{squeeze} = -\mu_{eff} \frac{s}{g} \dot{X} \times \frac{D^2}{g^2} \qquad (4.74)$$

and, more generally, for higher values of  $\sigma$ 

$$F_{squeeze} + \tau_F \dot{F}_{squeeze} = -\mu_{eff} \frac{s}{g} \dot{X} \times \frac{D^2}{g^2}$$
(4.75)

where  $\tau_F \propto \mu_{eff} D^2 / P_{amb} g^2$ .

For example, in the case of a flat surface  $S = b \times L$  with  $b \ll L$  (finger of a comb), one can choose D = b in the above expressions.



Fig. 2 Schematic diagram of perforated proof mass  $\mathbf{a}$  top view  $\mathbf{b}$  cross-sectional view



Fig. 4 Pressure distribution in the air gap under the proof mass for different structures

*Fig.* 4.32 - Optimization of the perforation size for a device that has to move out of plane (images taken from https://doi.org/10.1007/s10470-019-01560-5).

#### Remarks

1 - According to (4.69) and (4.74), as  $D \gg g$ , it is clear that the squeeze-film forces are much larger than the slide-film forces when the films have similar geometries.

2 - The fact that

$$\frac{F_{squeeze}}{F_{slide}} = \frac{D^2}{g^2} \gg 1$$
 (4.76)

is to be compared with (4.46). The price to pay for generating large electrostatic forces is therefore a high damping - unless, of course, a vacuum is created around the structure. In the case of squeeze-film, design trade-offs between high capacitance and high dissipation can also be found by playing on the characteristic distance D. This can be done, for example, by using perforations for out-of-plane movements (Fig. 4.32), or nonplanar electrodes for in-plane movements (Fig. 4.33).



Fig. 4.33 - Optimization of the finger geomety of a kinetic energy harvester (device converting ambient vibrations into electrical energy - dissipation must be minimized, but not at the expense of capacitance). Image from DOI 10.1038/s41378-018-0025-2.

3 - It should be noted that squeeze-film damping is a highly non-linear phenomenon (in the same way as gap-closing actuation), the damping coefficient having in fact a non-trivial dependence on X. This is masked in the above expressions by the assumption that  $X \ll g$ .

4 - The expressions (4.69) and (4.74) indicate a dependence as  $1/g^3$  for the squeeze-film coefficient, as 1/g for that of "slide-film". While this is true at low Knudsen numbers, one must keep in mind that  $\mu_{eff}$  also depends on the gap (4.65) (4.66). Thus, at moderate or large Knudsen numbers, the slide-film coefficient no longer depends on g. The squeeze-film coefficient varies as  $1/g^2$  for moderate Knudsen number, as 1/g for large Knudsen number.

5 - The relationship between the quality factor of a device and Knudsen number is represented classically, in an idealized way, as in Fig. 4.34. For a sufficiently large Knudsen number, fluid damping is no longer the dominant source of dissipation and many other phenomena, described below, must be taken into account. The paper by Mohanty and Imboden, especially dedicated to the case of NEMS resonators, is a good starting point on the subject.



Fig. 4.34 - Qualitative behavior of the quality factor of a MEMS structure as a function of the Knudsen number. For a sufficiently high Knudsen, Q is no longer set by viscous dissipation.

### 4.4.3 Thermoelastic losses

When a mechanical resonator is deformed, some of its parts expand, while others contract (see for example Fig. 4.3). These local deformations are sources of heat

$$q(x, y, z, t) \propto \frac{\partial}{\partial t} (Tr(\boldsymbol{\epsilon}))$$
 (4.77)

which diffuse within the resonator, according to the heat equation

$$\frac{\partial T}{\partial t} = \chi \Delta T + q \quad (4.78)$$

The local deformations therefore result in a non-uniform temperature field, in phase or in quadrature with the deformations, depending on whether or not the latter are fast with respect to the characteristic time constant of the heat exchanges (proportional to  $D^2/\chi$  where *D* is a characteristic dimension of the resonator, typically its vibrating height *h* in the case of a flexural resonator).

According to (4.6), these temperature variations give rise to a field of thermal stresses opposing the movement of the resonator, in phase or in quadrature with it. We find qualitatively the same behavior as in the previous examples: at low frequency, a "viscous damping" type behavior, at high frequency, a "dynamic spring" behavior. This last type of behavior is obviously more desirable when an important quality factor is targeted.

An ingenious approach to limit thermoelastic dissipation consists in reducing the effective thermal diffusivity of the material by defining slits in the resonator that slow down the heat exchanges between its hot and cold parts (Fig. 4.35).



Fig. 2. Computer-aided-design view of the designed resonator. (a) Modal shape-function of the flexural mode of the resonator under study. (b) Drive and sense ports have been hidden for the sake of clarity. The contour of the displacement field is shown in colour.



Fig. 3. SEM image of the designed resonator. The actuated slotted beams are 351  $\mu m$  long and 15  $\mu m$  wide.

Fig. 4.35 - MEMS time reference, work by PoliMi and ST Microelectronics. Operating principle (a), vibration mode used (b) and SEM view (c). The resonator is split along its entire length to limit thermoelastic losses (images from DOI 10.1109/TIE.2019.2938465). The same type of pattern can be seen on the SiTime resonator in Fig. 4.14.

# 4.4.4 Anchor losses

Where fluid losses can be ignored, acoustic radiation at the resonator anchors is usually a major cause of dissipation. These anchor losses depend on the acoustic impedance matching between the oscillation mode of the resonator and what is on the "other side" of the anchor(s), i.e. the resonator support. The energy leaving the resonator can be radiated and lost in the support, to the detriment of the resonator quality factor, or reflected and confined in the resonator.

An analytical expression of the dissipation coefficient can be obtained by assuming a simplistic geometry for the support (infinite half-space, etc., cf. Hao's article), to which the resonator is assumed to be weakly coupled. This type of model leads to the general conclusion that in order to limit dissipation in the anchor, the resonator must have a large aspect ratio (ratio of length to vibrating height). In other words, the greater the oscillation wavelength relative to the

characteristic dimension of the anchor point, the less energy the resonator dissipates through the anchor.

Several approaches exist to limit this phenomenon, such as the use of acoustic reflectors at the anchor (Fig. 4.36), that of double resonators oscillating in phase opposition so that the total force seen by the support is zero (tuning fork principle, to which the resonators in Figs. 4.13 and 4.35 are related), or more generally that of resonators anchored at their nodal points.



Figure 1. 3D model of a thin-film piezoelectric-on-silicon (TPoS) resonator with acoustic reflectors.



Figure 2. First (*a*) and third harmonic (*b*) mode shapes for a lateral-mode resonator. The color in the pictures denotes changes in strain along the axis of movement (width). Blue signifies compressive strain while red denotes tensile strain.





Fig. 4.37 - Modeling of the finite conductivity of a resonator by a parasitic resistor r.

### 4.4.5 Electrical losses

Ohmic losses may also be significant contributors to dissipation in MEMS resonators. These can result not only from the finite conductivity of the MEMS structure, but also from imperfections in the electronics associated with the resonator, such as an op-amp finite gain.

For example, in Fig. 4.37, the diagram in Fig. 4.26 is reproduced, with the addition of a small resistor r between the resonator and the input of the circuit, representing the finite

conductivity of the resonator and the interconnections with the electronics. As a result, the voltage across the variable capacitor *C* is no longer equal to  $V_{bias}$  but to  $V_{bias} - v$  where  $v \ll V_{bias}$ . Neglecting electrostatic softening, the force on the resonator verifies

$$F_{elec} \propto (V_{bias} - v)^2 - (V_{bias} - V_{ctrl})^2 \approx 2V_{bias}V_{ctrl} - 2V_{bias}v$$
(4.79)

The total force is therefore the sum of a "drive" term and an ohmic term due to resistor *r*. It is easy to show that

$$v \approx r \times \frac{\partial C}{\partial X} V_{bias} \times \dot{X}$$
 (4.80)

so that the ohmic term of the electrostatic force behaves as a viscous force, whose dissipation coefficient is proportional to  $V_{bias}^2$ .

The same type of behavior can be attributed to the imperfections of the AFE. For example, if we take into account the finite gain *A* of the amplifier of the transimpedance circuit in Fig. 4.38, one may show that

$$v = V_{-} = -\frac{V_{out}}{A} \approx \frac{Z}{A} \times \frac{\partial C}{\partial X} V_{bias} \times \dot{X}$$
 (4.81)

or

$$v + R_f C_f \dot{v} \approx \frac{R_f}{A} \times \frac{\partial C}{\partial X} V_{bias} \times \dot{X} (4.82)$$

Thus, a viscous behavior is observed below the cut-off frequency of the circuit, a dynamic spring behavior above it.



Fig. 4.38 - Transimpedance amplifier. Ohmic losses in  $R_f$  affect the quality factor of the resonator due to the finite gain of the op amp.

### 4.4.6 Other losses

As one dissipation phenomenon is minimized, another one takes over and becomes preponderant. Thus, energy can be dissipated because of the roughness of the resonator surface ("surface" losses, all the more marked as the characteristic dimensions decrease and the surface / volume ratio increases), because of irregularities in the crystal structure ("volume" losses, all the more marked as the crystal lattice has defects), because of quantum phenomena (phononphononon interactions or "Akhiezer effect").

#### Remarks

1 - A figure-of-merit commonly used to compare different resonators is the  $Q \times f$  product, i.e. the product between the quality factor of the resonator and its resonance frequency. This  $Q \times f$  product is naturally bounded by the ratio between the intrinsic stiffness of the material and its intrinsic dissipation, i.e.

$$Q \times f \le \frac{1}{2\pi} \times \left(\frac{K}{B}\right)_{max}$$
 (4.83)

For (monocrystalline) silicon, the maximal  $Q \times f$  product is of the order of  $3 \times 10^{13}$  Hz, as limited by the Akhiezer effect. Typical MEMS flexural resonators have  $Q \times f$  values well below this upper limit. This is explained by the preponderance of the loss phenomena described above.

2 - As previously pointed out, except in the case of simple resonator geometries, it is difficult to predict the dissipation coefficient of a MEMS resonator accurately using analytical approaches alone. The use of numerical models of dissipation phenomena can obviously be useful to adapt the parameters of analytical models to more complex geometries.

3 - A semi-empirical approach to damping phenomena consists in using the Rayleigh model (or the Caughey model, which is a generalization of it). This approach consists in postulating that dissipation has structural/intrinsic causes, on the one hand, and inertial/extrinsic causes, on the other hand, which is translated by

$$B_i = \alpha K_i + \beta M_i \tag{4.84}$$

where the  $K_i$ ,  $B_i$  and  $M_i$  are respectively the stiffness, dissipation and mass coefficients of the *i*<sup>th</sup> mode of the structure (see appendix D), and  $\alpha$  and  $\beta$  are the Rayleigh coefficients (common to all modes). These coefficients are determined empirically by measuring the resonance frequencies and quality factors of the different modes and by "fitting" the coefficients  $\alpha$  and  $\beta$  as well as possible to these measurements. This only makes sense if one suspects that the same dissipation phenomena are at work for the modes under consideration. One may assume that these coefficients vary little if the geometry under study is close to the geometry on which the measurements were made, and thus reuse the empirically-estimated coefficients in the models used for the design. This approach is a leap of faith if it is not supported by a good knowledge of the physics of the studied devices, but it can be very useful.

# 4.5 References

The calculation of classical MEMS suspensions is discussed in detail in Gary Fedder's thesis, which is a reference work on the subject. It is available here:

[1]G. Fedder, "Simulation of microelectromechanical systems", ECE Berkeley, 1994 https://users.ece.cmu.edu/~fedder/phd\_thesis/

These structures are assessed with respect to gyroscope applications in Andrei Shkel's book, cited in the previous chapter. A calculation form for "complex" beams is also available in appendix C of this handout.

Concerning micro-scale transduction methods, some rather old, but useful, books are

[2] G. Kovacs, "Micromachined transducers sourcebook", McGraw Hill, 1998.

[3] M.H. Bao, " Micromechanical transducers: pressure sensors, accelerometers, and gyroscopes ", Elsevier, 2000

Good starting points for understanding dissipation phenomena affecting MEMS and NEMS are

[4] M. Imboden, P. Mohanty, " Dissipation in nanoelectromechanical systems ", Physics Reports, vol. 534, 2014

5] M.H. Bao, H. Yang, "Squeeze film air damping in MEMS", Sensors and Actuators A, vol. 136, 2007

[6] A. Frangi, et al, "Near vacuum gas damping in MEMS: simplified modeling", vol. 26, 2017

[7] R. Lifschitz, M. L. Roukes, "Thermoelastic damping in micro- and nanomechanical systems ", Physical Review B, vol. 61, 2000

[8] Z. Hao, A. Erbil, and F. Ayazi, " An analytical model for support loss in micromachined beam resonators with in-plane flexural vibrations ", Sensors and Actuators A, vol. 109, 2003

Just open a recent issue of the Journal of Microelectromechanical Systems (IEEE/ASME), Sensors Journal (IEEE), Sensors and Actuators (Elsevier) or the Journal of Micromechanics and Microengineering (IOP) to discover the "trends" and the state of the art in the field, which are evolving in line with advances in microfabrication. At present, the teams of T. Kenny (Stanford U.) and G. Langfelder (PoliMi) are producing numerous, often original and interesting works on MEMS, especially on resonators.

# 5 MEMS design and modeling

# 5.1 Introduction

In the industry, the design of an integrated electronic system generally follows a top-down approach, from the highest level, the system specifications, to the lowest level: mask design, manufacturing, assembly and packaging (Fig. 5.1). In the case of a sensor, a major difficulty lies in the need for a dialogue between electronic design and (mechanical/multiphysical) transducer design, which require different tools, fields of knowledge and technological processes. Due to the dichotomy between electronics and transduction, the design flow is usually split in two, with co-simulation of the two parts being possible only at a high level.



Fig. 5.1 - "Structured" design of an integrated MEMS sensor. Image from the works of Gary Fedder (10.1109/MEMSYS.1999.746742).

In the *idealized* design flow shown in Fig. 5.1, the sensor specifications appear at the highest level. From these specifications, the system *architecture* is described, in the form of subblocks each performing a *function* (transduction, amplification, filtering, modulation, A/D conversion, etc.). The inputs and outputs of each of the blocks being defined and their performances specified, the blocks can be designed at *circuit* level, where choices of structures are made, and at the level of the *components of* these structures. Once all the active (transistors) or passive (resistors, capacitors, masses, suspensions) elements have been dimensioned, the *masks* can be laid out before being sent to manufacturing. As one goes down the design flow, the description of the system is more and more precise, and the tools used, in particular the models, are more and more precise and complex. Although globally top-down, the design flow also includes locally "bottom-up" phases, making it possible to adjust the tools and models used at higher levels, to re-evaluate design choices, and even specifications.

As far as the electronic part is concerned, one can generally conform to this very hierarchical approach to design, as the effects of the relative arrangement of the circuit components and their geometrical design (the mask layout) only marginally challenge the choices made at higher levels. On the other hand, for the transduction part, the distinction between the

component level and the mask level is more blurred, as the properties of MEMS components are so dependent on their actual geometry.

# 5.2 MEMS design flow

In a more concrete way, we may formalize a top-down design flow for the transducer part as follows:

- (1) Different transducer geometries are selected to implement the function to be performed.
- (2) A high-level model of "reasonable" order is assembled for each geometry, based on the description of the geometry.
- (3) A constrained optimization procedure of the model parameters is launched.
- (4) Low-level simulations are performed to validate the results and the underlying assumptions of the optimized high-level model.

# 5.2.1 Geometry selection

Step (1), geometry selection, consists not only in choosing a geometry, but also the alternative ways of detecting the motion of the structure (capacitive, resistive, resonant transduction, etc.) or actuating it if necessary. These are absolutely crucial choices, since they determine the space in which the design/optimization of the transducer will take place. However, it is difficult to talk about them except in a very general way.

# 5.2.2 High-level modeling

Step (2), high level modeling, can be approached with different degrees of finesse. An analytical approach is generally only possible by making strong kinematic assumptions about the displacement of the structure (e.g. "the structure only deforms according to dimension X", or "only the first vibration mode of the structure is excited"). It also requires simplifying the physics of the transducer in all domains ("perfectly rigid masses", Euler-Bernoulli hypotheses, plane capacitors, etc.). However, it makes it possible to arrive at models of a completely "reasonable" complexity (1 or 2 degrees of freedom) that are easy to simulate and optimize, and sometimes offer a satisfactory degree of fidelity while remaining "intelligible". This approach also has the advantage that the models are fairly simple to "adjust" to the results of low-level simulations, or even characterizations of real devices. The alternative to this approach is to use a software such as MEMS+: it makes it possible to assemble a high-level model from pre-defined "primitives" (straight beam, curved beam, serpentine, rigid mass, etc.), each of which is described without any particular kinematic hypothesis - but with simplified physics. The resulting models typically have from 10 to 1000 degrees of freedom, depending on the complexity of the designed structure, with the lower limit being "reasonable" and the upper limit not, at least in terms of co-simulation with the electronic part. These models are obviously of better fidelity than those obtained "by hand", with an analytical approach, at the price of increased complexity. Their co-simulation with the electronics typically requires an "order reduction" step.

Additional elements for obtaining high-level analytical models of suspended structures are given in Appendix C. The analytical approach to modal analysis is described in Appendix D.

## 5.2.3 Optimization

Phase (3), optimization, consists in determining the best parameters of the high-level model with respect to the specifications (in terms of sensitivity, noise, systematic error, etc.). Some parameters can be continuous (length and width of a beam, for example, diameter of perforations), while others can take discrete values (number of fingers of a comb). The optimization must also take design constraints into account. These are related in particular to the manufacturing process (typically minimum gap size, minimum width and maximum length of the beams, see appendix H), to particular specifications (maximum size of the device, operating temperature range, measurement range), to physical assumptions underlying the high-level model (rigid fingers and masses, plane capacitors, beam, etc.). Except for some very simple geometries, there is no analytical solution to this type of problem and optimization must be done numerically (when it is not done heuristically or empirically). The optimization then repeatedly calls upon the simulation of the high-level model designed in step (2). Consequently, the lower the level of the model, the less complex the optimization is from a computational point of view (but the results of the latter may be less relevant than with a model of increased complexity). A "multigrid" approach, in which one first optimizes a high level analytical model, which is used as a starting point for the optimization of a finer model, makes it possible to obtain a good compromise between complexity and fidelity.

If the optimization is successful (i.e. if there is a set of optimal parameters for which the design objectives are met), a low-level model of the geometry is assembled: step (4) is then carried out. If the optimization does not succeed for any of the geometries tested, step (3) should be restarted with a relaxed set of specifications, even if it means tightening those on the electronic part, or new geometries should be selected.

# 5.2.4 Low-level modeling

Phase (4), low-level modeling, verifies the relevance of the optimization results. Phenomena occurring inside the structure ("intrinsic" to the structure) are modeled using the finite element method (FEM) applied to their fundamental governing equations: this is notably the case of elasticity and thermoelasticity, governed by the Navier's and Fourier's equations. The ("extrinsic") phenomena of interaction between the structure and its environment are typically modeled using the Boundary Element Method (BEM), applied to Poisson's equation for the calculation of electrostatic forces (see Appendix E) or to the Navier-Stokes or Boltzmann's equations for the calculation of fluid friction (at low or high Knudsen). This type of "high-fidelity" model can be composed of  $10^5 \ge 10^7$  degrees of freedom, coupled in a non-linear way; its simulation is therefore particularly complex and costly. This is why it is only used sparingly, and the optimization of the device, which is based on repeated simulations for different sets of parameters, is done on the basis of a high-level model.

If phase (4) makes it possible to validate the design choices made using the high-level model, a "reduced order model" of the transducer is generated, for co-simulation with the electronic part. This reduced-order model may simply consist in a "re-set" version of the high-level model of phase (2), with parameters re-adjusted to the results of the low-level simulations. Alternatively, there are many available numerical methods for reducing, in a semi-automatic way, the complexity of a large system. These methods typically consist in the projection of the low-level model on the basis of a few "suitably chosen" eigenmodes of the structure and the approximation of non-linearities on this same basis. These are described in somewhat more detail in Appendix D.

Finally, if phase (4) is not conclusive, the high-level model of phase (2) can be refined and the optimization procedure (3) restarted on this new model. Otherwise, the constraints of the optimization procedure can be re-evaluated, including those underlying the physical assumptions of the high-level model.

# 5.3 Co-simulation

Different levels of representation (figure).

# 5.3.1 Equivalent electrical models

Butterworth-Van Dyke. To be completed.

# 5.3.2 Hardware description languages

VHDL-AMS, Verilog-A. To be completed.

# 5.4 References

The books

[1] S. Senturia, "Microsystem design", Springer, 2001

[2] T. Bechtold et al, "System-level modeling of MEMS", Wiley, 2013.

offer interesting insights into the field of MEMS design, the first in an in-depth but slightly dated manner, the second in a slightly more superficial but more up to date manner.

The more or less automated generation of reasonable order models is a research area in itself. The tools presented in this handout (including the appendix) are sufficient to achieve analytical high-level models for many simple geometry MEMS structures. One may also refer to

[3] V. Zega et al., "Numerical modelling of non-linearities in MEMS resonators", Journal of Microelectromechanical Systems, vol. \*, 202\* (to be published)

for a recent take on the problem (in addition to the interesting approach, this article proposes an extensive bibliography of the subject).

Finally, one may browse with interest the technical documentation of Coventor's software, CoventorWare and MEMS+, which are industrial standards in the field.

# 6 Around sensors

# 6.1 Components

- Voltage supplies
- Voltage references
- Clocks
- Other sensors (T, multi-axis, etc.)
- ADC and DAC
- Digital

# 6.2 Processing

- Lock-in detection / switched capacitors.
- Testing, calibration and compensation
- Feedback control (at mass / accelerometer / gyro level, at readout level)
- Power management
- Data fusion

# 6.3 Environment

- Packaging

To be completed.

# 7 FM resonant sensors and energy harvesters

To be completed.

# 8 Appendices

# A - PSD, variance and Allan variance

The power spectral density (PSD) provides a frequency representation of the stability of a signal. v(t). It is equal, to the nearest multiplicative factor, to the squared modulus of the Fourier transform of v(t) and, under certain hypotheses, to the Fourier transform of its autocorrelation function. In this handout, we note it  $S_v(f)$  where f is the frequency (f > 0, it is therefore a "one-sided" PSD). It is expressed in "units of v "squared per Hertz.

When v(t) is stationary, its variance  $\sigma_v^2$  is then given by

$$\sigma_{\nu}^2 = \int_0^\infty S_{\nu}(f) df \qquad (A1)$$

In the event that v(t) is not stationary, for example if

$$S_{\nu}(f) = h_{\alpha} \times f^{\alpha} \qquad (A2)$$

with  $\alpha = -1$  (flicker noise) or  $\alpha = -2$  (Brownian noise), this integral is no longer defined. Other variance estimators can then be used to characterize temporally the instability of v(t). This is the case of the Allan variance, or two-sample variance. This quantity gives information on the difference between the averages of v(t) obtained on successive time windows of duration  $\tau$ . The definition of Allan's variance is

$$\sigma_{v}^{2}(\tau) = \frac{1}{2} \langle (\bar{v}_{k+1} - \bar{v}_{k})^{2} \rangle \qquad (A3)$$

where the operator ( ) designates an average over an infinite time and

$$\bar{v}_k = \frac{1}{\tau} \times \int_{t_k}^{t_{k+1}} v_k(t) dt \qquad (A4)$$

and  $t_k = k\tau$  (see Fig. A1).

#### Fig. A1 - Notations for Allan variance

Allan's variance provides information on the short or long term fluctuations of v(t) depending on whether  $\tau$  is small or large, respectively. If v(t) is the output of a sensor, the minimum of the Allan variance indicates the time over which to average v(t) to get the most precise measurement, i.e. whose value changes the least two successive samples. The Allan variance  $\sigma_v^2(\tau)$  is related to the spectrum  $S_v(f)$  by the formula

$$\sigma_{\nu}^{2}(\tau) = 2 \int_{0}^{\infty} \frac{\sin^{4}(\pi\tau f)}{(\pi\tau f)^{2}} S_{\nu}(f) df \qquad (A5)$$

For more details, see the websites of David Allan (http://www.allanstime.com/AllanVariance/), Enrico Rubiola (rubiola.org) and the article :

A. Makdissi, F. Vernotte, E. De Clercq, "Stability variances: a filter approach. ", ARXIV : 0904.2660, 2009

# B - From Hooke's law to Euler-Bernoulli

# B.1 Derivation of Euler-Bernoulli's equation

A "bending" deformation of a beam is a deformation where the curvature of the beam changes due to an external force. The effect of this curvature can be described as an axial strain gradient on either side of a so-called "neutral" axis, with no net elongation, as shown in Fig. B1.

#### Fig. B1 - Bending deformation and Euler-Bernoulli hypothesis

Euler-Bernoulli's hypothesis, widely verified for small deformations of thin beams, is that the straight sections of the beam remain perpendicular to the axis during deformation. Under this hypothesis, the study of the deformation of the beam is reduced to the determination of the shape w(x, t) taken by the axis, from which the stress and strain fields in the whole beam can be deduced.



Fig. B2 - Displacement field in the beam

Euler-Bernoulli's hypothesis implies that the displacement field in direction x can be written (Fig. B2)

$$u_x \approx -z \frac{\partial w}{\partial x}$$
 (B1)

from which we deduce the longitudinal strain and stress along x

$$\epsilon_{xx} = \frac{\partial u_x}{\partial x} = -z \frac{\partial^2 w}{\partial x^2} \quad (B2)$$
$$\sigma_{xx} = E \epsilon_{xx} = -E z \frac{\partial^2 w}{\partial x^2} \quad (B3)$$

using Hooke's law and assuming that  $\sigma_{xx}$  is very large compared to other stress components.

#### Fig. B3 - Notations for internal forces and moments.

We are interested in the forces and moments resulting from the stresses acting on each straight section of the beam: normal force N, shear force V and bending moment M (Fig. B3). The latter is written

$$M = \iint z\sigma_{xx}dydz = -EI\frac{\partial^2 w}{\partial x^2}$$
(B4)

where  $I = \iint z^2 dy dz$  is the moment of inertia of the beam. We can now write the balance of the moments according to y, the balance of forces according to x and the fundamental law of dynamics according to z (i.e. we ignore inertia according to x and rotational inertia according to y). According to Fig. B3, we have

$$\frac{\partial M}{\partial x} - V + N \frac{\partial w}{\partial x} = 0 \quad (B5)$$
$$\frac{\partial N}{\partial x} = 0 \quad (B6)$$
$$f + \frac{\partial V}{\partial x} = \rho S \frac{\partial^2 w}{\partial t^2} \quad (B7)$$

where *S* is the cross-section of the beam. The internal forces and moments of (B4-B7) are easily eliminated to arrive at the Euler-Bernoulli equation

$$EI\frac{\partial^4 w}{\partial x^4} - N\frac{\partial^2 w}{\partial x^2} + \rho S\frac{\partial^2 w}{\partial t^2} = f(x, t)$$
(B8)

### **B.2** Boundary conditions

We can see that the absence of a couple at one end of the beam is well translated, through (B4), into the nullity of the second derivative of *w*. Equation (B5) implies that the condition "zero force at one end" is written as follows

$$-EI\frac{\partial^3 w}{\partial x^3} + N\frac{\partial w}{\partial x} = 0 \quad (B9)$$

which reduces to the nullity of the third derivative of *w* when the normal effort is zero or when the slope is zero at the considered end.

Two other boundary conditions are useful: the case where the end of the beam is held by (i) a spring (ii) a torsional spring. The first case corresponds to a shear force at the end that verifies

$$V = K \times w \qquad (B10)$$

Using (B5), this boundary condition is written as follows

$$-EI\frac{\partial^3 w}{\partial x^3} + N\frac{\partial w}{\partial x} = K \times w \Leftrightarrow w = -\frac{1}{K} \times \left(EI\frac{\partial^3 w}{\partial x^3} - N\frac{\partial w}{\partial x}\right)$$
(B11)

One may verify that, when the spring stiffness tends towards infinity, the boundary condition on the shear force "degenerates" into a boundary condition displacement *w*, which must then be zero.

The second case is similar, it corresponds to a couple at the end such that

$$M = -C \times \frac{\partial w}{\partial x}$$
(B12)

which, using (B4), becomes

$$\frac{\partial w}{\partial x} = \frac{EI}{C} \frac{\partial^2 w}{\partial x^2} \qquad (B13)$$

When the torsional stiffness tends towards infinity, the "degenerate" boundary condition of zero slope at one end is obtained.

A fixed end is therefore equivalent to a spring combined with a torsion spring, both of infinite stiffness.

# C - From simple beams to complex suspensions



Fig. C1 - Configurations of the bending beam seen in this section

We are interested here in the case of beams in pure bending or torsion, in the absence of normal stress, loaded at their ends with "non-trivial" boundary conditions. Many suspensions can be modeled as combinations of these different cases.

# C.1Torsion beam – Fixed at x = 0, couple at x = L

This case is modeled as a torsion spring, with torsional stiffness

$$C_{torsion} = \frac{GJ}{L}$$
 (C1)

where G = E/2(1 + v) is the torsional modulus of the material, and *J* the torsional moment of the beam around the axis *x*. For a beam with a rectangular cross-section, we have

$$J = Dd^{3} \times \left(\frac{1}{3} - 0.21 \frac{d}{D} \left(1 - \frac{1}{12} \frac{d^{4}}{D^{4}}\right)\right) \quad (C2)$$

where  $D = \max(b, h)$  is the largest of the lateral dimensions and  $d = \min(b, h)$  the smallest.

# C.2 Bending Beam – Fixed at x = 0, force at x = L

Under the effect of a force *F* the displacement at the end is

$$w(L) = \frac{L^3}{3EI} \times F \tag{C3}$$

and the slope at the end is

$$\theta(L) = \frac{dw}{dx}\Big|_{x=L} = \frac{L^2}{2EI} \times F$$
 (C4)

# C.3 Bending Beam – Fixed at x = 0, couple at x = L

Under the effect of a moment *M*, the displacement at the end is

$$w(L) = \frac{L^2}{2EI} \times M \tag{C5}$$

and the slope at the end is

$$\theta(L) = \frac{L}{EI} \times M \tag{C6}$$

# C.4 Bending beam –Fixed at x = 0, sliding + force at x = L

Under the effect of a force *F*, the displacement at the end is

$$w(L) = \frac{L^3}{12EI} \times F \tag{C7}$$

and the slope at the end is

$$\theta(L) = 0 \tag{C8}$$

This result can be derived from the two previous load cases. A fixed-sliding beam subject to a force F can effectively be seen as a cantilever subject to a force F and to a couple M = -FL/2 which guarantees (C8), by linear combination of (C4) and (C6).

# C.5 Bending beam – Spring at x = 0, sliding + force at x = L

According to (B11), the boundary condition in x = 0 is written

$$K \times w(0) - EI \frac{d^3 w}{dx^3}\Big|_{x=0} = 0 \quad (C9-a)$$
$$\frac{dw}{dx}\Big|_{x=0} = 0 \quad (C9-b)$$

Under the effect of a force *F* the displacement at the end is :

$$w(L) = \frac{L^3}{12EI} \times F \times \left(1 + \frac{12EI/L^3}{K}\right)$$
(C10)

By developing this expression (written in this form to emphasize the stiffness ratio), we highlight the fact that we can consider this case as two simple springs put in series.

C.6 Bending beam – Torsional spring at x = 0, sliding + force at x = L

According to (B12), the boundary condition in x = 0 is written

$$w(0) = 0$$
 (C11-a)  
 $C \times \frac{dw}{dx}\Big|_{x=0} - EI \frac{d^2w}{dx^2}\Big|_{x=0} = 0$  (C11-b)

Under the effect of a force *F*, the displacement at the end is

$$w(L) = \frac{L^3}{12EI} \times F \times \left(1 + 3\frac{\frac{EI/L}{C}}{1 + \frac{EI/L}{C}}\right)$$
(C12)

In this form, one recognizes in the right member the ratio of torsional stiffness C and that of a cantilever with a couple at its end (C6). One finds the fixed-sliding case (C7) in the limiting case of large C, and the cantilever case (C3) in the limiting case of small C.

### C.7 General model of a beam with end loads

The kinematic hypotheses underlying the above calculations (bending in a single direction, for example) do not make it possible to perfectly capture the behavior of suspensions with complex geometry. Armed with a little courage, it is however possible to establish a general model of a Euler-Bernoulli beam, with 6 degrees of freedom (3 translations, 3 rotations) and

subject to 6 concentrated loads (forces and couples in the 3 dimensions) at each end, in the form of

	$\left[ \frac{ES}{L} \right]$	0	0	0	0	0	$-\frac{ES}{L}$	0	0	0	0	0	
$\begin{bmatrix} F_{x}(0) \\ F_{y}(0) \\ F_{z}(0) \\ M_{x}(0) \\ M_{z}(0) \\ F_{x}(L) \\ F_{y}(L) \\ F_{z}(L) \\ M_{x}(L) \\ M_{y}(L) \\ M_{z}(L) \end{bmatrix} =$	0	$\frac{12EI_z}{L^3}$	0	0	0	$\frac{6EI_z}{L^2}$	0	$-\frac{12EI_z}{L^3}$	0	0	0	$\frac{6EI_z}{L^2}$	$\begin{bmatrix} u_x(0) \\ u_y(0) \\ u_z(0) \\ \theta_x(0) \end{bmatrix}$
	0	0	$\frac{12EI_y}{L^3}$	0	$-\frac{6EI_y}{L^2}$	0	0	0	$-\frac{12EI_y}{L^3}$	0	$-\frac{6EI_y}{L^2}$	0	
	0	0	0	$\frac{GJ_x}{L}$	0	0	0	0	0	$-\frac{GJ_x}{L}$	0	0	
	0	0	$-\frac{6EI_y}{L^2}$	0	$\frac{4EI_y}{L}$	0	0	0	$\frac{6EI_y}{L^2}$	0	$-\frac{2EI_y}{L}$	0	
	0	$\frac{6EI_z}{L^2}$	0	0	0	$\frac{4EI_z}{L}$	0	$-\frac{6EI_z}{L^2}$	0	0	0	$\frac{2EI_z}{L}$	$\theta_y(0) \\ \theta_z(0)$
	$-\frac{ES}{L}$	0	0	0	0	0	$\frac{ES}{L}$	0	0	0	0	0	$\begin{bmatrix} u_x(L) \\ u_y(L) \\ u_z(L) \\ \theta_x(L) \\ \theta_y(L) \\ \theta_z(L) \end{bmatrix}$
	0	$-\frac{12EI_z}{L^3}$	0	0	0	$-\frac{6EI_z}{L^2}$	0	$\frac{12EI_z}{L^3}$	0	0	0	$\frac{6EI_z}{L^2}$	
	0	0	$-\frac{12EI_y}{L^3}$	0	$\frac{6EI_y}{L^2}$	0	0	0	$\frac{12EI_y}{L^3}$	0	$-\frac{6EI_y}{L^2}$	0	
	0	0	0	$-\frac{GJ_x}{L}$	0	0	0	0	0	$\frac{GJ_x}{L}$	0	0	
	0	0	$-\frac{6EI_y}{L^2}$	0	$-\frac{2EI_y}{L}$	0	0	0	$-\frac{6EI_y}{L^2}$	0	$\frac{4EI_y}{L}$	0	
	0	$\frac{6EI_z}{L^2}$	0	0	0	$\frac{2EI_z}{L}$	0	$\frac{6EI_z}{L^2}$	0	0	0	$\frac{4EI_z}{L}$	

this relation being valid for a straight beam of length L, cross-section S, torsional moment  $J_x$  and moments of inertia  $I_y$  and  $I_z$  (and neglecting the effect of normal stress on flexural stiffness). One may then determine the static characteristics of a suspension made up of straight segments by applying this model to each of its parts. Thus one gains in generality what is lost in intelligibility.

This is the approach used in Coventor's MEMS+ software, which makes it possible to "build" MEMS transducers assembled like Lego, element by element (see Fig. C3 the available element library), and thus to generate small numerical models, offering an excellent compromise between computational complexity and accuracy. From this point of view, MEMS+ is a form of "missing link" between (i) simplified models with 1 or 2 degrees of freedom, essential for an intuitive approach to the dimensioning and behavior of a transducer, as well as for transducer-electronic co-simulation, and (ii) finite element models based on 3D elasticity.

An example of a structure modeled with MEMS+ is shown in Fig. C4: it is a resonator – central beam + "H" for capacitive actuation and detection – associated with electrostatically actuated microlevers – these make it possible to adjust the normal force in the resonator, and thus its resonance frequency.



Fig. C3 - Library of elements available in MEMS+. Several underlying physical models are available (Bernoulli, Timoshenko, etc.).



Fig. C4 - Structure modeled with MEMS+ and stiffness matrix associated with one of the elements (Euler-Bernoulli beam). The model includes 55 elements (only 16 beam-type deformable elements) and 239 degrees of freedom (mechanical, but also fluid and electrostatic).
# D - Eigenmodes and model order reduction D.1 Galerkin's method

We are trying to determine an approximation of the solution of a differential equation

$$\Lambda w = f + BC \quad (D1)$$

where  $\Lambda$  is a linear differential operator (spatial and temporal), f a source/force term (possibly dependent on w) and BC designates boundary conditions.

Galerkin's method - more precisely Bubnov-Galerkin's method - is one method to achieve this. It consists in

- (ii) approximating w in the form of a linear combination  $\hat{w}$  of "suitably chosen" functions forming a basis of the solution space of (D1).
- (iii) determining the coefficients of this linear combination by projection on the same basis of functions. This guarantees that the residual  $w \hat{w}$  is orthogonal to the chosen basis.

Galerkin's method, combined with the choice of "eigenmodes" as basis functions, is nothing more than an elaborate form of Fourier analysis. It leads to the "spectacular" result that the modal coefficients can be determined independently of each other when f does not depend on w in a non-linear way. It simplifies the initial problem by decomposing it into n simpler subproblems. In many case studies, n can be chosen small or even equal to 1.

For an engineer who can be contented with approximations, Bubnov-Galerkin's methods also have the appeal of providing results that are extremely robust to the exact choice of basic functions. Hereafter we present its application in the case of a bending beam, for different load cases: this study has a strong practical interest (the sensitive element of many resonant MEMS sensors is often a simple beam), while being educational, being treated in a quasi-analytical way. It may also be put into perspective with the numerical approach of model order reduction, discussed in section D.3.

## D.2 Modal analysis of a simple resonator

In this section we are interested in a beam of length *L*, width *b*, height *h*, moment of inertia  $I = bh^3/12$ , cross-section *S*, made of a material with Young's modulus *E* and density  $\rho$ . When subject to a linear density of force f(x, t), the beam takes a shape w(x, t) which is governed by

$$EI\frac{\partial^4 w}{\partial x^4} + \rho S\frac{\partial^2 w}{\partial t^2} = f(x,t) + BC$$
(D2)

where *BC* designates 4 boundary conditions that are assumed to be "trivial" (4.8), i.e. no concentrated force or couple at the beam ends.

### D. 2.1 - Eigenmodes of a pure bending beam

The eigenmodes  $w_k(x)$  of a pure bending beam can be defined as the non-trivial solutions of

$$EI\frac{d^4w_k}{dx^4} - \rho S\omega_k^2 w_k = 0 + BC \quad (D3)$$

where  $\omega_k > 0$  is the angular eigenfrequency associated with the mode. It can be shown that they form a complete base for the solutions of (D2) and that this base is orthogonal with respect to the scalar product

$$\langle u.v \rangle = \frac{1}{L} \times \int_0^L u(x) \times v(x) dx$$
 (D4)

which represents the average value of the product of u and v on [0, L]. As the  $w_k(x)$  verify the boundary conditions of (D2) and as these are trivial, it is clear that any linear combination of these modes

$$\widehat{w}(x,t) = \sum_{k=1\dots n} a_k(t) w_k(x)$$
 (D5)

also verifies the boundary conditions.

According to (D3),  $w_k(x)$  has the general form

$$w_k(x) = A_k \sin(\gamma_k \frac{x}{L}) + B_k \cos(\gamma_k \frac{x}{L}) + C_k \sinh(\gamma_k \frac{x}{L}) + D_k \cosh(\gamma_k \frac{x}{L})$$
(D6)

where we defined  $\gamma_k$  such that

$$\omega_k^2 = \frac{EI}{\rho SL^4} \gamma_k^4 \tag{D7}$$

It should be noted that, according to our definitions, the modes  $w_k$  are dimensionless, as are the coefficients  $\gamma_k$ ; as a result, the modal components  $a_k$  are homogeneous to displacements. The (dimensionless) coefficients  $A_k$ ,  $B_k$ ,  $C_k$  and  $D_k$  are solutions of a homogeneous linear system, which depends on the boundary conditions. For example, for a beam fixed at both ends

$$w_{k}(0) = 0 \Leftrightarrow B_{k} + D_{k} = 0 \quad (D8-a)$$

$$w_{k}'(0) = 0 \Leftrightarrow A_{k} + C_{k} = 0 \quad (D8-b)$$

$$w_{k}(L) = 0 \Leftrightarrow A_{k} \sin(\gamma_{k}) + B_{k} \cos(\gamma_{k}) + C_{k} \sinh(\gamma_{k}) + D_{k} \cosh(\gamma_{k}) = 0 \quad (D8-c)$$

$$w_{k}'(L) = 0 \Leftrightarrow A_{k} \cos(\gamma_{k}) - B_{k} \sin(\gamma_{k}) + C_{k} \cosh(\gamma_{k}) + D_{k} \sinh(\gamma_{k}) = 0 \quad (D8-c)$$

$$d)$$

This system has a trivial solution, except in the case  $\gamma_k$  verifies

$$1 - \cos(\gamma_k) \cosh(\gamma_k) = 0$$
 (D9)

One finds  $\gamma_1 \approx 4.730$ , and  $\gamma_{k+1} \approx \gamma_k + \pi$ .

For a cantilever beam

$$w_{k}(0) = 0 \Leftrightarrow B_{k} + D_{k} = 0 \quad (D10-a)$$

$$w_{k}'(0) = 0 \Leftrightarrow A_{k} + C_{k} = 0 \quad (D10-b)$$

$$w_{k}''(L) = 0 \Leftrightarrow -A_{k} \sin(\gamma_{k}) - B_{k} \cos(\gamma_{k}) + C_{k} \sinh(\gamma_{k}) + D_{k} \cosh(\gamma_{k}) = 0 \quad (D10-c)$$

$$w_{k}^{(3)}(L) = 0 \Leftrightarrow -A_{k} \cos(\gamma_{k}) + B_{k} \sin(\gamma_{k}) + C_{k} \cosh(\gamma_{k}) + D_{k} \sinh(\gamma_{k}) = 0 \quad (D10-c)$$

This system has a trivial solution, except in the case  $\gamma_k$  verifies

$$1 + \cos(\gamma_k) \cosh(\gamma_k) = 0 \qquad (D11)$$

One finds  $\gamma_1 \approx 1.875$ , and  $\gamma_{k+1} \approx \gamma_k + \pi$ .

Regardless of the boundary conditions, one can then choose  $A_k$  in an arbitrary manner, and the coefficients  $B_k$ ,  $C_k$  and  $D_k$  accordingly. In the following, we will assume that we have

chosen  $A_k$  so that  $\max|w_k| = 1$ . The functions  $w_k(x)$  corresponding to this choice are shown in Fig. D1, and the values of the coefficients  $\gamma_k$ ,  $\langle w_k \rangle$  and  $\langle w_k^2 \rangle$  are given in Table D1.



Fig. D1 – Pure bending eigenmodes of a bridge (top left, first three symmetric modes; top right, first three antisymmetric modes) and of a cantilever (bottom).

		k = 1	k = 2	<i>k</i> = 3	k = 4	k = 5
Bridge	$\gamma_k$	4.730	7.853	11.00	14.14	17.28
	$\langle w_k \rangle$	0.523	0	0.241	0	0.153
	$\langle w_k^2 \rangle$	0.397	0.439	0.437	0.437	0.437
Cantilever	$\gamma_k$	1.875	4.694	7.855	11.00	14.14
	$\langle w_k \rangle$	0.392	-0.217	0.127	-0.091	0.071
	$\langle w_k^2 \rangle$	0.25	0.25	0.25	0.25	0.25

Table D1 - Characteristics of bridge and cantilever eigenmodes.

We can look for an approximation of the solution of (D2) as a linear combination of eigenmodes (D5). By replacing w by  $\hat{w}$  in (D2), one obtains

$$\sum_{k=1\dots n} EIa_k \frac{d^4 w_k}{dx^4} + \rho S w_k \ddot{a}_k = f(x, t)$$
$$\Leftrightarrow \sum_{k=1\dots n} \left( \omega_k^2 a_k + \ddot{a}_k \right) \times w_k = \frac{1}{\rho S} \times f(x, t) \quad (D12)$$

By projecting, in the sense of (D4), this equation on the basis formed by the n eigenmodes , a set of n equations is obtained

$$\omega_k^2 a_k + \ddot{a}_k = \frac{1}{\langle w_k^2 \rangle} \times \frac{1}{\rho S} \times f_k(t)$$
(D13)

where  $f_k(t) = \langle f. w_k \rangle$ .

The *n* equations governing the  $a_k$  are independent (as long as *f* does not depend on *w*) and highlight a "mass-spring" type behavior for each of the modes. The coefficients

$$M_{k} = \langle w_{k}^{2} \rangle \times \rho S L \qquad \text{(D14)}$$
$$K_{k} = \omega_{k}^{2} \times M_{k} = \gamma_{k}^{4} \times \langle w_{k}^{2} \rangle \times \frac{EI}{L^{3}} \qquad \text{(D15)}$$

are generally referred to as "modal mass" and "modal stiffness", respectively.

The larger the order k of a mode, the higher its eigenfrequency and the larger its modal stiffness. For a cantilever, for example, the second eigenmode is about 40 times stiffer than the first. For a bridge, the second eigenmode is about 8 times stiffer than the first, and the third is about 30 times stiffer than the first. Indeed, from Fig. D1, one may well understand that a much larger force is required to give a beam the shape of a high order mode. Note also that, by construction, the mass of each mode is less than the mass of the real system.

In practice, the relevant modes to represent the solution of (D2) are the modes

- (i) that are most compliant.
- (ii) whose natural frequency is greater than or equal to the characteristic frequency of f(t).
- (iii) for which the modal force  $F_k = f_k L$  is significant.

The following sections illustrate the use of modal analysis for modeling a beam in different load cases f(x, t).

#### D.2.2 - Uniform load

The boundary value problem is written

$$EI\frac{\partial^4 w}{\partial x^4} + \rho S\frac{\partial^2 w}{\partial t^2} = f(t) + BC \quad (D16)$$

By projecting this equation, in the sense of (D4), on the basis formed by the eigenmodes, we obtain for every k

$$K_k a_k + M_k \ddot{a}_k = F_k(t) = f(t)L \times \langle w_k \rangle \tag{D17}$$

where  $\langle w_k \rangle$  is the average value of the mode on [0, L]. For a cantilever,  $\langle w_k \rangle$  decreases monotonously with k. For a bridge, the  $\langle w_{2k} \rangle$  are all zero, and the  $\langle w_{2k+1} \rangle$  decrease in a monotonous way. Thus, in the case of uniform loading, modal force  $F_k$  is a decreasing function of k.

### D.2.3 - Viscous friction and thermoelastic losses

The beam subject to a viscous friction and the same uniform loading as above is governed by

$$EI\frac{\partial^4 w}{\partial x^4} + c\frac{\partial w}{\partial t} + \rho S\frac{\partial^2 w}{\partial t^2} = f(t) + BC \quad (D18)$$

where c is homogeneous to a viscosity. By projecting this equation on the basis formed by the eigenmodes we obtain, for any k

$$K_k a_k + B_k \dot{a}_k + M_k \ddot{a}_k = F_k(t) \tag{D19}$$

where

$$B_k = \langle w_k^2 \rangle \times cL \tag{D20}$$

is a modal dissipation coefficient.

If thermoelastic losses are also taken into account, we find, below the thermoelastic cutoff frequency (see section 4.4 and Appendix H)

$$EI\frac{\partial^4}{\partial x^4}(w+\tau\frac{\partial w}{\partial t}) + c\frac{\partial w}{\partial t} + \rho S\frac{\partial^2 w}{\partial t^2} = f(t) + BC$$
(D21)

By projection on the basis of eigenmodes, we obtain, for every k

$$K_k a_k + (K_k \tau + B_k) \dot{a}_k + M_k \ddot{a}_k = F_k(t)$$
 (D22)

This form is quite interesting, since it allows a more physical interpretation of the Rayleigh damping, presented at the end of section 4.4: it highlights the fact that the coefficient  $\alpha$  is characteristic of intrinsic losses (such as TED) of the resonator, while coefficient  $\beta$  is characteristic of extrinsic losses (such as fluidic losses).

## D.2.4 – Influence of normal force

A bridge subject to a normal force *N* is governed by

$$EI\frac{\partial^4 w}{\partial x^4} - N\frac{\partial^2 w}{\partial x^2} + \rho S\frac{\partial^2 w}{\partial t^2} = f(t) + BC$$
(D23)

Note that, despite the slightly more complex form of (D23) compared to (D2), it is relatively simple to find a basis of eigenmodes  $v_k$  associated to angular eigenfrequencies  $\varpi_k$  which verify

$$EI\frac{d^4v_k}{dx^4} - N\frac{d^2v_k}{dx^2} - \rho S\varpi_k^2 v_k = 0 + BC$$
(D24)

and thus allowing to "diagonalize" rigorously (D23), assuming that N is independent of w (which excludes the case of a normal stress due to a lengthening of the beam). Rather than taking this rigorous approach, one can use the pure bending eigenmodes  $w_k(x)$  to find an approximate solution to (D23) using Galerkin's method - as pointed out at the beginning of this part, this method is robust to the choice of basis functions. Moreover, it can be verified a posteriori that this choice gives quantitatively correct results.

The projection of (D23) on the basis of the  $w_k$  gives

$$K_k a_k + \frac{N}{L} \times \sum_j R_{kj} a_j + M_k \ddot{a}_k = F_k(t)$$
 (D25)

where

$$R_{kj} = R_{jk} = L^2 \times \langle \frac{dw_k}{dx} \cdot \frac{dw_j}{dx} \rangle \quad (D26)$$

The derivatives of  $w_k(x)$  do not form an orthogonal basis: the matrix **R** is therefore not diagonal - as illustrated below in the case n = 5

$$\boldsymbol{R} = \begin{bmatrix} 4.88 & 0 & 4.05 & 0 & 3.17 \\ 0 & 20.2 & 0 & 7.50 & 0 \\ 4.05 & 0 & 43.2 & 0 & 10.64 \\ 0 & 7.50 & 0 & 75.0 & 0 \\ 3.17 & 0 & 10.6 & 0 & 115 \end{bmatrix}$$

The equations governing the modal components are therefore no longer independent of each other. By restricting our basis to a single eigenmode, k = 1, we have

$$\left(K_1 + \frac{N}{L}R_{11}\right)a_1 + M_1\ddot{a}_1 = F_1(t)$$
 (D27)

which illustrates the influence of normal stress on the effective stiffness of the bridges and their resonance frequency. The coefficients appearing in (4.26) are in fact obtained from (D27).

On the other hand, any diagonalization becomes impossible, whatever the basis functions, as soon as the system becomes non-linear, e.g. when the normal force reflects the effect of the elongation of the bridge, i.e.

$$N = \frac{ES}{2L} \int_0^L \left(\frac{\partial w}{\partial x}\right)^2 dx \qquad (D28)$$

By replacing w by  $\hat{w}$  in (D28), one can write

$$N \approx \frac{ES}{2L^2} \sum_i \sum_j R_{ij} a_i a_j \quad (D29)$$

and the equations governing the modal components are obtained by replacing N in (D25) by the latter expression. By restricting oneself to the first eigenmode, one obtains

$$\left(K_1 + \frac{ES}{2L^3}R_{11}^2a_1^2\right)a_1 + M_1\ddot{a}_1 = F_1(t)$$
 (D30)

This highlights a hardening spring effect (Duffing phenomenon).

### D.2.5 - Electrostatic force

If the beam is separated by a gap g from a plane electrode with lateral dimensions  $b, L \gg g$ , it is subject to electrostatic force density

$$f_{elec}(x,t) = \frac{\epsilon_0 b}{2} V^2 \times \frac{1}{\left(g - w(x,t)\right)^2}$$
(D31)

where *V* is the voltage applied across the beam and the electrode. The projection of  $f_{elec}(x, t)$  on the basis of eigenmodes results in modal forces

$$F_k = \frac{1}{2}V^2 \times \epsilon_0 bL \times \langle w_k, \frac{1}{(g-\widehat{w})^2} \rangle = -\frac{1}{2}V^2 \frac{\partial C}{\partial a_k}$$
(D32)

There is no simple analytical expression of these modal forces. The most common approach, which is well-adapted when w is small compared to the gap g, consists in expanding  $f_{elec}$  in Taylor series. If we restrict ourselves to the first eigenmode of the beam, we can approximate  $F_1$  by

$$\hat{F}_1 = \frac{\epsilon_0 bL}{2g^2} V^2 \times \langle w_1 \rangle \times \frac{1 + 0.017 \tilde{a}_1}{(1 - \tilde{a}_1)^{3/2}}$$
(D33)

for a bridge, or by

$$\hat{F}_1 = \frac{\epsilon_0 bL}{2g^2} V^2 \times \langle w_1 \rangle \times \left( \frac{1 + 0.783 \tilde{a}_1}{1 - \tilde{a}_1} + (0.531 + 0.114 \tilde{a}_1) \log(1 - \tilde{a}_1) \right)$$
(D34)

for a cantilever, where  $\tilde{a}_1 = a_1/g$  is the maximal displacement of the beam non-dimensionalized with respect to the gap. These expressions, obtained in a "semi-analytical" manner from the asymptotic expansion of  $F_1$ , have the advantage of being valid even when  $\tilde{a}_1 \rightarrow 1$ . They also highlight the fact that, for a given amplitude  $\tilde{a}_1$ , electrostatic non-linear effects are less marked for cantilevers than for bridges (and for bridges than for plane capacitors).

It should be noted that, in the general case, modal forces  $F_k$  depend on all modal components  $a_1, ..., a_n$ . This illustrates once again the fact that non-linearity generates coupling between modes.

### D.2.6 - Squeeze-film damping

We can model the effect of a film of air with viscosity  $\mu_{eff}$  squeezed between the beam and the electrode by a force density

$$f_{flu}(x,t) = -\mu_{eff} \frac{b^3}{(g - w(x,t))^3} \times \frac{\partial w}{\partial t}$$
(D35)

A significant number of assumptions, which are assumed to be fulfilled, are necessary to arrive at this expression (see Appendix G). The projection of  $f_{flu}(x, t)$  on the basis of eigenmodes results in modal forces

$$F_k = -\mu b^3 L \langle w_k \cdot \frac{w_k}{(6\Lambda + g - \hat{w})(g - \hat{w})^2} \rangle \dot{a}_k$$
(D36)

in which the dependence of  $\mu_{eff}$  to w and mean free path  $\Lambda$  has been made explicit. Thus the modal forces are written  $F_k = -B_k \times \dot{a}_k$  where coefficients  $B_k$  depend in a non-trivial way on all the set of modal components (another inter-modal coupling).

In the case where one can restrict oneself to the study of the first mode and where  $\Lambda \ll g$ , the following approximations are valid

$$\hat{F}_1 = -\frac{\mu b^{3L}}{g^3} \times \langle w_1^2 \rangle \times \frac{1 + 0.012 \tilde{a}_1}{(1 - \tilde{a}_1)^{5/2}} (\text{D37})$$

for a bridge, and

$$\hat{F}_1 = -\frac{\mu b^3 L}{g^3} \times \langle w_1^2 \rangle \times \left( \frac{1+0.381\tilde{a}_1}{(1-\tilde{a}_1)^2} + (0.165 + 0.091\tilde{a}_1)\log(1-\tilde{a}_1) \right)$$
(D38)

for a cantilever.

## D.3 Numerical models and modal analysis

For complex transducer geometries, obtaining a model of small dimensions - the so-called "reduced order model" - is not as immediate as in the case of a simple geometry, where eigenmodes can be expressed analytically. A preliminary step to obtaining the reduced order model is the construction of a more or less complex numerical model of the structure and its physical environment.

Here, we assume that this numerical model (obtained for example with the finite element method) is put in the form

$$Kx + B\dot{x} + M\ddot{x} = f(x, t)$$
 (D39)

where  $x \in \mathbb{R}^N$  represents the *N* degrees of freedom of the structure (displacements, rotations), *K* and *M* are respectively stiffness and mass matrices of size  $N \times N$ . The dissipation matrix is Rayleigh type, i.e.

$$\boldsymbol{B} = \alpha \boldsymbol{K} + \beta \boldsymbol{M} \tag{D40}$$

and f(x, t) includes the effect of forces acting on the structure (linear or non-linear) and other non-linearities (elongations for example). More generic formulations can of course be considered to take into account, for example, the effect of non-linear damping phenomena, etc.

The direct simulation of (D39), i.e. the determination of the solution x(t) at time steps  $t_1, t_2$  ... is generally not easy because of the dimension N of the problem and because of non-

linearity; at each calculation step  $t_i$ , a non-linear set of N equations must be solved, typically with an iterative method like the Newton-Raphson method. At each iteration of this method, a linear system of dimensional N must be resolved and N non-linear functions must be evaluated (which may in turn require... solving a nonlinear system). These last operations can be extremely costly - in terms of resources and computing time - to such an extent that switching to a reduced order model, of dimension  $n \ll N$ , is highly desirable.

A basis generated from *n* "properly chosen" eigenmodes of the structure,  $n \ll N$ , is the most common choice to approximate the solution of (D39). In this numerical context, an eigenmode  $u_k$  associated with an angular eigenfrequency  $\omega_k$  verifies

$$\boldsymbol{K}\boldsymbol{u}_k - \omega_k^2 \boldsymbol{M}\boldsymbol{u}_k = \boldsymbol{0} \quad (D41)$$

which is the discrete counterpart of (D3). As in the continuous case, the eigenmodes form an orthogonal base, which can be "normalized" arbitrarily, for example so that

$$\boldsymbol{u}_i^T \boldsymbol{M} \boldsymbol{u}_k = \delta_{ki} \qquad (D42)$$

By letting  $\boldsymbol{U} = [\boldsymbol{u}_1, ..., \boldsymbol{u}_n]$  the basis of *n* modes considered, one may approximate the solution of (D39) as

$$\widehat{\boldsymbol{x}}(t) = \boldsymbol{U}\boldsymbol{a}(t) \tag{D43}$$

where  $\boldsymbol{a}(t)$  is the vector of *n* modal components. By injecting  $\hat{\boldsymbol{x}}$  in (D39), we get

$$KUa + BU\dot{a} + MU\ddot{a} = f(Ua, t)$$
  
$$\Leftrightarrow MUWa + (\alpha MUW + \beta MU)\dot{a} + MU\ddot{a} = f(Ua, t)$$
(D44)

where

$$\boldsymbol{W} = diag(\omega_1^2, \dots, \omega_n^2) \tag{D45}$$

is a diagonal matrix of size  $n \times n$ . Equation (D44) is the discrete counterpart of (D12). To determine the equations governing the n unknowns of the problem, Galerkin's method is used, i.e. (D44) is projected onto U, which yields

$$\boldsymbol{U}^{T}\boldsymbol{M}\boldsymbol{U}\boldsymbol{W}\boldsymbol{a} + (\boldsymbol{\alpha}\boldsymbol{U}^{T}\boldsymbol{M}\boldsymbol{U}\boldsymbol{W} + \boldsymbol{\beta}\boldsymbol{U}^{T}\boldsymbol{M}\boldsymbol{U})\boldsymbol{\dot{a}} + \boldsymbol{U}^{T}\boldsymbol{M}\boldsymbol{U}\boldsymbol{\ddot{a}} = \boldsymbol{U}^{T}\boldsymbol{f}(\boldsymbol{U}\boldsymbol{a},t)$$
(D46)

As, according to (D42),  $\boldsymbol{U}^T \boldsymbol{M} \boldsymbol{U} = \boldsymbol{I}_n$ , this last equation reduces to

$$Wa + (\alpha W + \beta)\dot{a} + \ddot{a} = \phi(a, t)$$
 (D47)

where  $\boldsymbol{\phi}(\boldsymbol{a},t) = \boldsymbol{U}^T \boldsymbol{f}(\boldsymbol{U}\boldsymbol{a},t)$ . Equation (D47) is the discrete equivalent of (D13).

The left-hand side of (D47) is nice-looking, since it is diagonalized. In the absence of nonlinearity, the equations governing the modal components could therefore be studied independently. The whole difficulty lies in the right-hand side: although, formally,  $\phi$  is a function of  $\mathbb{R}^n$  in  $\mathbb{R}^n$ , its calculation remains expensive. Indeed, its evaluation requires expand a into  $\mathbb{R}^N$ and to calculate the *N* components of f (as mentioned above, this operation may be very costly). Similarly, in the case of the bending beam, the modal force expressions (D32) and (D36) required the (costly) evaluation of integrals on [0, L].

To get over this last difficulty - which is in fact the only real difficulty of this type of procedure - it is necessary to resort, as in the continuous case, to approximations of modal forces, based on Taylor series, asymptotic developments or any other "machine learning" method adapted to the exact form of non-linearity (electrostatic, fluidic, elongations, etc.). This constitutes a more or less complex / costly task depending on the method used and the validity domain being sought for the reduced order model. Once this task has been performed, the reduced order model is ready to be used.

# E - Beyond plane capacitors

## E.1 Analytical approach

The following formulas are particularly useful when the validity of the plane capacitor hypothesis

$$C = \epsilon_0 \frac{bL}{g}$$
 (E1)

- where g is the gap, and b and L are the lateral dimensions of the conductors - must be questioned.



*Fig. E1 - Flat capacitor type geometry (top), Palmer type (bottom left) and Leus-Elata type (bottom right).* 

The best known is Palmer's formula, which is valid for two facing electrodes of which only one of the lateral dimensions *L* is supposed to be very large compared to the gap, without any particular hypothesis on *b*. The capacitance is written

$$C = \epsilon_0 \frac{bL}{g} \times \left(1 + \frac{g}{\pi b} + \frac{g}{\pi b} \log\left(\frac{2\pi b}{g}\right)\right) \quad (E2)$$

This formula has been adapted by Elata and Leus to take into account the finite height. h of the facing conductors

$$C = \epsilon_0 \frac{bL}{g} \times \left( 1 + \frac{g}{\pi b} + \frac{g}{\pi b} \log\left(\frac{2\pi b}{g}\right) + \frac{g}{\pi b} \log\left(1 + \frac{2h}{g} + 2\sqrt{\frac{h}{g} + \frac{h^2}{g^2}}\right) \right)$$
(E3)

[1] H. B. Palmer, " Capacitance of a parallel-plate capacitor by the Schwartz-Christoffel transformation ", Trans. AIEE, vol. 56, 1927

[2] V. Leus, D. Elata, "Fringing field effect in electrostatic actuators", Technical Report ETR-2004-2, Technion, 2004.

## E.2 Electrostatic boundary element method

The Boundary Element Method is a method for numerically solving certain types of partial differential equations. It is mainly applied to homogeneous (no source term), static (no time dependence) and linear problems. In a MEMS context, it is used to solve electrostatic or fluidic problems. It has the advantage of requiring only a meshing of the boundaries of the problem (unlike finite elements, which also require meshing the inside of the problem) and naturally makes it possible to take into account boundary conditions "at infinity".

A typical example of use is the resolution of Laplace's equation in 3 dimensions

$$\Delta V = 0 \text{ in } \Omega \quad \text{(E4-a)}$$
$$V = \Phi_n \text{ on } \partial \Omega_n, n = 1...N \quad \text{(E4-b)}$$

which corresponds to the equation governing the electrostatic potential created in  $\Omega$  by N conductors with surface  $\partial \Omega_n$  at potential  $\Phi_n$  in the absence of space charge density. In a MEMS context, the resolution of this problem is necessary to determine the value of the capacitance between several structures. The boundary element method (which can be formalized as a Petrov-Galerkin method) makes it possible to numerically compute the charge distribution on the different conductors by requiring only a mesh of the boundaries  $\partial \Omega_n$  of the domain.



Fig. E2 – Principle of the BEM in electrostatics. The potential at any point P is written as the sum of the contributions of all the panels with unknown charge density. By taking P to be at the center of each panel, where the potential is known, as many equations as required can be assembled and solve for the charge densities.

To derive intuitively the numerical formulation of the problem, one can proceed as follows:

1) Discretize the surfaces  $\partial \Omega_n$  of the *N* conductors into "small panels" noted  $\partial \Omega_n^j$ ,  $j = 1 \dots J_n$  on which it is assumed that the load density  $\sigma_n^j$  is uniform.

The potential created at any point *P* of  $\Omega$  by panel  $\partial \Omega_n^j$  is then written

$$V_n^j(P) = \frac{\sigma_n^j}{4\pi\epsilon_0} \times \iint_{\partial\Omega_n^j} \frac{dM_n^j}{r(P,M_n^j)} = \sigma_n^j \times I(P,\partial\Omega_n^j)$$
(E5)

where  $r(P, M_n^j)$  represents the distance between the point *P* and the current point  $M_n^j$  of panel  $\partial \Omega_n^j$ . The double integral  $I(P, \partial \Omega_n^j)$  represents the influence of the charge of panel  $\partial \Omega_n^j$  on the potential at point *P*. The total potential at point *P* created by all the panels is written

$$V(P) = \sum_{n=1\dots N} \sum_{j=1\dots J_n} \sigma_n^j \times I(P, \partial \Omega_n^j)$$
(E6)

The potential constructed in this way necessarily verifies (E4-a), but not necessarily (E4-b). It is then a question of finding the values of the densities  $\sigma_n^j$  ensuring that the boundary conditions are met "at best". For example:

2) Use (E6) to write the value of the potential at the center  $C_m^k$  of each panel  $\partial \Omega_m^k$ , where, by (E4-b), it is known that the potential is equal to  $\Phi_m$  i.e.

$$\forall m = 1 \dots N, \ \forall k = 1 \dots J_n$$

$$V(\mathcal{C}_m^k) = \sum_{n=1\dots N} \sum_{j=1\dots J_n} \sigma_n^j \times I(\mathcal{C}_m^k, \partial \Omega_n^j) = \Phi_m$$
(E7)

All that remains is to solve the linear system (E7) to determine the charge densities  $\sigma_n^j$ , the values of the voltages  $\Phi_m$  being known, and the coefficients of influence  $I(C_m^k, \partial \Omega_n^j)$  which depend only on the geometry of the problem, being easily computable.

There are many variants of this method, whether in the assumptions (non-uniform load density on each panel), in the choice of collocation points (elsewhere than in the center of the panels), in the way of approximating the influence coefficients, or in the way of numerically solving the system (E7).

FASTCAP, a reference boundary element code developed by MIT, is available at this address:

#### https://www.rle.mit.edu/cpg/research codes.htm

For simple geometries (e.g. in 2 dimensions), a few lines of Matlab are sufficient to implement this method and check the range of validity of the expressions (E1) (E2) or (E3).

# F - From Navier-Stokes to squeeze-film damping

+ modal analysis of squeeze-film

To be completed

## G - From Fourier to thermoelastic damping

+ modal analysis of TED

To be completed

## H - Manufacturing processes

Step by step description of the SOIMUMPS process and design rules

http://www.memscap.com/\_\_data/assets/pdf\_file/0019/1774/SOIMUMPs.dr.v8.0.pdf

Step-by-step description of the PiezoMUMPS process and design rules

http://www.memscap.com/\_data/assets/pdf\_file/0020/5915/PiezoMUMPs.DR.1.3a.pdf

Video presentation of the THELMA process from ST Microelectronics

https://www.youtube.com/watch?v=KFBOmLwJyP4

Presentation of the MIDIS process (Teledyne DALSA)

https://www.teledynedalsa.com/download/10592c83-ffca-4ec7-8e5a-eff027d62a78/

General presentation of Bosch products (very nice pictures)

https://www.bosch.com/stories/bosch-mems-sensor-applications/

# I - Summary of poly-Si / mono-Si / Air properties

To be completed

# J - Labs K - Projects